The Pod People: Understanding Manipulation of Social Media Popularity via Reciprocity Abuse

Janith Weerasinghe*
janith@nyu.edu
New York University

Bailey Flanigan*
bgf35@drexel.edu
Drexel University

Aviel Stein
ajs568@drexel.edu
Drexel University

Damon McCoy
mccoy@nyu.edu
New York University

Rachel Greenstadt
greenstadt@nyu.edu
New York University

ABSTRACT

Online Social Network (OSN) Users' demand to increase their account popularity has driven the creation of an underground ecosystem that provides services or techniques to help users manipulate content curation algorithms. One method of subversion that has recently emerged occurs when users form groups, called pods, to facilitate reciprocity abuse, where each member reciprocally interacts with content posted by other members of the group. We collect 1.8 million Instagram posts that were posted in pods hosted on Telegram. We first summarize the properties of these pods and how they are used, uncovering that they are easily discoverable by Google search and have a low barrier to entry. We then create two machine learning models for detecting Instagram posts that have gained interaction through two different kinds of pods, achieving 0.91 and 0.94 AUC, respectively. Finally, we find that pods are effective tools for increasing users' Instagram popularity, we estimate that pod utilization leads to a significantly increased level of likely organic comment interaction on users' subsequent posts.

ACM Reference Format:
Janith Weerasinghe, Bailey Flanigan, Aviel Stein, Damon McCoy, and Rachel Greenstadt. 2020. The Pod People: Understanding Manipulation of Social Media Popularity via Reciprocity Abuse. In *Proceedings of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3366423.3380256

1 INTRODUCTION

Online Social Networks (OSNs) have emerged as one of the primary forums of personal, political, and commercial discourse. As of 2019, Facebook, YouTube and Instagram had 4.5 billion combined active users [24]. To increase users' engagement with content on their platforms, most major OSNs reorder the content shown to their users using undocumented *content curation algorithms*, which act as 'gatekeepers' to the visibility and influence of content. A large-scale ecosystem of services and techniques has emerged to manipulate these curation algorithms to artificially amplify the reach of content on these platforms. The harms of these manipulation attacks range from increasing the reach of influencers' deceptively promoted content [20] to the amplification of extreme political rhetoric by nation states to influence elections [8, 21].

OSNs have been deploying ad-hoc mitigation strategies as attacks arise [1]. However, one method of subversion that is not well-addressed by these defenses is manual reciprocity abuse performed at scale by *pods*. Pods are online groups designed to facilitate systematic *reciprocity abuse*, a term coined by DeKoven *et al.* [10] describing an agreement between users to interact with each other's content, thereby increasing its popularity and consequent importance to the content curation algorithm. Reciprocity abuse is difficult to defend against in general, because the resulting interaction is generated by real users [10]. Reciprocity abuse using pods is similarly difficult to defend against, but also especially important to address: as shown in this paper, pods are increasingly prevalent and provide effective means of quickly gaining influence. The growing importance of this issue is evident too, in the responses of OSN platforms themselves: in 2018, Facebook removed 10 Facebook-hosted pods, which were confirmed to have violated Instagram's terms of service by facilitating reciprocity abuse between Instagram users. [13]. However, despite sporadic efforts by platforms to instate and enforce regulations against pods, there has been no systematic study of reciprocity abuse facilitated by pods. In this paper, we perform the first quantitative characterization of pods as a method of content curation algorithm manipulation. We study reciprocity abuse through the lens of *Instagram pods*, which are pods used to increase the popularity of users' Instagram content. Our analysis shows that the number of posts advertised in pods is increasing over time, indicating that this type of content curation algorithm manipulation is growing. We collected 1.8 million Instagram post URLs that were advertised in pods. Based on our understanding of how different pods operate, we identify a set of features and train machine learning models to detect Instagram posts that have likely gained interaction as a result of reciprocity abuse through pod usage. Finally, we quantify the efficacy of pods at increasing the popularity of users' subsequent content after using a pod. Our key contributions are:

(1) **Pod Landscape Characterization.** We provide, to our knowledge, the first characterization of a portion of the pod ecosystem. Having collected a dataset of 1.8 million Instagram posts advertised across over 400 Instagram pods hosted on Telegram, we summarize these distinguishing features, usage patterns, and rules of operation.

*Both authors contributed equally to this research.

WWW '20, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7023-3/20/04.
https://doi.org/10.1145/3366423.3380256

(2) **Pod Interaction Detection.** We present classifiers that predict, with AUCs greater than 0.9, whether an Instagram post has been posted pods. With these classifiers, we illustrate that it is possible to distinguish between Instagram posts that have and have not been posted in pods, even when the posts have similar levels of interaction and are posted by users with similar levels of platform engagement.
(3) **Pod Efficacy Assessment.** Pods work. Although we do not have a random experiment to establish causality, we explore the efficacy of pods at increasing popularity by examining changes in post interaction over time across Instagram users' profiles. We show that posting in pods is associated with a significant profile-level increase in organic post interaction.

2 RELATED WORK

Detecting and mitigating inauthentic engagement is a problem that OSNs have been working to address for some time. When OSNs deploy new countermeasures against fake engagements, individuals who are trying to game the platform come up with new approaches to bypass the current mitigations. Over previous years, the OSN developers have focused on preventing content curation algorithm manipulation through the purchase of engagement through fake accounts and automated actions. Stringhini *et al.* [25] investigates Twitter follower markets which sell Twitter follows using automated methods. De Cristofaro *et al.*[9] explored the landscape of "like farms", a related curation algorithm manipulation technique. Since these fake engagements occur nearly simultaneously, approaches that temporally cluster such interactions are successful [6]. Recent advances in scalably detecting this lockstep behaviour (orchestrated actions by sets of users which have a very low likelihood of happening spontaneously or organically) [3] have allowed OSN platforms such as YouTube to successfully address this problem [16]. While these mitigation methods are effective against automated engagement, they cannot easily be applied to detect the manual engagement on posts that comes from pod users.

Another approach to fake engagement uses collusion networks that collect OAuth access tokens from colluding members and use them to provide fake likes or comments to their members. A 2017 study by Farooqi *et al.* [11] first described this thriving ecosystem of large-scale reputation manipulation services on Facebook that leverage the principle of collusion. DeKoven *et al.* [10] characterizes similar collusion networks on Instagram. Viswanath *et al.* [27] proposed a Principal Component Analysis (PCA) based anomaly detection system that uses temporal features such as like counts over time and categorical features such as page types to detect Facebook pages that received likes via similar collusion networks. Farooqi *et al.* showed that mitigation strategies based on temporal clustering are not effective against activity resulting from these collusion networks because the collusion networks tend to spread fake engagements originating from the same account over a large span of time and use a large pool of accounts to engage with a given post. Furthermore, this study shows that although the interactions from collusion networks are performed by a third-party service with access to the user's OAuth tokens, the fact that the colluding accounts belong to real users make them difficult to detect, because these accounts contain a mix of organic and fake activity. Instead, their mitigation strategy was to block the IP address subnets used by the collusion services which would not be effective against pods.

