Supporting a Storm: The Impact of Community on #GamerGate's Lifespan

J. David Smith, *Student Member, IEEE* and My T. Thai, *Senior Member, IEEE*

Abstract—Over the past half-decade, we have seen repeated examples of "firestorms" on social media. These often-negative events focus enormous bursts of attention on particular individuals or topics. While they often die out within days, in some cases they persist far longer. In this work, we study one such storm, #GamerGate, with the goal of identifying key differences between it and other, smaller events. We find that this storm has a distinct pattern of growth that differentiates it from previous events: rather than exploding quickly and burning out, it grows more slowly and develops an underlying community structure. Further, we find that participation in this community structure is a key indicator of whether a user will continue using the hashtag over the following year. Lastly, we find that users that become active later in the storm's lifetime contribute significantly to this community structure, though they appear to interact less overall with "older" users. Our results are the first to suggest that the formation of network structures play a significant and fundamental role in determining the lifecycle of these large-scale firestorms.

1 Introduction

As online social networks have grown, so to has grown the scale of online *firestorms*: social media events typically consisting of an explosion of negative activity directed at a particular user or topic. Most of the time, these storms are but a flash in the pan–lasting a few days in the majority of cases [1]. There are, however, exceptions to this rule. In this work, we study one such exception: #GamerGate.

A storm of harassment born of a smear campaign against game developer Zoë Quinn [2], [3], #GamerGate is the largest and longest-lasting firestorm ever observed. The instigating blog post made its way to 4chan on 16 August, 2014 [4], [5]. It quickly morphed into a conspiracy theory, which spread to Twitter and Reddit. It was not until August 28th that actor Adam Baldwin first coined #GamerGate in a tweet linking two videos on *The Quinnspiracy*, giving the burgeoning campaign a name divorced from its roots.

Nearly four years later, the hashtag still sees limited activity: 144 tweets using the hashtag appeared between the 20th and 21st of May, 2018. While the furor has undoubtedly died down, we are interested in understanding what led it to last so long in the first place. Despite its notoriety, there has been little study of the large-scale patterns present in data on #GamerGate. Chatzakou *et al.* [6] present the first study of #GamerGate users viewed through their Twitter posts. They find that ca. 2016 these users tend to be more active and have more followers and "friends" (Twitter's official term for users one follows) than a random sample of users. We build on this by going back to the beginning: August 2014.

In particular: we take advantage of archival data to study the structure of this event during its first year. Through this, we develop an understanding of *structural features* that distinguish it from events that have been studied previously (e.g. [1]). Our focus on structural features also distinguishes this from the prior work on #GamerGate, which has a heavy emphasis on textual features due to the harassment attached to the event [6]. In summary, our main contributions are:

- We show that community structure quickly developed among #GamerGate users, and that community participation is a key predictor of long-term activity.
- We find that, unlike events studied previously, #GamerGate does not feature explosive growth. Rather, it grows slowly and steadily—and in doing so alters our understanding of social network dynamics.
- We demonstrate the *conductance* of a graph partition is robust under the kind of sub-sampling present in the Twitter data, even on a 1% sample.

Roadmap. Using historical data available on archive.org [7], we study the participants of #GamerGate in 2014. We begin in Sec. 2 by showing both that #GamerGate is different from storms studied in prior work. Based on an intuition from contemporary reporting on #GamerGate, we next study the presence of community structure among #GamerGate users (Sec. 3). We find that #GamerGate users interact with one another at a significantly higher rate than random users or users involved in other storms. We follow this with a comparison of the user-base in 2014 to August–September 2015 (Sec. 4), finding that interacting with other users was highly correlated with continued activity. This is followed by a brief survey of related works (Sec. 5) and a discussion of the broader implications of this work (Sec. 7).

1.1 Data Collection

We draw data from the archival data hosted on archive.org [7]. This data is a complete reproduction of the 1% "Spritzer" Twitter sample. The Spritzer sample has been shown to be an *unbiased* representation of the Twitter population [8]. We collect data from the entire period of #GamerGate activity in 2014 (August–December), as well as a set from Aug–Sep 2015. For comparison, we use the tweets of random users active in the Aug–Sep 2014 time period.

