



# Post-hoc Evaluation of Nodes Influence in Information Cascades: The Case of Coordinated Accounts

NICCOLÒ DI MARCO, Sapienza University of Rome, Rome, Italy

SARA BRUNETTI, University of Siena, Siena, Italy

MATTEO CINELLI, Sapienza University of Rome, Rome, Italy

WALTER QUATTROCIOCCI, Sapienza University of Rome, Rome, Italy

---

In the last few years, social media has gained an unprecedented amount of attention, playing a pivotal role in shaping the contemporary landscape of communication and connection. However, Coordinated inauthentic Behaviour (CIB), defined as orchestrated efforts by entities to deceive or mislead users about their identity and intentions, has emerged as a tactic to exploit the online discourse. In this study, we quantify the efficacy of CIB tactics by defining a general framework for evaluating the influence of a subset of nodes in a directed tree. We design two algorithms that provide optimal and greedy post-hoc placement strategies that lead to maximising the configuration influence. We then consider cascades from information spreading on X (formerly known as Twitter) to compare the observed behaviour with our algorithms. The results show that, according to our model, coordinated accounts are quite inefficient in terms of their network influence, thus suggesting that they may play a less pivotal role than expected. Moreover, the causes of these poor results may be found in two separate aspects: a bad placement strategy and a scarcity of resources.

CCS Concepts: • **Networks** → *Network algorithms*; • **Applied computing**; • **Mathematics of computing** → *Trees*; **Graph algorithms**;

Additional Key Words and Phrases: Trees, influence, coordinated inauthentic behaviour

## ACM Reference Format:

Niccolò Di Marco, Sara Brunetti, Matteo Cinelli, and Walter Quattrociochi. 2025. Post-hoc Evaluation of Nodes Influence in Information Cascades: The Case of Coordinated Accounts. *ACM Trans. Web* 19, 2, Article 21 (May 2025), 19 pages. <https://doi.org/10.1145/3700644>

---

## 1 Introduction

Social media has transformed how we connect and share information, thus reshaping some of the mechanisms that we use for engaging with the world. Platforms that have been initially designed

---

The work is supported by IRIS Infodemic Coalition (UK government, grant no. SCH-00001-3391), SERICS (PE00000014) under the NRRP MUR program funded by the European Union - NextGenerationEU, project CRESP from the Italian Ministry of Health under the program CCM 2022, PON project “Ricerca e Innovazione” 2014-2020 and project SEED n. SP122184858BEDB3.

Authors’ Contact Information: Niccolò Di Marco, Sapienza University of Rome, Rome, Italy; e-mail: niccolo.dimarco@uniroma1.it; Sara Brunetti, University of Siena, Siena, Italy; e-mail: sara.brunetti@unisi.it; Matteo Cinelli, Sapienza University of Rome, Rome, Italy; e-mail: matteo.cinelli@uniroma1.it; Walter Quattrociochi, Sapienza University of Rome, Rome, Italy; e-mail: walter.quattrociochi@uniroma1.it.



This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2025 Copyright held by the owner/author(s).

ACM 1559-1131/2025/05-ART21

<https://doi.org/10.1145/3700644>

for entertainment, are increasingly becoming the principal environment in which opinions and views take form, determining a new tendency in information consumption [1–3].

However, this radical change came at the cost of a number of downsides such as online harassment, toxicity and hateful speech [4–6], polarisation [7–11] and the spreading of misleading information [1, 12–18]. In this landscape, special attention has been posed to investigate the phenomenon of **Coordinated Inauthentic Behaviour (CIB)** which, according to Meta’s definition, is “the use of multiple Facebook or Instagram assets, working in concert to engage in inauthentic behaviour, where the use of fake accounts is central to the operation”. More in detail, according to the platform’s Community Standards (<https://transparency.fb.com/en-gb/policies/community-standards/inauthentic-behavior/>), the concept of *inauthentic behaviour* refers to people who “misrepresent themselves on Facebook, use fake accounts, artificially boost the popularity of content or engage in behaviours designed to enable other violations”. Related to CIB researchers investigated the phenomenon of **Coordinated Behaviour (CB)**, which can be defined as an unexpected, suspicious, or exceptional similarity among users of a group [19].

Recent studies highlight that accounts displaying CIB, CB, and social bots (i.e., software agents that communicate autonomously on social media and other platforms, having different uses and purposes [20–23]) may have played a role in relevant events [24] such as political elections [25–27], and in disseminating false information [28–32]. It follows that, beyond detection challenges [19, 33–35], one key reason to study such accounts relies on their potential to manipulate public opinion by swaying the narrative and influencing the perceptions and views of a large audience [36].

In this work, we measure the influence of such potentially malicious accounts as the numerosity of the audience they interact with. In particular, we introduce a general framework that allows us to compare their influence with two different theoretical (post-hoc) models for coordinated accounts placement in information cascades:

- (1) first, we consider an optimal model, in which we identify the number and disposition of coordinated accounts to ensure the maximal influence over the tree, without imposing any constraints on the number of coordinated accounts to be used;
- (2) second, we focus on a greedy strategy in which we have a fixed number of coordinated accounts to place to maximise the influence (i.e., limited capabilities case).

Notably, the algorithms provided can be extended to any other influence problem modelled using binary labels on trees.

Using simulations on synthetic data, we show that, on average, it is possible to exert maximum influence using a limited number of coordinated accounts whose value highly depends on the height of the considered tree. Furthermore, the labelling that corresponds to the optimal placement results is rare compared to the whole set of possible arrangements of node labels. As a case study, we compare the influence obtained by our models in a large dataset made of  $\sim 50K$  X (formerly known as Twitter) cascades, built starting from 1.4M tweets about the 2019 UK political elections [26].

Our results show that, according to our modelling, coordinated accounts exert a much lower influence than the one obtained using both the unconstrained algorithm and the greedy strategy. The reasons for this behaviour have to be searched in two distinct factors: a limited number of available resources (i.e., coordinated accounts) and a bad placement strategy. We conclude by noticing that, in general, the observed placement closely resembles a random placement rather than a specific strategy. Despite the limitations of our post-hoc models, our results suggest that CB may exert a limited influence, contrary to what is expected.

The article is structured as follows: in Section 2, we provide the state-of-the-art of the actual research on the field of CIB and information cascade modelling. Then, in Section 3, we set the notation of the work, starting with some definitions of graph theory and then focusing on our specific model of influence. In Section 4, we present our data and we develop two algorithms to maximise the influence of an information cascade. Finally, in Section 5, we present the results of our simulations on real X information cascades using the developed framework.

## 2 Related Works

This work fits into the research areas of information cascade modelling and an emergent new field studying the effects and properties of CB. To give the right context, in this section, we present relevant literature about these two research areas.

### 2.1 Information Cascades

With the advent of the internet, the modelling of information cascades has gained large importance. In fact, social networks have exposed users to an unprecedented amount of information, thus requesting an understanding of the mechanisms that determine information diffusion. In this context, many techniques have been adopted [37, 38], and machine-learning approaches have been proven to be valid tools [39–41].

Interestingly, recent research focused on X. In fact, the simplest action that can be used to spread posts or messages is the *retweet* [42], providing (in principle) an easy way to follow the posts spreading. However, the X API service does not provide all the required information to reconstruct exactly the information paths. In this landscape, many articles propose heuristics to obtain an accurate approximation of it [43–46]. Notably, some attempts have been made on other social media such as Reddit [47, 48], GitHub [49] Instagram [50] and Facebook [51, 52].

### 2.2 CB

Following the advent of social media and the modelling of information cascades, researchers tried to understand the impact of coordination on information diffusion. This strand of research arises from the observation that coordination of online users may facilitate the aim of disinformation campaigns (or other online manipulations) to reach a larger audience and, thus, exert a greater influence [53, 54]. Although coordination, inauthenticity and harmfulness appear often together, they are distinct concepts [19]. In fact, there are activists (or similar figures) who can promote initiatives characterised by coordinated but authentic behaviour. Also, an opposite example may exist, since a single ill-intentioned user may create fake accounts, thus, providing an inauthentic but non-CB. To detect CB online, recent works used similarities extracted using Network Science tools [16, 19, 33, 36, 55, 56] or temporal synchronicity between users' actions [31, 57–60, 60]. However, only partial results have been obtained for the detection of inauthentic behaviour [33, 61]. In this landscape, a branch of research is focused on two types of malicious actors acting online, i.e., bots and trolls [62]. These two types of actors have a main difference, i.e., the former is automated while humans mostly drive the latter. Even if some works suggest that removing these accounts could effectively reduce online manipulations [63, 64], the results for social bots are debated and often conflicting [24]. Some studies report their significant contribution in spreading low-credibility content [28] and in extremising online communities [65] while others reported a limited role [66]. Moreover, many studies found extensive bot activity in online discussions [67–70]. However, there is a general agreement on the fact that such accounts are becoming more sophisticated due to the evolution of technological tools [71, 72]. This creates new challenges for their detection and removal, with current techniques often criticised [73]. To handle these new problems, recent

approaches avoid classifying the nature of individual accounts (i.e., automated or human-driven) and rather inspect suspicious patterns of coordination. Therefore, for testing our framework, we decide to employ this latter approach.