We seek to study a type of manipulation that is *even more difficult* to detect: reciprocity abuse performed manually by real OSN users in pods. Notably, existing mitigation techniques such as access token or IP-based rate limiting does not work against engagements from these types of users, because they are not using third-party Account Automation Services or relying on bot accounts or access to a user's OAuth tokens.

There is some anecdotal evidence on reciprocity abuse through pods [5, 26], in the form of online blog posts and personal experiences of pod users. While this anecdotal evidence suggests that participating in pod activity does increase the interactions a post receives, if this activity generates organic interaction as a result of the increased pod interaction is not known. To our knowledge, there has been little-to-no public research on the occurrence, characterization, or detection of manual reciprocity abuse through pods.

3 BACKGROUND

Pods are groups of OSN users who have gathered to increase the popularity of their content by reciprocally interacting with content posted by other members of the group. To discover pods and better understand how they are joined, we first played the role of an individual trying to discover pods and searched for Instagram pods on Google. These searches led us to numerous blog posts and forums on this topic, from which we learned that pods are predominantly hosted on Telegram, a privacy-focused messaging platform.

If a user wants to use one of these pods, they must first join the Telegram group on which the pod is hosted. Many of these groups can be found through Google searches for curated lists of Instagram pods. To receive engagement on an Instagram post from other pod users, the user must send a message to the group containing the link to their Instagram post. Before doing so, however, the user must interact with other users' posts in the quantity and manner specified by the pod's rules. These rules are often enforced by bots, which may remove pod users for failing to comply with the pod's specified rules of reciprocal interaction. Telegram may have become the preferred pod-hosting platform because there is no member cap for groups (more specifically Telegram Channels), and because their rich API allows group admins to deploy custom bots that can moderate the group activity. These features of Telegram allow pods to scale with minimum admin effort while maintaining engagement quality.

In our initial exploration, we investigated pods hosted on platforms other than Telegram, such as Facebook, Instagram Direct Message groups, and Reddit. While Facebook does host some pods, Facebook is known to remove engagement groups from their platform [13], causing them to be ephemeral and smaller in scale. Instagram Direct Message groups are capped at 50 members, meaning that these groups also operate at limited scale. Moreover, Instagram direct message groups are private and group members are vetted manually by group admins, potentially hindering data collection. Reddit is used primarily as a platform to advertise pod groups hosted on other platforms (most often, Telegram). Given these considerations and our observation of Telegram's overwhelming popularity as a platform for hosting pods, in this study we focus on Instagram pods hosted on Telegram.

While many of these pods seem to be operated by individuals, we also discovered a handful of web-based companies that maintain multiple pod groups and offer pod-related resources, such as video tutorials on how to utilize pods to increase popularity. These companies monetize pods by promoting related products such as "auto likers" and dashboards for measuring profile popularity growth. They also sell access to "elite" or "VIP" pods, and charge fees to get un-banned from groups or clear "leach warnings", which result from failing to comply with the rules of reciprocal interaction.

There are several terms and concepts specific to the pod ecosystem, which we define here and discuss quantitatively in Section 5:
Quantity of required interaction: Pods require the participants to make a certain number of interactions with the other users in the pod. Most commonly, pods use a "Do times N" or "D$x$$N$" system, in which any user who posts an Instagram link on the pod must interact with the previous N links posted on the group. Alternatively, in other pods users are required to interact with all previous links posted in the last H hours.
Type of required interaction: Most pods are focused on increasing their participants' Instagram like and comment counts. There are pods that are designated as increasing likes only, or comments only, or likes and comments both. While rare, we also discovered pods that are designated as follow and save (Instagram allows users to "save" photos on their feed to a private collection).
Entry requirements: Some pods require their members to have a minimum follower count. Most companies who manage pods have multiple pods with varied entry requirements.
Special interest groups: Some pods are designated for Instagram users who post about specific topics such as fashion, food, travel, etc, while other pods are generic.

4 DATA COLLECTION

In this section, we describe the multiple datasets we collected from Telegram (TG) and Instagram (IG). To understand the landscape of Instagram pods, we started by systematically collecting data from public pods hosted on Telegram. We identified 1.8 million Instagram posts that been posted in these Telegram-hosted Instagram pods. We collected data on all of these Instagram posts as well as the 111,455 Instagram users that had posted them. We also collected data on posts from a random, activity-matched set of Instagram users to be used as a control set to train a classifier to predict whether an Instagram post has received interaction through a pod (Section 6).

4.1 Telegram-Hosted Pods

Our dataset contains records of 432 Telegram-hosted Instagram pods. We used an iterative approach to discover Instagram pod groups by starting with a seed set of pods and related groups discovered via a Google search. The key observation driving this exploration was that often, pods are advertised on the message boards of other pods, allowing us to search pod message boards to discover new pods.

A flow diagram of our data collection method is provided in Figure 1. On February 26th, 2019, we performed a systematic Google search (Step(1) on Figure 1) of the following phrases: "Instagram engagement pod list," "Instagram engagement group list," "Instagram engagement groups telegram," and "Instagram pods telegram". These search terms were designed to employ terminology we had observed to be commonly associated with pods, and to emulate what an OSN user seeking to join a pod might search. On the first four pages of Google search results–we found later pages to be irrelevant–we visited each page and collected all Telegram group URLs of Instagram pods and other groups in which individuals discussed Instagram engagement. We then used the Telegram API to download the group metadata and all the messages from these public Telegram groups (Steps (2) and (4) on Figure 1). The group metadata we collected includes: group username, title, description, the group creation date, and the number of members in the group. The message data includes the message content, timestamp, and the user identifier of the Telegram user who sent the message. Most groups also include a *pinned-message* which is displayed at the top and usually specifies group rules. We discovered 163 Telegram group URLs through this process. Of these URLs, 88 of them pointed to Telegram groups that were public and valid (un-expired and referencing a currently-existing group).

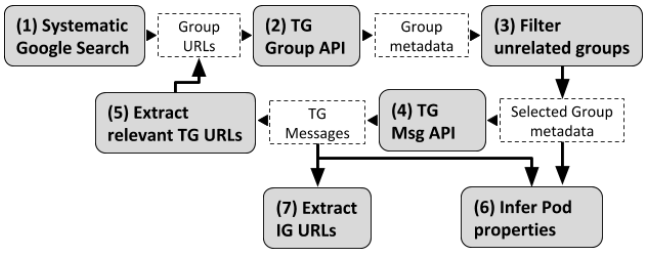


Figure 1: Data collection flow diagram. Solid gray boxes represent major steps, dotted boxes represent resulting intermediate data.