We additionally collect data for 16 of the 20 storms studied in [1] from the same 1% sample to use in our

J. David Smith and My T. Thai are with the Computer & Information Science & Engineering Department, University of Florida, Gainesville, FL, 32601 USA. E-mail: {jdsmith,mythai}@cise.ufl.edu.

TABLE 1: Storms from Lamba *et al.* [1] to which we compare. All data is from the 1% stream to ensure accurate comparisons. "# Rel. Tweets" is the number that use or mention the hashtag or user in question, and "Observed Dur." is the amount of time between the first and last such tweet. Note that only 6 weeks of post-event data are collected for other storms.

Storm	Event Date	First Tweet	Last Tweet	Observed Dur.	# Tweets	# Rel. Tweets	# Users
#whyimvotingukip	20 May 2014	6 May 2014	30 Jun 2014	6 weeks	44,590	4,263	4,125
@TheOnion	24 Feb 2013	10 Feb 2013	6 Apr 2013	6 weeks	40,870	3,557	4,453
@David_Cameron	5 Mar 2014	19 Feb 2014	15 Apr 2014	6 weeks	57,830	2,898	3,468
#myNYPD	22 Apr 2014	1 Apr 2014	1 Jun 2014	6 weeks	50,039	1,850	1,749
#muslimrage	17 Sep 2012	3 Sep 2012	28 Oct 2012	6 weeks	23,817	1,543	1,466
@KLM	29 Jun 2014	15 Jun 2014	9 Aug 2014	6 weeks	19,327	1,457	1,676
#CancelColbert	27 Mar 2014	1 Mar 2014	1 May 2014	5 weeks	38,512	1,453	1,378
#AskThicke	30 Jun 2014	16 Jun 2014	10 Aug 2014	6 weeks	24,824	1,194	1,167
@GaelGarciaB	29 Jun 2014	15 Jun 2014	9 Aug 2014	6 weeks	10,626	1,016	1,107
@celebboutique	20 Jul 2012	6 Jul 2012	30 Aug 2012	6 weeks	11,357	718	749
#AskJPM	6 Nov 2013	23 Oct 2013	17 Dec 2013	5 weeks	7,053	491	461
#VogueArticles	10 Sep 2014	27 Aug 2014	21 Oct 2014	5 weeks	5,605	301	290
@SpaghettiOs	6 Dec 2013	22 Nov 2013	1 Jan 2014	4 weeks	3,522	297	296
@fafsa	25 Jun 2014	11 Jun 2014	5 Aug 2014	6 weeks	4,637	271	290
#McDStories	18 Jan 2012	4 Jan 2012	28 Feb 2012	6 weeks	2,764	253	252
#AskBG	17 Oct 2013	3 Oct 2013	27 Nov 2013	1 weeks	2,793	218	213
#GamerGate	28 Aug 2014	1 Aug 2014	1 Jan 2015	22 weeks	565,826	69,676	14,984
#GamerGate (2015)	-	1 Aug 2015	1 Oct 2015	8 weeks	61,378	7,494	2,615
Random	-	1 Aug 2014	1 Oct 2014	-	277,834	-	7,112

comparisons in Sections 2 and 3. Three are excluded because data is not available from the time period in which they occur (#qantas, #QantasLuxury and #NotIntendedToBeAFactualStatement, all from 2011). A fourth (@UKinUSA) is excluded because too few tweets appear in the 1% sample. We collect data from two weeks before to six weeks after each event. When comparing #GamerGate to other storms, we use a similar interval: $[-2, +6]$ weeks from 28 Aug 2014.

A two-pass process is used to extract data from the stream. First, we identify the set of *related users*–i.e. those that used the hashtag or mentioned the listed user. We then collect every tweet by these users in the specified observation period. Table 1 shows detailed information about the collected data. Due to the large amount of activity on the #GamerGate hashtag during the 2014 time period, we elected not to use snowball sampling to get related hashtags.

Taking a cue from [6], we follow a simple scheme for spam removal: if a user uses more than two hashtags per tweet on average, or has an average normalized inter-tweet Levenshtein distance of less than 0.7 we label them a *spam user* and remove them. These values were selected based on manual inspection of the filtered accounts, and end up removing roughly 2% of users from each dataset.