### 3 Definitions

In this section, we are going to define a framework for representing formally information cascades together with coordinated and non-coordinated users.

#### 3.1 Framework

We consider a directed tree  $T = (V, E)$ , where  $V$  is the set of nodes with  $|V| = n$ , and  $E$  is the set of directed edges. As usual in graph theory, we denote by  $T - v$  the induced subgraph of  $T$  on the vertex set  $V \setminus \{v\}$ . In particular, we may consider each node as a user, while a directed edge indicates the direction of the flow of information among them. We denote the out-neighbourhood of a node  $v$  (without  $v$  itself) by  $N(v)$  and its out-degree by  $d(v)$ . The height of a tree is the longest path from the root to any leaf.

To distinguish between coordinated and non-coordinated users, we assign a binary label  $\ell_v$  to each node, such that  $v$  belongs to the coordinated users if and only if  $\ell_v = 1$ . To keep us general, we say that  $v$  is a 1-node if  $\ell_v = 1$  and a 0-node, otherwise. We collect the labels in a vector  $\ell$ . The labelling determines a partition of the nodes  $V = V_1^\ell \cup V_0^\ell$ , where  $V_i^\ell = \{v \in V \mid \ell_v = i\}$  is the subset of  $i$ -nodes, and its cardinality as  $k_i = |V_i^\ell|$ , for  $i = 0, 1$ . Moreover, we can specialize the out-degree of a node using  $d_i^\ell(v) = |\{w \in N(v) \mid \ell_w = i\}|$ , i.e., the number of  $i$ -nodes in the out-neighbourhood of  $v$ , for  $i = 0, 1$  so that  $d(v) = d_1^\ell(v) + d_0^\ell(v)$ .

Finally,  $p_\ell(w)$  is the binary variable indicating that the parent of  $w$  is a 1-node (i.e., it has a direct interaction with a coordinated account) if  $p_\ell(w)$  is true, else false.

We define the *influence* of a configuration  $(T, \ell)$  as

$$I_\ell(T) = \sum_{v \in V_1^\ell} d_0^\ell(v). \quad (1)$$

Equation (1) counts the number of non-coordinated users that are connected (i.e., share content) to a coordinated one. Therefore, it provides a measure of the impact that coordinated users exert on non-coordinated ones.

We are interested in defining an algorithm that, given a directed tree structure, searches the labelling (i.e., the disposition of coordinated accounts) that maximises (1). We denote the optimum influence among all the labellings by  $I_\ell^*(T)$  and its labelling using the minimum number of coordinated accounts by  $\ell^*(T)$ . Accordingly, we denote as  $V_1^*$  and  $k^*(T) = |V_1^*|$  the set of optimal 1-nodes and its cardinality. We highlight that, when the context is clear, we will omit  $\ell$  and  $T$  from all the previous notations.

For example, consider the tree depicted in Figure 1. Among all the possible labels, the optimal one has  $I^* = 6$  and  $k^* = 2$ . Indeed, even for any labelling of size  $k = 3$ , the influence reaches at most 6.

Interestingly, it is possible to give some constraints over the possible values of  $I^*$  and  $k^*$ . In more detail, let  $\mathcal{T}_n$  be the set of all directed trees with  $n$  vertices. Then,  $\max\{I^*(T) : T \in \mathcal{T}_n\} = n - 1$  and it is obtained for the configuration depicted in Figure 2(a). Note that it corresponds also to the lower bound for  $k^*$ . On the other hand,  $\min\{I^*(T) : T \in \mathcal{T}_n\} = \lfloor \frac{n}{2} \rfloor$  and it is obtained for the configuration depicted in Figure 2(b) which also provides the upper bound for  $k^*$ .

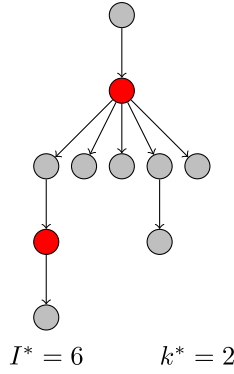


Fig. 1. Example of an optimal configuration. The coordinated accounts are highlighted in red.

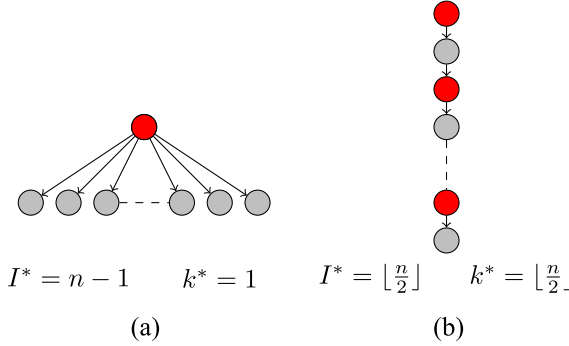


Fig. 2. (a) The configuration that maximizes  $I^*(n)$  and minimizes  $k^*(n)$ . (b) The configuration that minimizes  $I^*(n)$  and maximizes  $k^*(n)$ .

Therefore, for any  $T$  in  $\mathcal{T}_n$  with  $n \geq 2$  vertices, the following inequalities hold:

$$\lfloor \frac{n}{2} \rfloor \leq I^*(T) \leq n - 1, \tag{2}$$

$$1 \leq k^*(T) \leq \lfloor \frac{n}{2} \rfloor. \tag{3}$$

### 3.2 The switch Operator

We now define a useful operator that we will use in the following sections. In particular, the  $switch(v,w)$  operator starts from a labelling  $\ell$  and provides a new labelling  $\bar{\ell}$  that increases the influence obtained by a configuration.

*Definition 1.* Let  $\ell$  be a labelling of nodes of a given directed tree  $T$ . Suppose  $v \in V_1^\ell$  and  $w \in V \setminus V_1^\ell$ ; the switching operator  $switch(v, w)$  exchanges the labels between  $v$  and  $w$ , i.e., it sets  $\ell_v = 0$  and  $\ell_w = 1$ .

Thus, the operator provides a new labelling  $\bar{\ell}$ , where  $V_1^{\bar{\ell}} = (V_1^\ell \setminus \{v\}) \cup \{w\}$ .

To formally define the concept of increment, consider a node  $v \in V_1^\ell$  and  $w \in V \setminus V_1^\ell$ . Switching  $\ell(v)$  with  $\ell(w)$  leads to an increment if

$$d_0^\ell(v) < \Delta(v, w) := d_0^{\bar{\ell}}(w) - p_{\bar{\ell}}(w) + p_{\bar{\ell}}(v), \tag{4}$$

where  $\bar{\ell}$  is the label resulting from  $switch(v, w)$ .

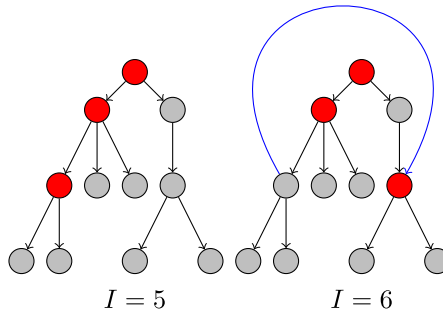


Fig. 3. An example of how *switch* is applied.

The motivation is easily obtained:  $d_0^{\bar{v}}(w)$  counts the number of 0-nodes in the out-neighbour of  $w$ . Moreover, when changing  $v$  with  $w$  it is possible to add 1 to the influence if the parent of  $v$  is a 1-node, i.e.,  $p_{\bar{v}}(v) = 1$ . Similarly, we subtract 1 from the influence if the parent of  $w$  is a 1-node i.e.,  $p_{\bar{v}}(w) = 1$ .

Summarising, if (4) is satisfied,  $switch(v, w)$  assures that  $I_{\bar{v}} > I_{\bar{v}}$ . In the example, we have  $d_0^{\bar{v}}(v) = 2 < 2 - 0 + 1 = 3 = \Delta(v, w)$ . Figure 3 shows a visual example of a *switch* application.