After this first search iteration, we collected Telegram group URLs posted in these groups in search of additional Instagram pods. Because collecting Telegram group metadata is computationally costly, we filtered out URLs unrelated to pods in two steps, the first being a coarser step that required only the Telegram group URL and the body of the message that mentioned the URL. For this step, we made use of the fact that the Telegram group URLs appear in two forms, one with the group's username in the URL and one with a join request hashed in the URL. For the URLs that included the username, it was easy in most cases to determine whether the group was pod-related from the group's username. More information could also be extracted from the content of the message that was associated with the URL. In Figure 1, this step is represented by Step (5), in which we used a Random Forest Classifier and the TF-IDF features of the message content and character n-grams of the group's username (if available) to determine whether a URL linked to a pod-related group. This classifier had a recall of 0.93 and a precision of 0.87. The application of this filter removed Telegram groups that were obviously not pod-related, such as those related to pyramid schemes, cryptocurrencies, online gambling, and porn.

After removing obviously non-pod-related group URLs, we filtered the remaining Telegram groups by examining their group metadata, collected using the Telegram API. This filtering step is depicted in Step (3) in Figure 1, and was done in a semi-automated manner. We used a heuristics-based filter to decide whether a group was an Instagram pod based on the group's username, title, and description. The heuristic marked groups as pods when these attributes contained phrases such as *pod, engagement, like, comment* and the phrase "*Dx*" followed by digits (Such as Dx5 and Dx10) as pods. The heuristic excluded groups containing phrases such as *market, porn, sex, crypto, bitcoin, bet, odds*. We manually labelled groups that were not included or excluded based on the above heuristics. Finally, with this filtered list of newly-discovered pod URLs, we repeated (Steps (2)-(5)), collecting all the messages posted in these groups and again searching them for more pods. We repeated this process for three iterations, ultimately collecting a total of 38,000 group URLs. After discarding URLs that were not pod related based on the predictions of the above classifier, we were left with 4,425 URLs. Out of them, 873 belonged to currently active, public Telegram groups. After the heuristics-based filtering and manual labelling, we identified 432 Telegram groups as pods.

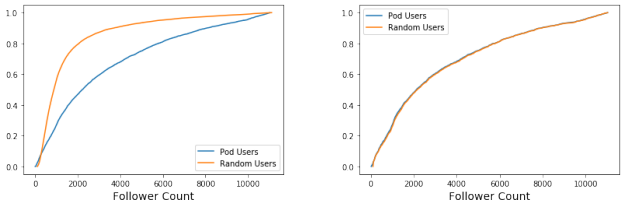
We inferred additional properties of the identified pods using a semi-automated approach (Step (6), which used regular expressions based heuristics to infer whether the pod had any entry requirements, and also its engagement type requirements (comment/like/follow/save), and its engagement level requirements (its Dx*N* value or time-based requirements). For example, to determine if a given pod is a likes and comments pod, we searched for strings such as "Like and Comments," "Likes & Comments," and "L+C" in the group's username, title, description, and the pinned post. We manually annotated the remainder of the groups whose properties could not be inferred via heuristics.

4.2 Instagram Posts by Pod Users

From the message boards of each Telegram-hosted pod identified in the last section, we extracted the URLs of all Instagram posts that had been posted in the pod. We collected data on these posts, including Instagram post identifier, post caption, likes count, comments count, date of post creation, and the Instagram identifier of the user who made the post. Since the Instagram's Developer API does not expose any methods to capture post details, we implemented a scraper that made requests to the Instagram web application and extracted the post metadata from the page response. We made sure to retrieve data only from Instagram profiles marked as public. We collected such data on 1,853,455 unique Instagram posts that belonged to 111,455 unique Instagram accounts.

4.3 Classifier Control Group

To train a classifier to distinguish between pod-interacting Instagram posts and other Instagram posts, we constructed a dataset of "other" Instagram posts. The natural definition of "other" posts would have been those that had not been posted in a pod; however, we could not ensure that a given post was not posted in a pod that we had not seen. Instead, we sought to compare pod-affiliated Instagram posts to the "average" Instagram post.



(a) Unmatched follower counts (b) Matched follower counts

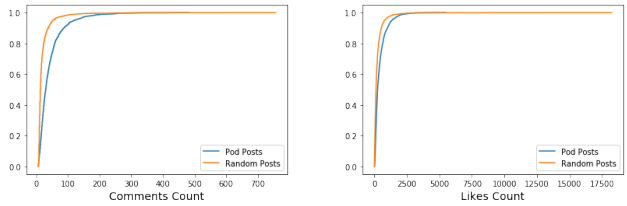
Figure 2: CDF of the number of followers of each pod and random user before and after matching on follower count

We initially tried to capture the "average" Instagram post by populating our control group with posts by randomly-chosen Instagram users. However, as one would expect, these posts had substantially lower engagement (comments, likes, posts, followers) than those by pod users. This large discrepancy in engagement between pod users and non-pod users, captured by follower count, is shown in Figure 2(a). While this difference illustrates that the level of engagement would be an excellent heuristic by which to identify pod-interacting posts, a classifier trained on these datasets would produce many false-positives among Instagram users who have high profile engagement but do not use pods. To demonstrate that it is possible to design a classifier that does not produce these types of false-positives, we chose control-group posts with engagement attributes that matched those of pod-affiliated Instagram users. As such, we show that it is possible to identify pod-interacting posts *even when the posts have similar levels of engagement.*

To collect engagement-matched posts, we first randomly sampled the Instagram user identifier space, in which each user is sequentially assigned an integer upon joining the platform. We sampled identifiers that were close to the identifiers of pod users, thereby ensuring that the random users we collected had created accounts in the same time period as pod users. We downloaded details only from public profiles with English content and at least five posts. Language was detected using the Python-ported version of Google's language-detector library [23] on the user's profile description. During this process, we polled nearly 1.5 million identifiers. After identifying valid and public user identifiers and applying the language and post-count filters, we were left with 16,669 accounts.

We decided to match random users with pod users by follower count, using follower count as a proxy for general engagement attributes. The follower count distribution among random users was extremely right-skewed, with a large fraction of users having follower counts much smaller than those of pod users. We therefore had to poll many random users per pod user to get a reasonable matching. Due to this imbalance, we could only include 1,800 pod users in our analysis.

The matching process was formulated as a linear sum assignment problem, where each user in the pod dataset was paired with random user from the pool (without replacement) such that the sum of the differences of paired users' follower counts was minimized. We used the Hungarian Algorithm as implemented in Scipy [7] to perform the optimization. The results are shown in Figure 2, which shows (a) the original imbalance of follower counts and (b) the quality of our follower-count matching.



(a) Comments Counts (b) Likes Counts

Figure 3: CDF of the number of comments and likes of posts in the *Comment+Like Pods vs Rand. Matched* dataset.

To get posts by each of the matched random users, we randomly selected a post from each user's top 50% of posts by number of comments. For each pod user, we collected one of the Instagram posts we had discovered in a pod. As shown in Figure 3, the posts collected for pod users and random users had similar levels of comments and likes, suggesting that follower count was a good proxy for general engagement attributes.

This process produced two datasets, in which the positive classes consisted of Instagram posts that were posted in *comment and like pods* and *like only pods*. We focused on these two types of pods because they were the most commonly observed in the pod ecosystem (See section 5.1). Each of these two datasets contains 1,800 pod-affiliated Instagram posts and 1,800 matched control Instagram posts. Throughout the rest of the paper, we will refer to the dataset pertaining to comment and like pods as *Comment+Like Pods vs Rand. Matched* dataset, and that pertaining to likes only pods as *Like Pods vs Rand. Matched* dataset.