1.2 Hypothesis Testing Methods & Coefficients

In this work, we rely on statistical hypothesis testing to quantify the accuracy of our claims. Two tests in particular are used throughout our work: the Mann-Whitney U (MWU) test, a nonparametric rank-sum test useful both for questions of ordering and homogeneity of integral and real variables [9]; and the χ^2 test, a nonparametric test of independence useful for categorical variables [10]. The null hypotheses for these tests are that neither the distribution of X nor of Y is stochastically greater than the other; and that the variables X and Y are independent, respectively. In this work, we generally leave the null hypotheses implied as they are dependent only on the test use and the groups X and Y being tested. We use the implementations of these available in the SciPy package [11], and in particular use the 2×2 contingency-table-based χ^2 test provided.

We note that the large sample sizes we use mean that these tests will mark even very small differences significant. Thus, we also present various *effect sizes* to quantify how large the impact is. For the χ^2 test, we use the standard Pearson's correlation coefficient ϕ, which varies from $[-1, 1]$ and indicates how well correlated the test statistic is with group membership; $\phi = 0$ indicates no correlation [10]. For the MWU test, we use two values. First, the common language effect size (CLES) c, which reports the fraction of all pairs that explicitly support the hypothesis. Second, the rank-biserial correlation r, which is analagous to ϕ [12].

We select a family-wise p-value (i.e. the probability of rejecting *any* null hypothesis tested in this work when it should be accepted) of $\alpha^* = 0.01$. After Bonferroni correction [10] for the $m = 202$ hypotheses tested in the development of this work, we have a corrected significance level of $\alpha = \alpha^*/m = 4.95 \times 10^{-5}$. Tests whose p-value is below α are marked in **bold**. Outside of tables, we write $\mathbf{p \approx 0}$ if the p-value is more than 100 times smaller than α.

1.3 Ethical Considerations

As noted previously, we make extensive use of social network information in this work. This information was posted publicly to Twitter, and thus is not sensitive or private information. According to the guidelines posted by the IRB at our institution, this study does not require additional review. Further, nearly all of our analysis is done and presented in aggregate form, where individual users cannot be identified. This further mitigates the risk of individual harm.

Later in our work, we perform some analysis on *notable users*. While any reader could reconstruct the list of notable

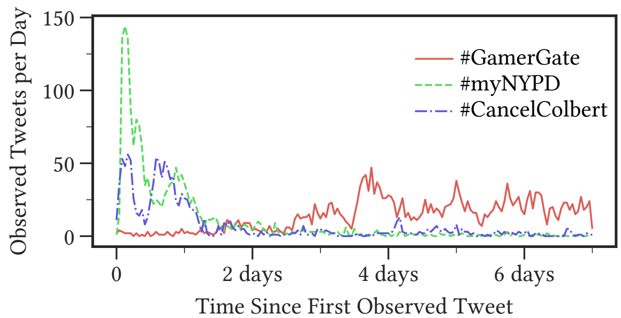


Fig. 1: Number of observed tweets per day for each of #GamerGate, #CancelColbert and #myNYPD.

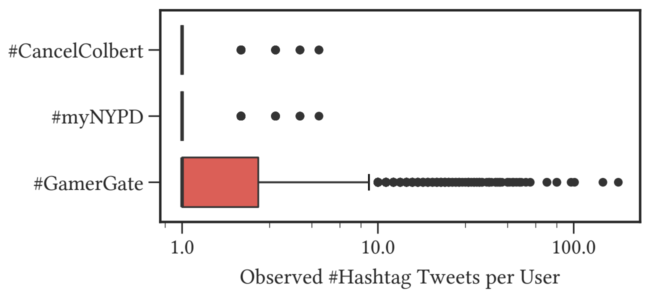


Fig. 2: Distribution of observed #hashtag-using tweets per user for #GamerGate, #CancelColbert and #myNYPD over the six weeks since the first hashtag usage. The box covers the inter-quartile range: from the 25th to 75th percentiles.

users we use based on our description, we do not list the exact groups. This is due to the level of harassment associated with #GamerGate; we do not wish to hand out a list of targets to those who may have participated in harassment and may later find this work. This concern is motivated in part by the weaponization of prior academic work by #GamerGate members [13]. We do name certain key individuals, such as Adam Baldwin and Zoë Quinn, as these individuals are already inextricably linked to #GamerGate.