## 4 Data and Methods

### 4.1 Data

Our dataset for this study is based on a collection of tweets related to the online debate about the 2019 United Kingdom general elections presented in [19]. In brief, data were collected using the official X API and performing both hashtag searches (#GeneralElection19, #VoteLabour, #VoteConservative etc.) and timeline downloads of the involved political parties and their leaders. The time window for the download lasted for one month, from 12 November to 12 December 2019, and the volume of downloaded data is 11,264,820 tweets published by 1,179,659 distinct users. The dataset is available at <https://doi.org/10.5281/zenodo.4647893>.

The set of coordinated accounts is also the same provided in [19] in which a method for coordinated account detection is presented. The method is made up of the following steps: (1) Selection of influential users or super-spreaders; (2) Selection of users similarity measure (e.g., cosine similarity); (3) Creation of a users similarity network using pairwise comparison; (4) Filtering of users similarity network; (5) Clustering of the similarity network. This method provides clusters of users provided with a continuous coordination score. The distinction between coordinated and non-coordinated accounts and the reconstruction of information cascades was performed in a later work [26] that we describe in the following lines. Coordinated accounts were selected by retaining only the nodes at the endpoints of the 1% of the links having the highest similarity scores. The number of tweets is 49,331, and for each of them, an information cascade based on retweets was reconstructed using a method proposed in [74] and used in several works [66, 75]. The motivation for using a specific method for cascade reconstruction is due to the fact that X API didn't (and do not) provide the full structure of the retweet trees so that if a user  $j$  retweets  $t$  by user  $i$  and user  $k$  retweets the same tweet  $t$  from the retweet of user  $j$  the edge list provided by the X API will be  $(i \rightarrow j; i \rightarrow k)$ , thus a star. However, this star is not a proper representation of the actual information flow, being it  $(i \rightarrow j; j \rightarrow k)$ . For this reason, the method proposed in [74] provides a deterministic way for reconstructing retweet-based information cascades that relies on the knowledge of friendship networks (i.e., following relationships) and on the assumption that a user will be retweeting a tweet from the latest retweeters of  $t$  contained in her/his list of friends. Therefore,

Table 1. Data Breakdown of the Cascades Collected from X

Number of cascades	49,331
Minimum number of nodes	1
Maximum number of nodes	9,066
Minimum number of coordinated accounts	0
Maximum number of coordinated accounts	236

assuming the same situation as before, but knowing that user  $k$  follows user  $j$ , the reconstructed cascade will be  $(i \rightarrow j; j \rightarrow k)$ . Moreover, when a user shares a post from someone not in her/his friendship network, it is added as a disconnected node. Therefore, in general, we can obtain a forest of directed trees. Using this procedure on the selected 49,331 tweets and their retweets the same number of information cascades was reconstructed. Details about the structure of such cascades are provided in Table 1.

We highlight that several issues may arise when working with this type of data. In particular, researchers must deal with privacy and consent issues when studying CIB, ensuring data is anonymised and used ethically. Additionally, they must comply with relevant laws, such as GDPR, and be aware of the broader societal impacts their findings might have on user experience and platform policies.

#### 4.2 Optimal Placement of Coordinated Accounts without Constraints on Their Number

In this section, we design an algorithm that determines the placement of coordinated accounts in an information cascade maximising the influence. Formally, starting from a directed tree  $T = (V, E)$  rooted at node  $r$ , the algorithm provides  $I^*(T)$  and  $V_1^*(T)$ . We highlight that here we are not imposing an upper bound to the number of 1-nodes, resembling a case of unlimited resource capability.

Let  $T_v$  denote the subtree of  $T$  rooted at  $v$ . In particular,  $T_r = T$ . Given a labelling of the vertices of  $T$ , the influence  $I(T)$  is given by (1). Therefore, the influence for  $T_r$  is obtained by the influences for the subtrees rooted at the children of  $r$ . We denote the children of  $r$  by  $w_1, \dots, w_{d(r)}$ , while  $T_i$  is the subtree rooted in  $w_i$ ,  $i = 1, \dots, d(r)$ . Since after the removal of the root, the trees in the obtained forest are disjoint, it holds  $I(T_r - r) = \sum_{w_i \in N(r)} I(T_i)$ , i.e., the maximum influence of the forest is given by the sum of the maximum influences of their trees. Thus,  $I(T_r)$  is obtained by  $I(T_r - r)$  adding  $d_0(r)$  if  $r \in V_1$  (0, otherwise). Then, it easily follows that the problem of finding a labelling for  $T$  maximizing the influence has an optimal substructure, that is, the optimal solution is constituted by the optimal solutions of the subproblems. Based on this property, we provide a recurrence relation for the maximum influence.

For a node  $v$ , let  $MI(v)$  denote the value of the maximum influence in the subtree of  $T$  rooted at  $v$ , and  $V_1$  denote the set of the coordinated accounts of  $T$ . Moreover, let  $MI_{yes}(v)$  and  $MI_{no}(v)$  be the values of the maximum influence of the subtree rooted at  $v$  that includes  $v$  in  $V_1$ , and that excludes  $v$  from  $V_1$ , respectively. Our aim is to compute  $MI(r) = \max\{MI_{yes}(r), MI_{no}(r)\}$ .

Interestingly, it is possible to give a recursive definition of  $MI(r)$  based on  $MI_{yes}(r)$  and  $MI_{no}(r)$ :

- if  $r$  is a leaf, then  $V = \{r\}$  and  $E = \emptyset$ . In this case,  $MI_{yes}(r) = MI_{no}(r) = 0$ , and  $MI(r) = MI_{no}(r) = 0$  since the maximum influence is 0;
- if  $r$  is not a leaf, then the optimal configuration is computed by considering the maximum influence of the subtrees rooted at the children of  $r$ . Using the same notation as before, the

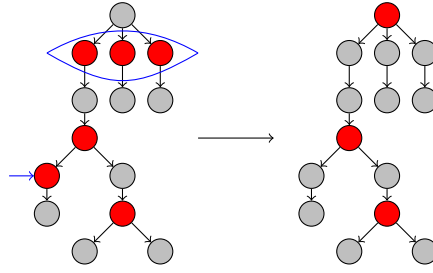


Fig. 4. (a) A directed tree and the label obtained by Algorithm 1. The 1-nodes highlighted by the blue arrows can be removed. (b) After a run of Algorithm 2 the nodes are deleted and  $V_1^*$  is returned.

following recurrence relation describes how to obtain  $MI(r)$ :

$$MI(r) = \max \begin{cases} MI_{\text{no}}(r) = \sum_{w \in N(r)} MI(w) \\ MI_{\text{yes}}(r) = \sum_{w \in N(r)} \max(MI_{\text{yes}}(w), MI_{\text{no}}(w) + 1) \end{cases} \quad (5)$$

Indeed, if the maximum influence  $MI(r)$  is obtained when  $r \notin V_1$  (i.e.,  $MI(r) = MI_{\text{no}}(r)$ ), then it is the sum of the values of the maximum influences for the subtrees rooted at its children. Otherwise, if  $MI(r)$  is obtained when  $r \in V_1$  (i.e.,  $MI(r) = MI_{\text{yes}}(r)$ ), it is the sum of the values of the maximum influences for the subtrees rooted at its children and adding one for every child that is not a coordinated node.

Note that, for instance, if  $r$  is not a leaf, and the height of the tree is one (i.e.,  $T$  is the tree depicted in Figure 2(a)),  $MI_{\text{yes}}(r) = \sum_{w \in N(r)} (MI_{\text{no}}(w) + 1)$ , since if  $r \in V_1$  the maximum influence is equal to the sum of the children nodes not belonging to  $V_1$ . In this case,  $MI_{\text{no}}(w) = 0$  for all  $w \in N(r)$  and so we get

$$MI(r) = MI_{\text{yes}}(r) = \sum_{w \in N(r)} 1 = d(r).$$

The computation of the recurrence relation for  $MI(r)$  can be sped up via dynamic programming and Algorithm 1 accurately describes our approach. We memoize the functions  $MI_{\text{yes}}$  and  $MI_{\text{no}}$  into the tree itself by defining two fields for each node  $v$ , that is  $v.MI_{\text{yes}}$  and  $v.MI_{\text{no}}$ . Notably, they are computed by the functions  $MI_{\text{yes}}$  and  $MI_{\text{no}}$  associated with the children  $w$  of  $v$ . Therefore, a post-order tree traversal realised by one recursive call evaluates both functions at each child in the specified order. Note that in the computation of  $v.MI_{\text{yes}}$ , the addition of one for every child that is not a coordinated node is obtained by counting them and summing them up once. The algorithm computes the label of each node  $v.l$  by comparing  $v.MI_{\text{yes}}$  and  $v.MI_{\text{no}}$ . In general, we include  $v$  in  $V_1$  if and only if  $v.MI_{\text{yes}} > v.MI_{\text{no}}$  (i.e., if  $d_0(v) > 0$ ). In the case where  $v.MI_{\text{yes}} = v.MI_{\text{no}}$ ,  $v$  is not included in  $V_1$  to “minimize” its size. To conclude, since each node is visited once and some constant-time operations are applied in each case, the algorithm runs in  $O(|V|)$  time.