5 POD LANDSCAPE

In this section, we provide an overview of the general properties of pods we observed. Most notably, we illustrate the recent rapid growth of the discovered pod ecosystem, a relatively high intensity of usage by pod users, and a low barrier to entry to these Telegram groups. We also see heterogeneity in the types of interaction yielded by these pods, and specifically find that half of discovered pods require comment interaction. This is promising because, as discussed in Section 6, the content of comments can be used to detect pod interaction on Instagram posts.

5.1 Pod Ecosystem Evolution

The ecosystem of 432 Telegram-hosted Instagram pods we observed came about largely in the past two years. The earliest-created pod in our dataset became active in late 2016. The ecosystem then grew steadily throughout the rest of 2016 and early 2017. By the end of 2017, it had begun growing more rapidly, and it has since continued to grow at an increasing pace. Figure 4 shows a frequency histogram of the Instagram posts posted in three major types of pods: comments only, likes only, and like and comment pods. Note that the dip in activity during March to July 2019 was due to a data collection issue where we were unable to collect the full number of messages from the Telegram API. This figure serves as a lower bound for the activity level of pods and shows that the activity in pods is growing rapidly.

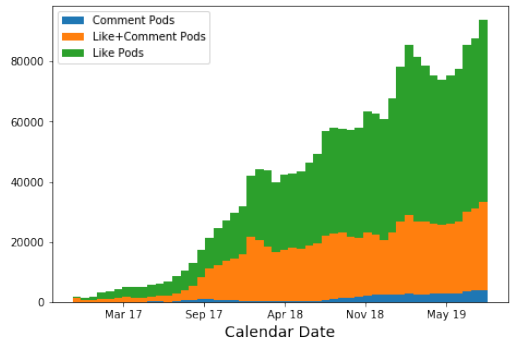


Figure 4: The number of IG posts posted on Comment, Comment+Like, and Like pods on each day.

5.2 Pod Attributes

We first observed that the pods we captured varied on several qualitative and quantitative attributes.

Quantity of Required Interaction. Pods require users to interact with different numbers of Instagram posts, and impose this constraint within three rule frameworks. The most common rule framework for requiring reciprocal interaction is "Do times N" or "DxN". This framework requires that before a pod users sends their own Instagram post to the pod, they first must interact with the previous N Instagram posts on the pod. Of the 385 total pods which specified engagement requirements in their rules, we found that 71% percent were Dx pods. Of these, almost all of them were either Dx5 or Dx10, comprising 34% and 30% of Dx pods, respectively. The remaining Dx pods required between 2 and 87 reciprocal interactions. Alternatively to Dx, another 24% of total pods were Time pods, which require users to interact with all Instagram posts sent to the pod message board in the previous H hours. 80% of Time pods required interaction with all posts in the last 24 hours, and the rest required it in the last 6 or 12 hours. The remaining 5% were Rounds pods. These operate via scheduled rounds, during which members are given a one-hour window to reciprocate engagements.

Type of Required Interaction. Pods also varied on the *type* of interaction with Instagram posts they require. There were four main ways pods required users to interact with Instagram posts: liking a post, commenting on a post, following a user, and saving a post. Of the 432 pods on which we collected data, 41% were Like pods, meaning they required users to interact with other pod users' Instagram posts by simply liking them. Another 37% of pods were Comment & Like pods, meaning they required either or both types of interaction. Another 7% were Comment Only pods.

Entry Requirements. Of the 432 pods we discovered, we found that only 4% required Instagram users to have some minimum number of followers prior to joining. For those that did have entry requirements on follower-count, these requirements ranged between 1,000 and 100,000 followers.

Special Interest Topics. Of the pods we discovered, we detected only five that focused on special interest topics such as fashion, photography, or entrepreneurship.

Quantitative Attributes. We present additional pod attributes in Table 1. Our dataset includes pods of of varying sizes, with six pods

Variable	Mean	Median	Range
POD ATTRIBUTES (per pod)			
Number of users	918	156	0 – 17,301
Lifespan of pods (years)	0.98	0.87	0 – 3.21
Mean number of pod messages in a day over lifespan	41	7	1 – 1,456
Max number of pod messages in a day over lifespan	165	57	1 – 4,218
POD USER ATTRIBUTES (per IG user)			
Number of pods used by each IG user	2.77	2.0	1 – 96
Number of messages per pod by IG user	11.43	3.0	1 – 5,838
Time between IG post and pod post	1.28 hrs	1 min	2 s – 8 yrs

Table 1: Table of pod and pod user features. The pod attributes were computed over 432 pod groups. The pod user attributes were computed over 111,455 Instagram users we discovered that participated in pod groups.

having more than 10,000 members and over half having fewer than 350 members. 95% of the pods in our dataset had fewer than 5000 members. The pods that had more than 10,000 members accounted for nearly 40% of the Instagram posts that we discovered and the top 25 pods, based on the membership count, accounted for 54% of the Instagram posts discovered. This suggests that the pod ecosystem is heavily skewed, with a small number of pods accounting for most of the pod activity. We computed the lifespan of a pod as the time duration between the first and the last message posted to its message board. The average pod lifespan was about 0.87 years. We measured the number of daily messages that were posted on a given pod to measure its activity and saw that activity in some pods had died down at the time of data collection while other pods are gradually increasing in activity. We also observed heterogeneity in how users use pods. The average user was recorded to have used more than two separate pods. In each distinct pod a user utilized, they sent an average of 11 messages, although half of users sent an average of three or fewer messages per pod. Over half of users advertised their posts in a pod within an average of one hour after posting them on Instagram.

6 PREDICTING POD USAGE

In this section, we present our approach to building a classifier that can detect Instagram posts that received engagement through pods, specifically when the pod posts and control posts have similar engagement-level distributions. We designed this classifier based on the hypothesis that the interactions users received through pods would be quantitatively and qualitatively different from interactions that a user would receive from an organic follower base.

6.1 Features Used

Engagement Features: We used the following values to measure the level of engagement on each post:

- *Comments Count*: Number of comments at time of data collection
- *Likes Count*: Number of likes at time of data collection
- *Comments Count : Likes Count (Ratio)*

Comment Timeseries Features: We anticipated that the distribution of interactions a post would receive over time through pods would be different from that of a post receiving interactions organically. Since comments were the only time-stamped interactions available to us, timeseries features were computed only for comments.

- *Time until Zero-Equilibrium*: Number of seconds until Zero-Equilibrium, which is defined as the first 24-hour span after the Instagram post is posted in which it receives zero comments.
- *Proportion of Comments Count before Zero Equilibrium*: Proportion of total comments (Comments Count) received before Zero-Equilibrium.
- *Seconds until Zero-Equilibrium : Comments Count (Ratio)*
- *Comments Count in Hour* $i \in \{1...24\}$: Proportion of total comments received in the first, second,..., twenty-fourth hour.