2 Is #GamerGate Different?

To ground our analysis, we first seek to confirm that #GamerGate has substantial differences from what has been previously observed. We accomplish this by comparing to a subset of the storms studied by Lamba *et al.* [1]. Note that since we do not have access to the same 10% stream they used, we collect the data anew from the 1% stream (c.f. Sec. 1.1) and reproduce their results. Thus, all comparisons between storms are made with data taken from the 1% stream. When drawing figures, we primarily show the storms #myNYPD and #CancelColbert as these are the events highlighted in [1]. Following this, we compare to the results of Chatzakou's study of #GamerGate users in 2016 [6]. In particular, we pose the following questions:

Research Question 1. Is #GamerGate significantly different than the storms studied by Lamba *et al.* [1]?

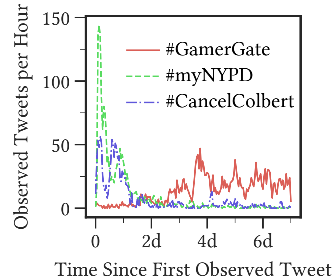


Fig. 3: Observed tweets per hour in the first week.

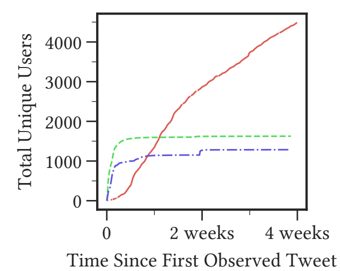


Fig. 4: Unique active users over the first 4 weeks.

Research Question 2. Did #GamerGate change significantly between its start in August 2014 and the observations of Chatzakou *et al.* [6] in 2016?

We first qualitatively show that #GamerGate was a much *larger* storm. It is immediately clear from Fig. 1 that the rate of activity outside of spikes is much higher for #GamerGate than for either of #myNYPD or #CancelColbert. Figure 2 indicates that this difference is caused by more than a small number of users in the tail of the distribution: even in the head of the distribution, #GamerGate users tend to be more active. Thus, we can conclude that #GamerGate has higher activity overall due to typical users being more active.

These qualitative observations on #GamerGate's size hold up under quantitative analysis. For the sake of readability, we only include comparisons to #myNYPD; unless otherwise specified, these results generalize to all storms in question. By the MWU test, we find that users that tweeted about #GamerGate did so a moderate amount more than those who tweeted about #myNYPD ($\boldsymbol{p \approx 0}$, $c = 0.38$, $r = 0.36$). Similarly, #GamerGate saw significantly more daily active users than other storms ($\boldsymbol{p \approx 0}$, $c = 0.95$, $r = 0.90$) and activity (in tweets-per-day) was higher overall ($\boldsymbol{p \approx 0}$, $c = 0.95$, $r = 0.91$). However, merely examining the magnitude of the storm misses subtler differences.

Looking more closely at the initial period of these storms, it becomes apparent that #GamerGate differs not only in activity level but also in the *shape* of its growth. Figure 3 shows that, contrary to the explosive expectations of typical firestorms, #GamerGate began with a slow burn. The first two-and-a-half days after the first mention have almost no activity, followed by a jump mid-way through day 3. Note that this spike rivals the peak activity on #CancelColbert. This is underscored by Figure 4: while #myNYPD and #CancelColbert rapidly gain users in the first day and then level out, #GamerGate gains users at a lower but steady rate. Even the second spike in #CancelColbert activity–visible just before the two week mark–does little to disrupt this pattern. The day-to-day change in unique users over the observation period is not significantly higher for #GamerGate than #myNYPD ($p = 0.59$, $c = 0.48$, $r = -0.03$) or any other storm. Upon investigation, this appears to be due to dips in the number of daily active users of #GamerGate after the initial growth shown in Fig. 4.