Although Algorithm 1 returns the optimal value of the influence, it is possible that  $|V_1^{\bar{l}}| \geq k^*$ . To solve this problem and obtain  $l^*$ , we use Algorithm 2.

The correctness of the algorithm easily follows: if all the children of  $r$  are 1-nodes and their out-degree is 1, we can simply put  $r$  in the 1-nodes and remove all its children. This procedure reduces the number of used coordinated accounts without changing the influence. Moreover, since each node  $v \in T$  has exactly one father (except for the root), the only case in which  $v$  does not contribute to the influence is when  $d_0(v) = 1$  and its father is a 1-node. The algorithm runs in  $O(k^*)$ .

A visual example of the running of Algorithm 2 is depicted in Figure 4.

**ALGORITHM 1:** TreeMaxInfluence

---

**Require:**  $v$   
**Ensure:**  $I^*(T_v), v.l$

```

 $v.MIno \leftarrow 0$ 
 $v.MIyes \leftarrow 0$ 
if  $v$  is a leaf then
     $v.l \leftarrow 0$  ▷  $v \notin V_1$ 
    return  $v.MIno$  ▷ the Max Influence value is zero
▷ It has at least one child
else
    for each child  $w$  of  $v$  do
         $treeMax \leftarrow TreeMaxInfluence(w)$ 
         $v.MIno \leftarrow v.MIno + treeMax$  ▷ Compute the field MIno of  $v$ 
        if  $treeMax = w.MIno$  then
            Increment count ▷ Count the children  $w \notin V_1$ 
        end if
    end for
     $v.MIyes \leftarrow v.MIno + count$  ▷ Compute the field MIyes of  $v$ 
    if  $v.MIyes > v.MIno$  then
         $v.l \leftarrow 1$  ▷  $v \in V_1$ 
        return  $v.MIyes$  ▷ Return the Max Influence value
    else
         $v.l \leftarrow 0$  ▷  $v \notin V_1$ 
        return  $v.MIno$  ▷ Return the Max Influence value
    end if
end if

```

---

**ALGORITHM 2:** clear1nodes

---

**Require:**  $T = (V, E), \ell$   
**Ensure:** A minimum 1-labelling  $\ell^*$

```

if  $d_0^\ell(r) = 0$  and  $d_0^\ell(w_i) = 1$  for each  $w_i \in N(r)$  then
     $r.l^* = 1$ 
     $w_i.l^* = 0$  for  $w_i \in N(r)$ 
end if
for each  $v \in V_1^\ell$  do
    if  $d_0^\ell(v) = p_\ell(v)$  then
         $v.l^* \leftarrow 0$  ▷  $v \notin V_1$ 
    end if
end for

```

---

**4.3 A Greedy Placement with Constraints on the Number of Coordinated Accounts**

In this section, we propose a greedy strategy to face the case in which we want to maximize  $I$  with a fixed number  $k$  of coordinated accounts, whose pseudocode is presented in Algorithm 4.

We rely on the *switch* operator defined in Section 3.2. In particular, we apply the following strategy: we start by including sequentially the  $k$  nodes which contribute the most to the influence. Then, we try to increment the influence by applying *trySwitch*. When it is no longer possible to increase  $I$  using a call of *switch*, the algorithm stops.

**ALGORITHM 3:** *trySwitch***Require:**  $T = (V, E), \ell$ **Ensure:**  $\tilde{\ell}$  $check = TRUE$ **while**  $check$  **do**▷ Try to apply *switch* $check = FALSE$ **for**  $v \in V_1^\ell$  **do****if**  $d_0(v) < \max_{u \in V \setminus V_1^\ell} \Delta(v, u)$  **then**

▷ It is possible to increase the influence

 $check = TRUE$  $S = \{w \in V \setminus V_1^\ell \mid \Delta(v, w) = \max_{u \in V \setminus V_1^\ell} \Delta(v, u)\}$ choose randomly  $u \in S$  $switch(v, u)$ ▷ Exchange  $\ell(v)$  and  $\ell(u)$ **end if****end for****end while** $\tilde{\ell} = \ell$ **ALGORITHM 4:** An heuristic for the optimal  $k$ -bot placement**Require:**  $T = (V, E), \ell, k$ **Ensure:** an approximation  $\tilde{\ell}$  of the optimal solution**for**  $i = 1 : k$  **do**▷ Place the first  $k$  nodes $W = \{w \in V \setminus V_1^\ell \mid d_0(w) - p_\ell(w) = \max_{u \in V} (d_0(u) - p_\ell(u))\}$ Select randomly a node  $v \in W$  $\ell(v) = 1$ **end for** $\tilde{\ell} = trySwitch(T, \ell)$ 

▷ Call Algorithm 3

## 5 Results

In this section, we conduct some experiments using the framework proposed in Section 4. In particular, we initially provide the growth rate of  $I^*$  and  $k^*$  using a set of randomly generated directed trees. Then, we use our algorithms to evaluate the impact of coordinated accounts over a set of real cascades (see Section 4.1 for further details) by applying Algorithms 1 and 4.

### 5.1 Growth Rate of $I^*$ and $k^*$

Here, we are interested in determining the growth rate of  $I^*$  and  $k^*$  depending on the number of nodes (i.e., users) and the height of the tree. Although those values are bounded, as shown in (2)–(3), finding an explicit expression for them is beyond the scope of this work.

To obtain an approximation, we consider  $N = 100$  random trees made up of  $5 \leq n \leq 100$  nodes. We apply to each of them Algorithm 1 and Algorithm 2 to obtain  $I^*$  and  $k^*$ . Then, we average the results over the  $N$  graph realisations for each of the  $n$  values. We plot the results in Figure 5(a) and (b).

Interestingly, we observe in both cases a linear relation ( $p < 0.001$ ,  $R^2 = 0.99$ ) in which the slope is much lower than 1 ( $\approx 0.28$ ) in the case of  $k^*$  while closer to the unity ( $\approx 0.6$ ) in the case of  $I^*$ . This suggests that it is possible to reach the point of maximum influence using approximately  $0.3 \cdot n$  coordinated accounts. Moreover, they are sufficient to exert a great influence on the tree (approximately  $0.6 \cdot n$ ).