Comment Content Features: We anticipated that the comments made as a result of reciprocal interaction agreements would have a different linguistic signature than comments that were made organically. To capture this difference, we use two topic modeling approaches, Latent Dirichlet Allocation (LDA) [4] and Non-negative Matrix Factorization (NMF) [15] to discover the underlying topics that the commenters use. LDA learns the relationships between words, topics, and documents by assuming documents are generated by a particular probabilistic model, whereas NMF learns from an unnormalized probability distribution over topics [14]. We concatenated all the English comments made on each post and ran both LDA and NMF algorithms on these comments. To evaluate if the topics learned by these models are stable, we used k-means clustering with two clusters and measured the purity of the clusters [19]. We trained the LDA and NMF model 100 times to account for randomness. The purity was consistently between 82.1-85.2% with an average of 84.1%, demonstrating the stability of the topics. We noticed that the topics learnt by the two approaches were complementary and decided to include features from both models in to our feature set. For NMF the regularization mixing parameter was set to 0.5, alpha was set to 0.1, and initialization was set to Non-negative Double Singular Value Decomposition. For LDA we set the maximum number of iterations to 5, used the 'online' learning method, set the learning offset to 50. All other parameters were set to the default. The posterior probabilities of each topic was used as a feature. We also computed the diversity of the comments made on each post. This is similar to the approach taken by Jang *et al.* [12], in which they computed the entropy of LDA topic probabilities to measure the diversity of a user's posts. Entropy has also been used to detect spam comments on social media.[2]. We computed diversity by measuring the entropy of the LDA and NMF topics and the entropy of the (Term Frequency - Inverse Document Frequency) TF-IDF values of the comments. The entropy is defined as follows:

$$E(X) = \sum_{i=0}^{N} P(X_i) log(P(X_i)$$

In the case of the topic models, $P(X_i)$ was taken as the posterior probability given by the model for each topic i. For the TF-IDF values, $P(X_i)$ was taken as the normalized TF-IDF value for ith token. The following is a list of all the comment content features that we computed: (1) *NMF Topic* $i \in \{0....19\}$ (2) *LDA Topic* $i \in \{0...19\}$ (3) *NMF Entropy* (4) *LDA Entropy* (5) *TFIDF Entropy*

6.2 Prediction Model and Results

We computed values for the above features on our two datasets and trained three different supervised classifiers: Support Vector Machine (SVM) with a Linear Kernel (SVM - Linear), SVM with an RBF Kernel (SVM - RBF), and a Random Forest Classifier. We report the performance of the three classifiers on these two datasets in Table 2.

Model	F1	Precision	Recall	AUC
Like Pods vs Rand. Matched				
SVM - Linear	0.75	0.76	0.75	0.83
SVM - RBF	0.81	0.81	0.81	0.88
Random Forest	0.83	0.83	0.83	0.91
Comment+Like Pods vs Rand. Matched				
SVM - Linear	0.80	0.80	0.80	0.87
SVM - RBF	0.86	0.86	0.86	0.93
Random Forest	0.87	0.87	0.86	0.94

Table 2: 10-Fold cross-validated classification results.

The results in Table 2 shows that the random forest classifier performs the best on both datasets. The performance results from the *Comment+Like Pods vs Rand. Matched* dataset performs better across all classifiers. This behaviour is expected since our time-series and content features are based on comments.

Since we were only able to train our classifier on a smaller dataset due to the limited number of random IG user accounts we could collect, we also ran our classifier on known pod posts to determine the True Positive Rate (TPR) (The number of true positives divided by the total number of positives) of our classifier on a larger dataset. To do this, we trained our classifier on the full *Comment+Like Pods vs Rand. Matched* dataset and tested it on 5000 posts that were posted in comment and like pods. The TPR on this bigger test set was 0.91. For comparison, on our smaller *Comment+Like Pods vs Rand. Matched* dataset, we achieved a TPR of 0.87 and a False Positive Rate (FPR) (The number of false positives, divided by the total number of negatives) of 0.14. This further indicates that our classifier would have a comparable-or-better level of performance to that in Table 2 on larger datasets.

6.3 Feature Analysis

We performed a feature analysis to determine which of the features were important in making these predictions. We measured the "importance" of each feature with Shapley Additive Explanations, or SHAP values [17, 18], which captures the contribution of each feature based on local explanations [22] and principles of game theory. The 15 features with the highest SHAP values for both of our datasets are plotted in Figure 5.

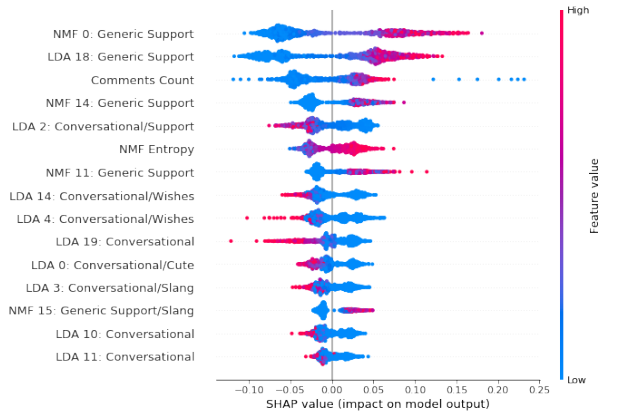
Most of the important features for the *Comment+Like Pods vs Rand. Matched* dataset (Figure 5 (a)), according to their SHAP values, are NMF or LDA topics. In Table 3, we display the top 10 tokens that describe several important topics for this dataset. Five topics fell into the category of "generic support," which contain words with positive connotations that are relevant to Instagram posts in general, but not necessarily to specific posts. Furthermore, most of the topics identified by NMF was different types of "generic support" topics. Therefore, having a high NME entropy, which suggests that the comments were from a variety of these topics, were predictive of pod affiliated posts. The prevalence of generic support words such as "shot," "great," "nice," and "love" among important features indicates that comments high in these types of generic encouragement increases the probability of a positive prediction. It is logical that this should be the case, as comments by strangers through obligatory reciprocal interaction agreements are likely to be more generic than those by organic and voluntary commenters, who are connected with the user through personal relationships. On the other hand, topics that describe more conversational language were important in predicting the control posts. Unsurprisingly, a higher number of comments, when compared to users in the control group who had a similar distribution of follower counts, was an important feature in predicting pod posts.

Figure 5 (b) shows the important features for the *Like Pods vs Rand. Matched* dataset. When compared to the *Comment+Like Pods vs Rand. Matched* dataset, fewer of the content related features were ranked as important. However, comments in positive posts of this dataset too included "generic support" type language and were predictive of pod affiliation. Apart from these features, having a high comments and likes count and skewed comments-to-like ratio were predictive of pod usage.

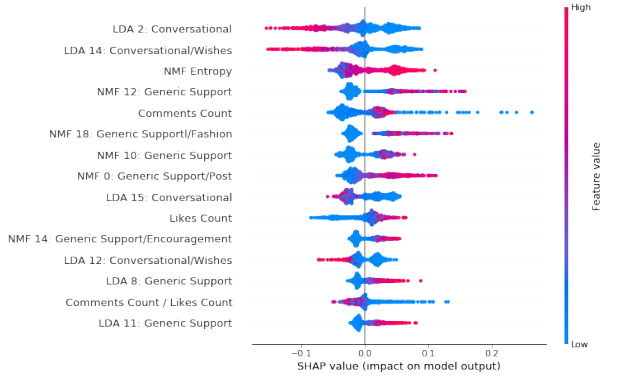
6.4 Misclassifications Analysis

To identify scenarios in which our classifier made mistakes, we trained our classifier on 70% of the data, leaving the remaining 30% for testing and misclassification analysis. We then studied the SHAP values of the misclassified posts to identify which features contributed the most to push the classifier towards an incorrect decision. We display these values in SHAP force plots, which show the contribution of each feature in pushing the classifier probability to its predicted value from the baseline, where baseline is the average of the prediction value in the training set.

We included only six SHAP force plots in Figure 6 for brevity, but ultimately evaluated a sample of 20 misclassified instances. In most force plots, false-positive posts had higher SHAP values for generic support topics (NMF topics 0, 14, and 11, and LDA topic 18). Force plots for false-negative posts show that, although these posts were posted on a comment pod, they had a low value for generic support topics probabilities. This suggests that some posts that were posted on pods did not receive as much comment content containing generic support topic words. Manual examination of some of the pods in these misclassified instances revealed that pods which produced interaction on false-positive posts often required the participants to post high-quality comments that either related to the post caption or met a length requirement. We performed a similar analysis for the *Like Pods vs Rand. Matched* dataset and



(a) Important Features for *Comment+Like Pods vs Rand. Matched* dataset



(b) Important Features for *Like Pods vs Rand. Matched* dataset

Figure 5: SHAP values for the most important features.

found that most misclassifications in this dataset were also caused by similar topic related features.

7 POD EFFICACY ANALYSIS

In this section, we provide evidence suggesting that utilizing pods increases users' Instagram popularity. We define an Instagram user's "popularity" by the level of interaction they receive that is not a direct result of pod-based reciprocal interaction agreements. We refer to interaction of this nature as "organic" interaction. To measure the effect of pod utilization on the popularity of users, we compare the average levels of organic interaction each user receives on their posts before and after they begin posting in pods. To measure changes in organic interaction, we focus our analysis on "control" posts — Instagram posts that have not been posted in a pod. We do so because by definition, all interaction on these posts is organic, so any measured increase in interaction after pod utilization *must* be due entirely to an increase in organic interaction. In contrast, organic interaction with non-control posts cannot be separated from interaction they have received through the pods in which they are known to have been advertised. Importantly, as previously discussed we cannot observe whether a post is truly

Topic	Top Word Content
NMF 0 (Generic Support)	great, shot, post, awesome, good, love, like, really, photo, picture
LDA 18 (Generic Support)	nice, great, wow, cool, shot, amazing, awesome, like, post, really
NMF 14 (Generic Support)	nice, pic, super, photo, sweet, friendship, für, future, funny, fun
LDA 2 (Conversational/Support)	cool, work, wow, amazing, people, beautiful, want, awesome, just, oh
NMF 11 (Generic Support)	wow, amazing, looks, dope, crazy, oh, omg, art, like, incredible
LDA 14 (Conversational/Wishes)	congrats, congratulations, lol, gonna, lmao, right, guys, bad, read, im
LDA 4 (Conversational/Wishes)	happy, birthday, day, year, hope, enjoy, boo, tea, anniversary, bday
LDA 19 (Conversational)	just, xx, omg, like, yes, let, thank, come, good, hi
LDA 0 (Conversational/Cute)	cute, lol, oh, know, like, omg, haha, just, got, adorable
LDA 3 (Conversational/Slang)	bro, shit, time, guys, man, got, fam, couple, stay, like
NMF 15 (Generic Support/Slang)	cool, really, post, sharing, super, interesting, like, dope, friday, fucking
LDA 10 (Conversational)	wait, miss, yay, sexy, ha, tysm, excited, baby, girl, handsome
LDA 11 (Conversational)	ya, funny, man, dope, thx, shirt, don, miss, tbh, dude

Table 3: Top words within the most important NMF and LDA topics for the *Comment+Like Pods vs Rand. Matched* dataset

a control post. In order to execute this analysis, we use our classifier to identify probable control posts and then account for the uncertainty of our classifier in our analysis.

7.1 Methods

Dataset Construction

We randomly 800 Instagram users for this analysis, all of whom who had posted in a pod, and collected all posts on their profiles. On each post, we collected the features necessary for the Random Forest classifier, as specified in Section 6.

At ground truth, every Instagram post is either a *pod post*—an Instagram post that has been posted in one or more pods, or a *control post*—a post that has never been posted in a pod. As previously discussed, we cannot know this ground truth about an Instagram post unless it was posted a pod that we discovered. To infer this information, we defined pod posts to be those that are known to be in a pod and/or are predicted by the random-forest classifier to have been sent to a pod message board with greater than 0.5 confidence. The remaining posts were defined as control posts.

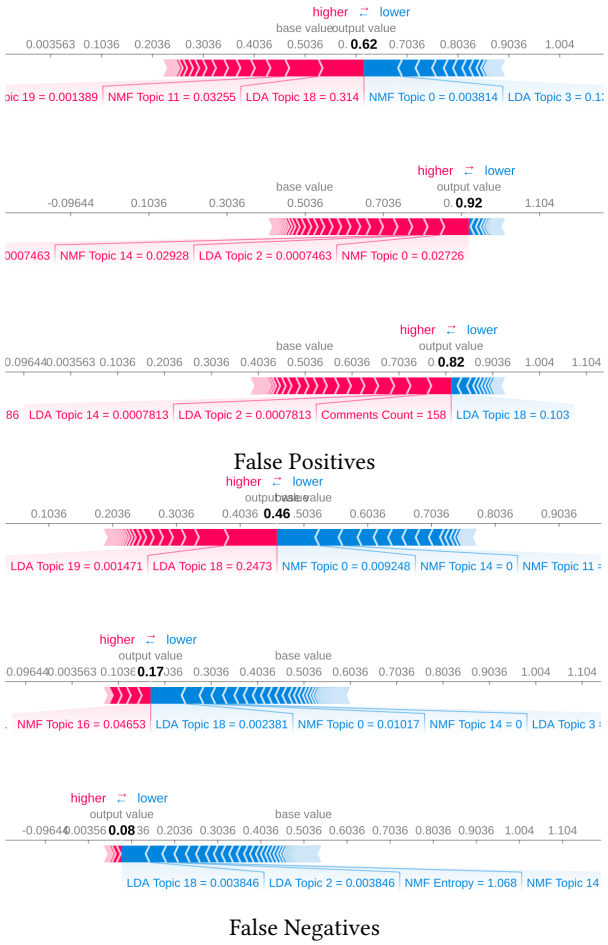


Figure 6: SHAP Force plots for three false positive and three false negative instances in the *Comment+Like Pods vs Rand. Matched* dataset that show the top features that are pushing the prediction from the base value to the model output

We constructed a chronological timeseries of posts for each user, where each timeseries was separated into two periods. The first period is the time before the user's first pod post. All posts in this period are control posts. This period can be considered a *baseline* period, because the average level of interaction attained on the posts in this period represents a baseline level of interaction the user received per Instagram post prior to using pods. The subsequent *pod-interacting* period spans the time *after* the user's first pod post. The posts made in this time may be control posts or pod posts.