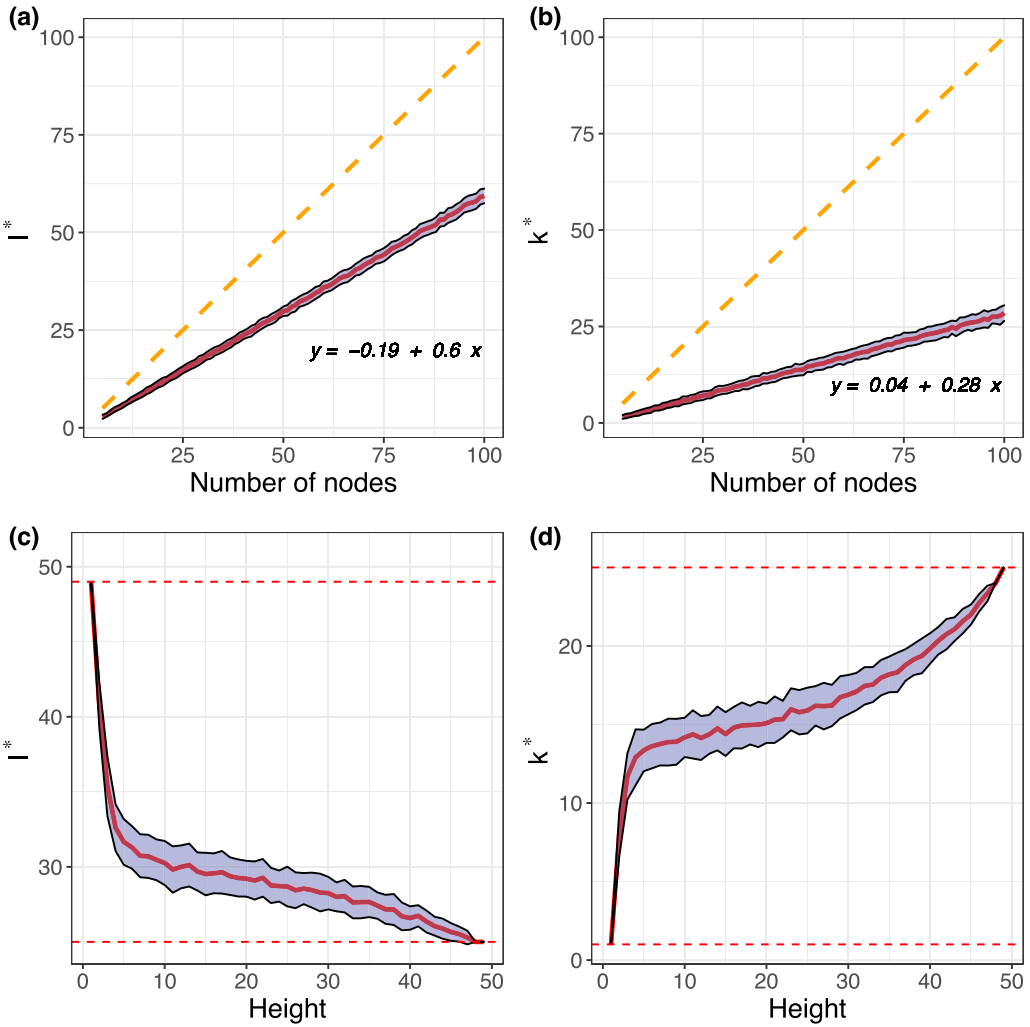


Fig. 5. Upper row: relation between the number of nodes of a tree and (a) mean optimal influence and (b) mean optimal number of bots. Bottom row: comparison between the height of the tree and (c) mean optimal influence and (d) mean optimal number of nodes. The two horizontal red lines show the maximum and minimum values of the measure, while the ribbon around the curves represents a standard deviation from the mean.

To evaluate the impact of the tree’s height on  $I^*$  and  $k^*$ , we consider several random directed trees having a fixed number of nodes ( $n = 50$ ), and height spanning between 1 and  $n$ . For each  $h$ , we generate  $N = 100$  random trees having height  $h$ . Similarly to the previous case, we then average the results obtained by our algorithms, which are shown in Figure 5(c) and (d). As depicted,  $I^*$  drops dramatically as soon as the height increases, and starts to decrease more softly when  $h \geq 10$ . A symmetrical result holds for  $k^*$ . This suggests an expected result, i.e., for a fixed number of users in the conversation, information cascades in which the majority of the retweets are close to the original user (i.e., the associated tree has a low height) can be covered with fewer coordinated accounts, obtaining a greater impact.

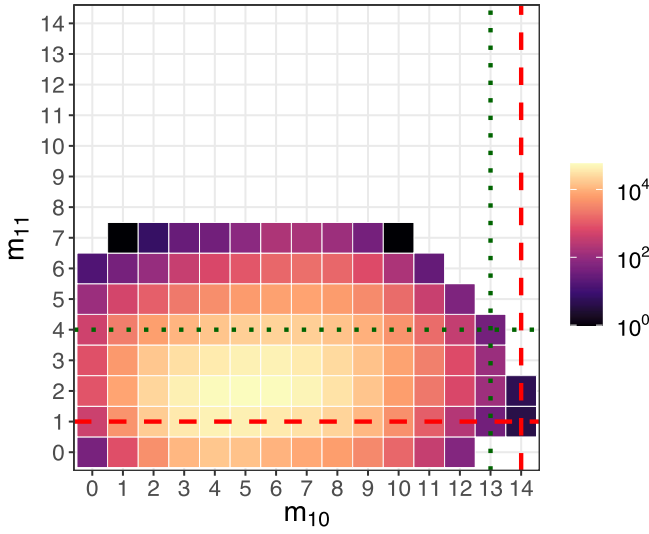


Fig. 6. Phase diagram of possible values of  $(m_{10}, m_{11})$ . The intersection of red (green) lines highlights  $(m_{10}, m_{11})$  for the optimal (greedy) placement. The colours that fill the bins use a logarithmic scale.

However, the Figure also suggests that from a certain height threshold ( $\approx 20 \cdot n$  in our simulation case)  $I^*$  starts to show greater stability with respect to the increase in the height of the tree.

Symmetrical results hold for  $k^*$ .

## 5.2 Phase Diagram of the Relative Positions of Node Labels

In Section 4, we defined an algorithm that finds one of the optimal placements node labels. Interestingly, the same problem can be translated into finding an assignment of binary node metadata such that the number of (directed) edges between 1–nodes and 0–nodes is maximised [76].

In such a framework, let's denote with  $m_{11}$  ( $m_{10}$ ) the number of directed edges between 1 and 1 (0). To better understand how rare the optimal placement is when compared to the set of all the labellings, we computed the phase diagram relative to  $m_{11}$  and  $m_{10}$  [77].

More in detail, we generate a random directed tree of  $n = 25$  nodes and compute its optimal influence, obtaining  $I^* = 14, k^* = 7$ . Then, we generated all the  $\binom{25}{7}$  possible placements, storing for each of them the corresponding values of  $m_{11}$  and  $m_{10}$ . We show a  $2d$ –histogram of the results in Figure 6.

We observe that both the optimal and greedy placements are located in low-frequency areas at the boundary of the diagrams, obtaining high values of  $m_{10}$ . This provides results related to the statistical significance of the obtained labellings and suggests that they are somewhat rare with respect to the whole sets of labels. In general, a placement without a specific strategy is less likely to obtain good results.

## 5.3 Results of the Algorithm on Real Cascades

In this section, we consider a dataset of X's cascades (modelled as directed trees) in which nodes are labelled as either coordinated or non-coordinated as explained in Section 4.1. In each cascade, we aim at detecting the influence exerted compared to the optimum placement provided by Algorithm 1. In particular, the *size* of a cascade is the number of its nodes (i.e., users).

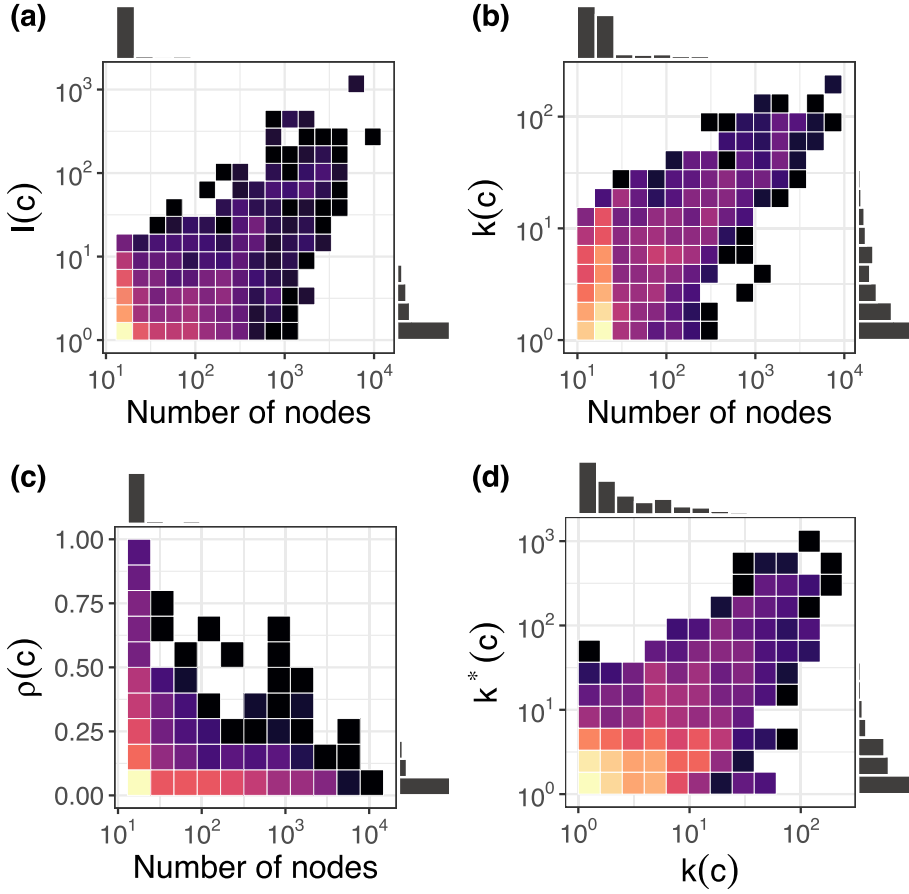


Fig. 7. Results of the algorithm on real cascades. (a) and (b) show the 2-dimensional density of the number of nodes and (a)  $I(c)$  or (b)  $k(c)$ . In (a) we add 1 to the influence of each tree for graphical reason. (c) 2-dimensional density of the number of nodes and  $\rho(c)$ . (d) 2-dimensional density of  $k(c)$  and  $k^*(c)$ . The colours that fill the bins use a logarithmic scale.