Analysis

We index each user's posts by i, ordered chronologically in time. We define the level of interaction on a user's ith post as the number of comments received on that post, C_i. We define indicator variables A_i and B_i, where $B_i = 1$ if post i occurs in the baseline period (*before* the user's first pod post), and $A_i = 1$ if post i occurs in the pod-interacting period. We represent the classifier's confidence on post i as ρ_i. From these variables, we compute α and β, where β and α are the average number comments per post before and after a user's first pod post, respectively. To account for the uncertainty of the classification, these averages are weighted by the certainty of prediction on each post.

$$\alpha = \frac{\sum_i A_i C_i \rho_i}{\sum_i A_i \rho_i} \qquad \beta = \frac{\sum_i B_i C_i \rho_i}{\sum_i B_i \rho_i}$$

We quantify the increase of activity on a user's profile due to pod posting with the ratio α/β, which represents the ratio of activity on control posts after the user began interacting with pods to that on control posts before pod interaction began. $\alpha/\beta > 1$ indicates increased interaction on control posts after pod posting began, $\alpha/\beta < 1$ indicates a decrease, and $\alpha/\beta = 1$ indicates no change. When computing this ratio, we drop all users for whom either α or β is computed over fewer than 5 posts. This results in us dropping 283 of our original 800 users.

We also sought to understand whether increased pod utilization leads to increased organic interaction, a relationship that would suggest that pods are effective at increasing a user's popularity. To examine this, we ran a linear regression with the increased interaction after first pod post (α/β) as the outcome variable. This regression was designed to test the relationship between this outcome variable and the proportion of each user's Instagram posts in the pod-interacting period that were posted in a pod. We hypothesized that many variables could confound this relationship, and thus controlled for them in the regression. These variables included inherent pod properties that could impact efficacy, including the number of members of the user's most-used pod and that pod's maximum post rate over its lifespan. To account for the possibility that the number of posts a user made in the past might increase their "importance" to the content curation algorithm, we included two additional variables: the proportion of each user's total posts that occurred in the baseline period, and the user's total number of posts. Finally, we controlled for calendar time, because we suspected that as the Instagram platform has grown, its increased membership could mechanically increase the level of overall interaction occurring. To control for this, we implemented as variables the calendar time of the user's first and last Instagram posts. After normalizing each variable through scaling by its mean and variance, we ran a linear regression on these variables with $Y = \alpha/\beta$ as the dependent variable. The resulting coefficients are in Section 7.2.

7.2 Results

As shown in Figure 7, we find that 70% of users experienced a 2-fold or greater increase in interaction level on control posts after they began posting in pods, and on average, these users saw a five-fold increase in comments. This increased interaction is likely to be organic (not from reciprocal interaction), because these control posts were, according to the classifier, not posted in pods. Posts identified as control posts with less certainty were given less weight in the computation of this ratio.

As shown in Table 4, only two regression coefficients are significant at a 5% level. Of these two coefficients, only one has a magnitude of practical significance: the proportion of IG posts in the pod-interacting period that were posted in pods. The coefficient of this variable, 11.79, indicates that, all else held constant, if a user who had never posted in pods began posting 50% of their posts in

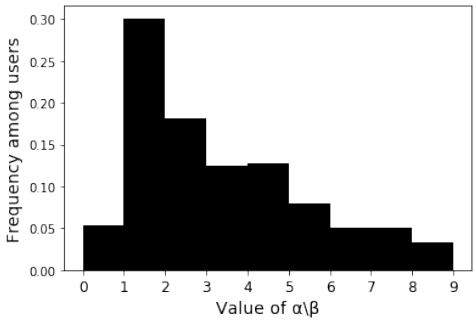


Figure 7: Value of α/β, signifying the ratio of the average organic activity on a user's posts after versus before the first time they utilized a pod. Frequency over all users (n = 577).

pods, they would see a more than 5-fold increase in organic interaction with the posts that they did not post in pods. This strong positive correlation and its practical interpretation suggests that posting in pods is an effective way of increasing overall organic interaction with the user's profile.

Variable	Regr. coeff. ± 95% CI
Prop. of IG posts in pod-interaction period posted in a pod*	$11.79^{*} \pm 2.82$
Prop. of total posts in baseline period	0.26 ± 9.19
Pod's max. post rate over lifespan	$(5.04 \pm 6.28) \times 10^{-4}$
User's last post time	$(3.89^{*} \pm 2.71) \times 10^{-8}$
User's first post time	$(-6.50 \pm 7.86) \times 10^{-9}$
Number of pod members	$(-1.41 \pm 1.41) \times 10^{-4}$
Number of posts	$(-9.47 \pm 12.8) \times 10^{-4}$

Table 4: Linear regression coefficients. * indicates statistical significance at 5% level.

8 DISCUSSION

8.1 Implications

We explored the ecosystem of Instagram pods hosted on Telegram, which appears to be the most popular platform for hosting Instagram pods. The ease with which we were able to find these pods via google search, the low barrier to joining them, and their consistency in rules and structure all increase the potential for these groups to continue to be rapidly adopted. Already, there is evidence of recently increasing adoption of this strategy: the pods we discovered have emerged at an accelerating pace over the last two years.

We find that these pods are not only easy to join, but are actually effective at increasing the organic interaction users receive. Strikingly, we find that if a user who had not been utilizing pods began advertising half their Instagram posts in pods, they would see more than a 5-fold increase in organic interaction with their Instagram content. The efficacy of these pods and their ease of use threaten the integrity of OSNs, as they present a method by which users can use these platforms to artificially and rapidly gain influence, potentially for nefarious purposes.

To address this emerging threat, we demonstrate that it is possible to detect posts that are likely to be getting interactions via pod usage. Importantly, we show that this detection is possible *even when engagement is high in both groups*, because there are attributes of pod interactions beyond engagement level, such as style of comments and interaction timing, that can be used to make these more granular discernments. Our results suggest that a pod detection tool might use engagement-level features as a screening heuristic before deploying a classifier trained to distinguish between high-organic-engagement users and pod users. In practice, OSN platforms could use this basic paradigm, together with much richer data and additional information such as topology of interaction networks, to build more accurate pod detection tools. These tools could ultimately be deployed to detect and penalizing pod users, or alternatively as a part of Instagram's content curation algorithm itself, where it could help the algorithm account for this type of interaction by down-voting posts that benefit from it.