For a given cascade  $c$ , we denote as  $I(c)$  the influence obtained by them and as  $k(c)$  the number of coordinated accounts acting in  $c$ . The optimal influence and number of coordinated accounts are denoted accordingly as  $I^*(c)$  and  $k^*(c)$ .

For our analysis, we only consider cascades having at least 15 nodes and at least 1 coordinated account. Such a restriction allows us to consider more complex properties and structures at the cost of losing some of the data. After this procedure, we obtain 4,119 trees to which we apply Algorithms 1 and 2 to compute  $I^*(c)$  and  $k^*(c)$ .

In each cascade, we measure how close  $I(c)$  is to  $I^*(c)$  using

$$\rho(c) = \frac{I(c)}{I^*(c)}. \tag{6}$$

Note that  $0 \leq \rho(c) \leq 1$ , where a value close to one indicates that the influence obtained by the detected coordinated accounts is close to the optimal value.

Table 2. Kullback–Leibler Divergence Values

$D_{KL}(p_{real}  q_{greedy})$	4.278
$D_{KL}(p_{real}  q_{random})$	0.097

As reported in Figure 7, the majority of the cascades have a limited size and, in general, the influence obtained from the coordinated accounts is low, as depicted in (a). In fact, in the biggest trees,  $I(c)$  is approximately 10% of the number of nodes, much lower than the expected 60% suggested by Figure 5(a). A somewhat similar result is observed in (b), which indicates the majority of the cascades contain a very low number of coordinated accounts. Also in this case, approximately 1% of the nodes are coordinated in the biggest cascades, while Figure 5(b) suggests that 30% of the number of nodes is needed to obtain the optimal results. Accordingly, Figure 7(c) shows that only small cascades succeed in obtaining influence values close to the optimal one. For larger cascades, we observe a dramatic collapse in  $\rho(c)$ , probably due also to the much more complex structures arising with a higher number of nodes. Finally, (d) shows that, apart from the smallest trees in which the two measures are approximately similar,  $k(c)$  largely differs from  $k^*(c)$ , yielding a possible explanation for the low values of  $\rho(c)$ .

In summary, we conclude that coordinated accounts exert an influence much lower than the upper bound provided by our algorithm. Moreover, there is evidence that this may be due to a limited number of coordinated accounts placed in each cascade.

#### 5.4 Comparison with Optimum Using a Fixed Number of Coordinated Accounts

In the previous section, we compared the behaviour of coordinated accounts with an optimal strategy that considers having, at prior, an undefined number of coordinated accounts to place. As highlighted before, the scarce results obtained by them could be due to a limited number of resources employed to influence other nodes.

Therefore, here we are interested in unveiling if the observed non-optimal behaviour changes if we compare it to the greedy strategy proposed in Algorithm 4, using exactly  $k(c)$  coordinated accounts. Observing poor results in this case could suggest that they are due also to a bad placement strategy of coordinated accounts.

We denote as  $I_k(c)$  the influence computed by Algorithm 4. Similarly, as before, we measure how close  $I(c)$  and  $I_k(c)$  are using

$$\rho_k(c) = \frac{I(c)}{I_k(c)}. \quad (7)$$

In Figure 8, we show a 2-dimensional histogram of  $\rho_k(c)$  versus the number of nodes in the cascade. Notably, the figure looks very similar to Figure 5(c): only small cascades get comparable results with the greedy algorithm but, when the size of the cascade increases, the placement of coordinated accounts in real cascades hardly reaches the influence obtained through the greedy strategy.

Taking into account all the previous results, it seems that, in real cascades, coordinated accounts may be placed randomly, without a specific rationale. To validate this observation, for each cascade, we generate  $N = 10$  random labellings and we compute the mean influence obtained by those random placements. This results in a new distribution of influence values, denoted as  $q_{random}$ . We also denote as  $p_{real}, q_{greedy}$  the real and greedy influence distributions.

We then compute Kullback–Leibler divergence [78] to compare  $p_{real}$  with the other two distributions, adding a small correction of  $10^{-4}$  to handle the cases in which  $q_{greedy}$  and  $q_{random}$  distributions have 0 probability. The results are shown in Table 2.

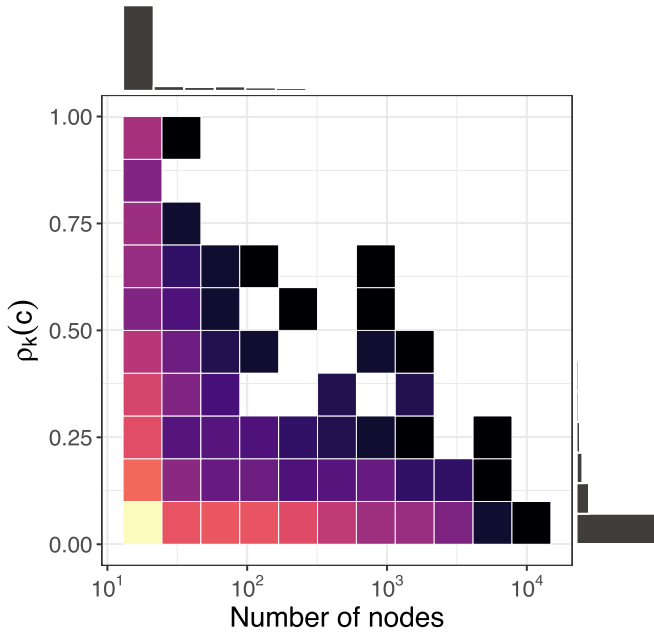


Fig. 8. 2–dimensional density of the number of nodes and  $\rho(c)_k(c)$ . The colours used to fill the bins use a logarithmic scale.

The values suggest that, as expected,  $p_{real}$  is much more similar to  $q_{random}$  than  $q_{greedy}$ , thus confirming that, according to our model, coordinated nodes are placed quite randomly.

Joining these results with the previous ones, we can conclude that, according to our framework, the results observed in real cascades are limited for two reasons: a scarcity of resources and a bad placement strategy. However, since our model takes into account only a part of the complex behaviour of coordinated accounts, several exogenous factors may influence the observed results, e.g., changes in the policies of the platform or geopolitical factors.

### 6 Limitations

In this work, we developed algorithms capable of finding the theoretical upper bound for the influence obtained by coordinated accounts on cascades. However, our model and results present several limitations that we highlight in this section.

First of all, as previously discussed, we used a heuristic to reconstruct information cascades, since restrictions imposed by X API prevent an exact reconstruction of them. Although its results may be inaccurate compared to the real dynamical scenario, this approach has been widely used [66, 74, 75] and, at least at the moment, is one of the more common approaches used to deal with data restrictions.

Moreover, due to a lack of data from other social platforms, we evaluate the performance of coordination only using X data. Thus, our findings may not apply in different contexts or social media. However, our model applies also in the general context being essentially based on the existence of trees having binary labelled nodes. Therefore, while data restrictions may limit the applicability of our framework it should not determine its theoretical validity.

Finally, we only consider one of the possible ways to model the influence over information cascades. In particular, different models may assume that influence spreads through directed paths

of arbitrary length instead of only path of length one (i.e., out-neighbour of each node). On the other hand, it could be interesting to add to each node some measure of their importance (e.g., the number of followers in the case of X) and weigh their influence accordingly. Future works may consider these scenarios and, with the right data available, it would be possible to take into account also the dynamism of real-time cascade, thus providing an even better estimation of the influence of coordinated accounts.

## 7 Conclusion

In this article, we propose a general framework that allows for the evaluation of the influence of coordinated accounts in real information cascades, providing an upper bound to its value.

First, we show (using synthetic data) that it is possible to exert the maximum influence with a low number of coordinated accounts. Moreover, the majority of node labelling gives far-from-optimum results, confirming that random placement is not likely to obtain good results.

As a case of study, we consider  $\approx 4K$  information cascades on X about 2019 UK political elections. We show that observed coordinated accounts exert a very low influence on the tree, compared to both greedy and optimal strategies. Our results suggest that this is due to a double reason: a scarcity of resources (i.e., used coordinated accounts) and an absence of a strategy in how coordinated accounts act.

Despite the possible limitations of our model and the limitations of the detection procedure and the reconstruction of X cascades, our results overall suggest that CIB may play a less pivotal role than expected in information diffusion. Nonetheless, as often reported in social media studies [24], different methodological settings and case studies may lead to different results thus future work may employ the proposed framework to consider the impact of CIB and other similar actors in a wider range of scenarios.

## Acknowledgement

We would like to thank Maurizio Tesconi for providing the dataset employed for the case study.

## References

- [1] Eytan Bakshy, Itamar Rosenn, Cameron Marlow, and Lada Adamic. 2012. The role of social networks in information diffusion. In *Proceedings of the 21st International Conference on World Wide Web*. 519–528.
- [2] Seth Flaxman, Sharad Goel, and Justin M. Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* 80, S1 (2016), 298–320.
- [3] Ana Lucia Schmidt, Fabiana Zollo, Michela Del Vicario, Alessandro Bessi, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. 2017. Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences* 114, 12 (2017), 3035–3039.
- [4] Elizaveta Zinovyeva, Wolfgang Karl Härdle, and Stefan Lessmann. 2020. Antisocial online behavior detection using deep learning. *Decision Support Systems* 138 (2020), 113362. DOI: <http://dx.doi.org/10.1016/j.dss.2020.113362>
- [5] Amaury Trujillo and Stefano Cresci. 2022. Make reddit great again: Assessing community effects of moderation interventions on r/the\_donald. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–28.
- [6] Michele Avalle, Niccolò Di Marco, Gabriele Eтта, Emanuele Sangiorgio, Shayan Alipour, Anita Bonetti, Lorenzo Alvisi, Antonio Scala, Andrea Baronchelli, Matteo Cinelli, and Walter Quattrociocchi. 2024. Persistent interaction patterns across social media platforms and over time. *Nature* 628, 8008 (2024), 582–589.
- [7] R. Kelly Garrett. 2009. Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication* 14, 2 (2009), 265–285.
- [8] Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115, 37 (2018), 9216–9221.
- [9] Joshua A. Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. Social media, political polarization, and political disinformation: A review of the scientific literature. *Political Polarization, and Political Disinformation: A Review of the Scientific Literature (March 19, 2018)* (2018).

- [10] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences* 118, 9 (2021), e2023301118.
- [11] Max Falkenberg, Alessandro Galeazzi, Maddalena Torricelli, Niccolò Di Marco, Francesca Larosa, Madalina Sas, Amin Mekacher, Warren Pearce, Fabiana Zollo, Walter Quattrociocchi, and Andrea Baronchelli. 2022. Growing polarization around climate change on social media. *Nature Climate Change* 12, 12 (2022), 1114–1121.
- [12] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences* 113, 3 (2016), 554–559.
- [13] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.
- [14] Alexandre Bovet and Hernán A. Makse. 2019. Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications* 10, 1 (2019), 7.
- [15] Fernando P. Santos, Yphtach Lelkes, and Simon A. Levin. 2021. Link recommendation algorithms and dynamics of polarization in online social networks. *Proceedings of the National Academy of Sciences* 118, 50 (2021), e2102141118.
- [16] Kristina Hristakieva, Stefano Cresci, Giovanni Da San Martino, Mauro Conti, and Preslav Nakov. 2022. The spread of propaganda by coordinated communities on social media. In *Proceedings of the 14th ACM Web Science Conference 2022*. 191–201.
- [17] Jonas L. Juul and Johan Ugander. 2021. Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences* 118, 46 (2021), e2100786118.
- [18] Francesco Pierri, Luca Luceri, Nikhil Jindal, and Emilio Ferrara. 2023. Propaganda and misinformation on Facebook and Twitter during the Russian invasion of Ukraine. In *Proceedings of the 15th ACM Web Science Conference 2023*. 65–74.
- [19] Leonardo Nizzoli, Serena Tardelli, Marco Avvenuti, Stefano Cresci, and Maurizio Tesconi. 2021. Coordinated behavior on social media in 2019 UK general election. In *Proceedings of the International AAAI Conference on Web and Social Media*. 443–454.
- [20] Carlene Lebeuf, Alexey Zagalsky, Matthieu Foucault, and Margaret-Anne Storey. 2019. Defining and classifying software bots: A faceted taxonomy. In *Proceedings of the 2019 IEEE/ACM 1st International Workshop on Bots in Software Engineering (BotSE)*. IEEE. DOI : <http://dx.doi.org/10.1109/botse.2019.00008>
- [21] Joseph Seering, Juan Pablo Flores, Saiph Savage, and Jessica Hammer. 2018. The social roles of bots: Evaluating impact of bots on discussions in online communities. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–29. DOI : <http://dx.doi.org/10.1145/3274426>
- [22] Lei (Nico) Zheng, Christopher M. Albano, Neev M. Vora, Feng Mai, and Jeffrey V. Nickerson. 2019. The roles bots play in wikipedia. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–20. DOI : <http://dx.doi.org/10.1145/3359317>
- [23] Lorenz Cuno Klopfenstein, Saverio Delpriori, Silvia Malatini, and Alessandro Bogliolo. 2017. The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS'17)*. ACM. DOI : <http://dx.doi.org/10.1145/3064663.3064672>
- [24] Derek Ruths. 2019. The misinformation machine. *Science* 363, 6425 (2019), 348–348.
- [25] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on Twitter during the 2016 US presidential election. *Science* 363, 6425 (2019), 374–378.
- [26] Matteo Cinelli, Stefano Cresci, Walter Quattrociocchi, Maurizio Tesconi, and Paola Zola. 2022. Coordinated inauthentic behavior and information spreading on Twitter. *Decision Support Systems* 160 (2022), 113819.
- [27] Franziska B. Keller, David Schoch, Sebastian Stier, and JungHwan Yang. 2020. Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Political Communication* 37, 2 (2020), 256–280.
- [28] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. *Nature Communications* 9, 1 (2018), 1–9.
- [29] Marcelo Mendoza, Maurizio Tesconi, and Stefano Cresci. 2020. Bots in social and interaction networks: Detection and impact estimation. *ACM Transactions on Information Systems* 39, 1 (2020), 1–32.
- [30] Stefano Cresci. 2020. A decade of social bot detection. *Communications of the ACM* 63, 10 (2020), 72–83.
- [31] Karishma Sharma, Yizhou Zhang, Emilio Ferrara, and Yan Liu. 2021. Identifying coordinated accounts on social media through hidden influence and group behaviours. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1441–1451.
- [32] Diogo Pacheco, Alessandro Flammini, and Filippo Menczer. 2020. Unveiling coordinated groups behind white helmets disinformation. In *Companion Proceedings of the Web Conference 2020 (WWW'20)*. ACM. DOI : <http://dx.doi.org/10.1145/3366424.3385775>