8.2 Limitations

Because we studied exclusively English pods, our results likely more closely reflect English communities. Due to the substantial computational time required to construct these datasets without back-end access to the OSNs we studied, we were limited in our investigation by the scope of our data. Our pod interaction detection classifier is limited in the sense that it relies heavily on the content of post comments. While it does learn topics from this content that are consistent with what one would expect, the reliance of the classifier on this content makes it less useful in the detection of interaction with the estimated 50% of pods that do not require comment interactions. Finally, while we show that pod-interacting posts can be distinguished from posts with high *organic* engagement, users with high organic engagement such as artists and influencers still remain at heightened risk of receiving false-positive predictions. Some of these false-positives can be mitigated by whitelisting such users, for example, using the "verified" badge.

9 CONCLUSIONS

In this paper, we show that Telegram-hosted Instagram pods have become increasingly prevalent, are utilized by many OSN users, for many reasons, and are likely to continue to be popular. We also show that pods are effective at increasing the popularity of user content, which affirms the threat they pose to the integrity, security, and resilience of OSNs against politically-motivated propaganda and other implications of artificially-garnered influence. To address this growing threat, we developed a supervised learning tool to detect posts with high likelihood of having gained popularity through pod engagement. We propose that this tool could be deployed as part of content curation algorithms.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their helpful comments and feedback, and Cynthia Gill and Jaime Richards for their contributions in the early stages of this project. Our work was supported by the National Science Foundation under grants 1931005 and 1814816.

REFERENCES

[1] 2012. News Feed FYI. https://newsroom.fb.com/news/category/news-feed-fyi/. [Online; Accessed 10-May-2019].

[2] Ling Huang Sadia Afroz Anthony D. Joseph J. D. Tygar Alex Kantchelian, Justin Ma. 2012. Robust Detection of Comment Spam Using Entropy Rate. (2012).

[3] Alex Beutel, Wanhong Xu, Venkatesan Guruswami, Christopher Palow, and Christos Faloutsos. 2013. Copycatch: stopping group attacks by spotting lockstep behavior in social networks. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 119–130.

[4] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[5] Emma Brown. 2018. Do Instagram Pods Work? The Truth Behind InstagramâĂŹs Latest Engagement Hack. https://blog.hootsuite.com/instagram-pods/. [Online; Accessed 10-May-2019].

[6] Qiang Cao, Xiaowei Yang, Jieqi Yu, and Christopher Palow. 2014. Uncovering Large Groups of Active Malicious Accounts in Online Social Networks. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14)*. ACM, New York, NY, USA, 477–488. https://doi.org/10.1145/2660267.2660269

[7] The Scipy community. 2016. Scipy - Linear Sum Assignment. https://docs.scipy.org/doc/scipy-0.18.1/reference/generated/scipy.optimize.linear_sum_assignment.html. [Online; Accessed 10-May-2019].

[8] Oliver Darcy. 2019. How Twitter's algorithm is amplifying extreme political rhetoric. https://edition.cnn.com/2019/03/22/tech/twitter-algorithm-political-rhetoric/index.html. [Online; Accessed 10-May-2019].

[9] Emiliano De Cristofaro, Arik Friedman, Guillaume Jourjon, Mohamed Ali Kaafar, and M. Zubair Shafiq. 2014. Paying for Likes?: Understanding Facebook Like Fraud Using Honeypots. In *Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14)*. ACM, New York, NY, USA, 129–136. https://doi.org/10.1145/2663716.2663729

[10] Louis F DeKoven, Trevor Pottinger, Stefan Savage, Geoffrey M Voelker, and Nektarios Leontiadis. 2018. Following Their Footsteps: Characterizing Account Automation Abuse and Defenses. In *Proceedings of the Internet Measurement Conference 2018*. ACM, 43–55.

[11] Shehroze Farooqi, Fareed Zaffar, Nektarios Leontiadis, and Zubair Shafiq. 2017. Measuring and mitigating oauth access token abuse by collusion networks. In *Proceedings of the 2017 Internet Measurement Conference*. ACM, 355–368.

[12] Jin Yea Jang, Kyungsik Han, and Dongwon Lee. 2015. No reciprocity in liking photos: analyzing like activities in Instagram. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. ACM, 273–282.

[13] Alex Kantrowitz. 2018. Facebook Removes 10 Instagram Algorithm-Gaming Groups With Hundreds Of Thousands Of Members. https://www.buzzfeednews.com/article/alexkantrowitz/facebook-removes-ten-instagram-algorithm-gaming-groups-with. [Online; Accessed 10-May-2019].

[14] David Andrzejewski David Buttler Keith Stevens, Philip Kegelmeyer. 2012. Exploring topic coherence over many models and many topics. (2012).

[15] Daniel D Lee and H Sebastian Seung. 2001. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*. 556–562.

[16] Yixuan Li, Oscar Martinez, Xing Chen, Yi Li, and John E Hopcroft. 2016. In a world that counts: Clustering and detecting fake social engagement at scale. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 111–120.

[17] Scott M Lundberg, Gabriel G Erion, and Su-In Lee. 2018. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888* (2018).

[18] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[19] Christopher Manning, Prabhakar Raghavan, and Hinrich Schütze. 2010. Introduction to information retrieval. *Natural Language Engineering* 16, 1 (2010), 100–103.

[20] Arunesh Mathur, Arvind Narayanan, and Marshini Chetty. 2018. Endorsements on Social Media: An Empirical Study of Affiliate Marketing Disclosures on YouTube and Pinterest. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 119 (Nov. 2018), 26 pages. https://doi.org/10.1145/3274388

[21] Alex Pasternack. 2018. ItâĂŹs not over: RussiaâĂŹs divisive Instagram memes are still racking up likes. https://www.fastcompany.com/90283167/russia-instagram-war-facebook-memes. [Online; Accessed 10-May-2019].

[22] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 1135–1144.

[23] Nakatani Shuyo. 2010. Language Detection Library for Java. http://code.google.com/p/language-detection/

[24] Dustin Stout. 2019. Social Media Statistics 2019: Top Networks By the Numbers . https://dustinstout.com/social-media-statistics/. [Online; Accessed 10-May-2019].

[25] Gianluca Stringhini, Gang Wang, Manuel Egele, Christopher Kruegel, Giovanni Vigna, Haitao Zheng, and Ben Y. Zhao. 2013. Follow the Green: Growth and Dynamics in Twitter Follower Markets. In *Proceedings of the 2013 Conference on Internet Measurement Conference (IMC '13)*. ACM, New York, NY, USA, 163–176. https://doi.org/10.1145/2504730.2504731

[26] Alex Tooby. 2018. The Truth About Instagram Pods âĂŞ Do They Really Work To Increase Your Engagement? https://alextooby.com/instagram-pods/. [Online; Accessed 10-May-2019].

[27] Bimal Viswanath, M. Ahmad Bashir, Mark Crovella, Saikat Guha, Krishna P. Gummadi, Balachander Krishnamurthy, and Alan Mislove. 2014. Towards Detecting Anomalous User Behavior in Online Social Networks. In *23rd USENIX Security Symposium (USENIX Security 14)*. USENIX Association, San Diego, CA, 223–238. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/viswanath