- [33] Diogo Pacheco, Pik-Mai Hui, Christopher Torres-Lugo, Bao Tran Truong, Alessandro Flammini, and Filippo Menczer. 2021. Uncovering coordinated networks on social media: Methods and case studies. In *Proceedings of the International AAAI Conference on Web and Social Media*. 455–466.
- [34] Luca Lucherì, Valeria Pantè, Keith Burghardt, and Emilio Ferrara. 2024. Unmasking the web of deceit: Uncovering coordinated activity to expose information operations on twitter. In *Proceedings of the ACM Web Conference 2024 (WWW'24)*, Association for Computing Machinery, Singapore, Singapore, 2530–2541. DOI: <https://doi.org/10.1145/3589334.3645529>
- [35] Alexander C. Nwala, Alessandro Flammini, and Filippo Menczer. 2023. A language framework for modeling social media account behavior. *EPJ Data Science* 12, 1 (2023), 33.
- [36] David Schoch, Franziska B. Keller, Sebastian Stier, and JungHwan Yang. 2022. Coordination patterns reveal online political astroturfing across the world. *Scientific Reports* 12, 1 (2022), 4572.
- [37] Shayan Alipour, Niccolò Di Marco, Michele Avalle, Gabriele Etta, Matteo Cinelli, and Walter Quattrociocchi. 2024. The drivers of global news spreading patterns. *Scientific Reports* 14, 1 (2024), 1519.
- [38] Fan Zhou, Xovee Xu, Goce Trajcevski, and Kunpeng Zhang. 2021. A survey of information cascade analysis: Models, predictions, and recent advances. *ACM Computing Surveys* 54, 2 (2021), 1–36.
- [39] Xovee Xu, Fan Zhou, Kunpeng Zhang, and Siyuan Liu. 2022. Ccgl: Contrastive cascade graph learning. *IEEE Transactions on Knowledge and Data Engineering* 35, 5 (2022), 4539–4554.
- [40] Hongliang Qiao, Shanshan Feng, Xutao Li, Huiwei Lin, Han Hu, Wei Wei, and Yunming Ye. 2023. RotDiff: A hyperbolic rotation representation model for information diffusion prediction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 2065–2074.
- [41] Aravind Sankar, Xinyang Zhang, Adit Krishnan, and Jiawei Han. 2020. Inf-VAE: A variational autoencoder framework to integrate homophily and influence in diffusion prediction. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 510–518.
- [42] Syeda Nadia Firdaus, Chen Ding, and Alireza Sadeghian. 2018. Retweet: A popular information diffusion mechanism—A survey paper. *Online Social Networks and Media* 6 (2018), 26–40.
- [43] Soroush Vosoughi, Mostafa ‘Neo’ Mohsenvand, and Deb Roy. 2017. Rumor gauge: Predicting the veracity of rumors on Twitter. *ACM Transactions on Knowledge Discovery from Data* 11, 4 (2017), 1–36.
- [44] Qi Cao, Huawei Shen, Keting Cen, Wentao Ouyang, and Xueqi Cheng. 2017. Deephawkes: Bridging the gap between prediction and understanding of information cascades. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1149–1158.
- [45] Cazabet Remy, Nargis Pervin, Fujio Toriumi, and Hideaki Takeda. 2013. Information diffusion on Twitter: Everyone has its chance, but all chances are not equal. In *Proceedings of the 2013 International Conference on Signal-Image Technology and Internet-Based Systems*. IEEE, 483–490.
- [46] Jiang Yang and Scott Counts. 2010. Predicting the speed, scale, and range of information diffusion in Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*. 355–358.
- [47] Jasser Jasser, Ivan Garibay, Steve Scheinert, and Alexander V. Mantzaris. 2022. Controversial information spreads faster and further than non-controversial information in Reddit. *Journal of Computational Social Science* 5, 1 (2022), 111–122.
- [48] Maria Glenski, Emily Saldanha, and Svitlana Volkova. 2019. Characterizing speed and scale of cryptocurrency discussion spread on reddit. In *Proceedings of the World Wide Web Conference*. 560–570.
- [49] Ramya Akula, Niloofar Yousefi, and Ivan Garibay. 2019. Deepfork: Supervised prediction of information diffusion in github. arXiv:1910.07999. Retrieved from <https://arxiv.org/abs/1910.07999>
- [50] Kristo Radion Purba, David Asirvatham, and Raja Kumar Murugesan. 2022. Influence maximization diffusion models based on engagement and activeness on instagram. *Journal of King Saud University-Computer and Information Sciences* 34, 6 (2022), 2831–2839.
- [51] Pei Li and Yunchuan Sun. 2014. Modeling and performance analysis of information diffusion under information overload in Facebook-like social networks. *International Journal of Communication Systems* 27, 9 (2014), 1268–1288.
- [52] Dong Yang, Tommy W. S. Chow, Lu Zhong, Zhaoyang Tian, Qingpeng Zhang, and Guanrong Chen. 2018. True and fake information spreading over the Facebook. *Physica A: Statistical Mechanics and its Applications* 505 (2018), 984–994.
- [53] Serena Tardelli, Marco Avvenuti, Maurizio Tesconi, and Stefano Cresci. 2022. Detecting inorganic financial campaigns on Twitter. *Information Systems* 103 (2022), 101769.
- [54] Serena Tardelli, Leonardo Nizzoli, Maurizio Tesconi, Mauro Conti, Preslav Nakov, Giovanni Da San Martino, and Stefano Cresci. 2024. Temporal dynamics of coordinated online behavior: Stability, archetypes, and influence. *Proceedings of the National Academy of Sciences* 121, 20 (2024), e2307038121.
- [55] Mustafa Alassad, Billy Spann, and Nitin Agarwal. 2021. Combining advanced computational social science and graph theoretic techniques to reveal adversarial information operations. *Information Processing and Management* 58, 1 (2021), 102385.

- [56] Derek Weber and Frank Neumann. 2020. Who's in the gang? Revealing coordinating communities in social media. In *Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 89–93.
- [57] Thomas Magelinski, Lynnette Ng, and Kathleen Carley. 2022. A synchronized action framework for detection of coordination on social media. *Journal of Online Trust and Safety* 1, 2 (2022). DOI : <http://dx.doi.org/10.54501/jots.v1i2.30>
- [58] Fabio Giglietto, Nicola Righetti, Luca Rossi, and Giada Marino. 2020. It takes a village to manipulate the media: Coordinated link sharing behavior during 2018 and 2019 Italian elections. *Information, Communication and Society* 23, 6 (2020), 867–891.
- [59] Yizhou Zhang, Karishma Sharma, and Yan Liu. 2021. Vigdet: Knowledge informed neural temporal point process for coordination detection on social media. *Advances in Neural Information Processing Systems* 34 (2021), 3218–3231.
- [60] Derek Weber and Frank Neumann. 2021. Amplifying influence through coordinated behaviour in social networks. *Social Network Analysis and Mining* 11, 1 (2021), 111.
- [61] Luis Vargas, Patrick Emami, and Patrick Traynor. 2020. On the detection of disinformation campaign activity with network analysis. In *Proceedings of the 2020 ACM SIGSAC Conference on Cloud Computing Security Workshop*. 133–146.
- [62] Kate Starbird. 2019. Disinformation's spread: bots, trolls and all of us. *Nature* 571, 7766 (2019), 449–450.
- [63] Paolo Fornaciari, Monica Mordonini, Agostino Poggi, Laura Sani, and Michele Tomaiuolo. 2018. A holistic system for troll detection on Twitter. *Computers in Human Behavior* 89 (2018), 258–268.
- [64] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Communications of the ACM* 59, 7 (2016), 96–104.
- [65] Massimo Stella, Emilio Ferrara, and Manlio De Domenico. 2018. Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences* 115, 49 (2018), 12435–12440.
- [66] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.
- [67] Samuel C. Woolley. 2016. Automating power: Social bot interference in global politics. *First Monday* 21, 4 (2016).
- [68] Mehrnoosh Mirtaheri, Sami Abu-El-Haija, Fred Morstatter, Greg Ver Steeg, and Aram Galstyan. 2021. Identifying and analyzing cryptocurrency manipulations in social media. *IEEE Transactions on Computational Social Systems* 8, 3 (2021), 607–617.
- [69] Xiaoyi Yuan, Ross J. Schuchard, and Andrew T. Crooks. 2019. Examining emergent communities and social bots within the polarized online vaccination debate in Twitter. *Social Media+ Society* 5, 3 (2019), 2056305119865465.
- [70] Michele Mazza, Stefano Cresci, Marco Avvenuti, Walter Quattrociocchi, and Maurizio Tesconi. 2019. Rtbust: Exploiting temporal patterns for botnet detection on Twitter. In *Proceedings of the 10th ACM Conference on Web Science*. 183–192.
- [71] Stefano Cresci, Marinella Petrocchi, Angelo Spognardi, and Stefano Tognazzi. 2018. From reaction to proaction: Unexplored ways to the detection of evolving spambots. In *Companion Proceedings of the Web Conference 2018*. 1469–1470.
- [72] Facebook Removes Accounts With AI-Generated Profile Photos. Retrieved June 12, 2024 from <https://www.wired.com/story/facebook-removes-accounts-ai-generated-photos/>
- [73] Adrian Rauchfleisch and Jonas Kaiser. 2020. The false positive problem of automatic bot detection in social science research. *PloS One* 15, 10 (2020), e0241045.
- [74] Tom De Nies, Io Taxidou, Anastasia Dimou, Ruben Verborgh, Peter M. Fischer, Erik Mannens, and Rik Van de Walle. 2015. Towards multi-level provenance reconstruction of information diffusion on social media. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. 1823–1826.
- [75] Sharad Goel, Ashton Anderson, Jake Hofman, and Duncan J. Watts. 2016. The structural virality of online diffusion. *Management Science* 62, 1 (2016), 180–196.
- [76] Matteo Cinelli, Leto Peel, Antonio Iovanella, and Jean-Charles Delvenne. 2020. Network Constraints on the Mixing Patterns of Binary Node Metadata. *Physical Review E* 102, 6 (2020), 062310.
- [77] Juyong Park and Albert-László Barabási. 2007. Distribution of node characteristics in complex networks. *Proceedings of the National Academy of Sciences* 104, 46 (2007), 17916–17920.
- [78] Solomon Kullback and Richard A. Leibler. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* 22, 1 (1951), 79–86.

Received 29 December 2023; revised 21 August 2024; accepted 20 September 2024