Based on this, we test whether the number of followers amongst #GamerGate users grew faster than the follower counts in other storms. Specifically: we calculate the difference between follower counts at the first and last tweet of each user that tweeted at least twice in the eight-week observation period. This approach is taken both because (1) the true follower accounts are contained in each tweet object, and (2) a simpler time-binning approach is infeasible due to data sparsity (almost all bins are empty). Unlike previous tests, which are consistent across all storms, this one is highly variable. While it is significant for many storms (e.g. #whyimvotingukip ($\boldsymbol{p \approx 0}$, $c = 0.56$, $r = 0.13$)), it is not for all—including #myNYPD ($p = 1.00$, $c = 0.44$, $r = -0.10$). The effect sizes for this test are uniformly small, indicating that–despite significance–the dependence of follower count on storm participation is small. We observe similar results when considering the number of "friends" a user has ($p = 3.28 \times 10^{-2}$, $c = 0.51$, $r = 0.03$). These outcomes together lead to our first result:

Result 1. #GamerGate is significantly different from prior storms, with a slower start, continued growth over time, and higher activity overall.

Lastly, we briefly compare the network properties of #GamerGate users to random users. Where Chatzakou *et al.* [6] find that #GamerGate users in 2016 tended to have more friends and followers and to tweet more than a uniform random sample of users, we find that in August 2014 the opposite holds: #GamerGate users had slightly fewer followers ($\boldsymbol{p \approx 0}$, $c = 0.58$, $r = 0.16$) and tweets ($\boldsymbol{p = 1.89 \times 10^{-5}}$, $c = 0.52$, $r = 0.04$) than our random sample. They didn't have a significantly different number of friends ($p = 1.43 \times 10^{-4}$, $c = 0.52$, $r = 0.04$). There are a number of potential explanations for this. We will examine one such explanation—that low-activity users simply did not continue participating—in Sec. 4. Regardless of the cause, the effect is clear and leads to the following:

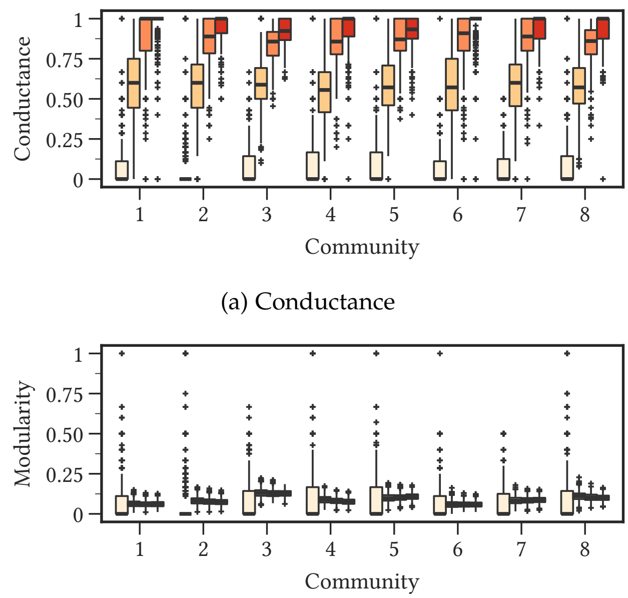
Result 2. #GamerGate users in 2014 had fewer followers and tweets than a random sample. Although the difference is small, this is the opposite of the relation seen in 2016 [6].

3 The Presence of Community

Having established that #GamerGate is different, we next explore one potential cause: community structure. Specifically, we hypothesize that a community formed among #GamerGate users early in the storm, which could help explain its longevity. This hypothesis is motivated by two observations. First, we see that discussion of the event in its historical context often includes discussion of certain key users, such as @Int_Aristocrat, Adam Baldwin, Zoë Quinn, and Anita Sarkeesian. It is natural to wonder whether communities formed around these notable pro- and anti-#GamerGate users, or whether they are merely proxies for discussion of the participants. Second, and unlike the other events we study, a sub-reddit (`r/KotakuInAction`) became a hot-bed of discussion related to the hashtag early in its lifetime. These observations lead us to several questions:

Research Question 3. Were one or more communities present amongst #GamerGate users? If so, does this distinguish it from other events?

If so, it leads to two follow-up questions:



(a) Conductance

(b) One-vs-Rest Modularity

Fig. 5: Distribution of conductance $\psi(S)$ and modularity Q as measured on each $G' \in \mathcal{G}$ for 8 arbitrary ground-truth synthetic communities (leftmost within each group) and those same communities with (left-to-right) 33%, 66% and 100% of members replaced by random users.

Research Question 4. Was this community structure newly-formed or pre-existing?

Research Question 5. Is this community structure related to certain "notable" users?

3.1 Measuring Community Structure on a 1% Sample

At a high level, there are two approaches we could take to test our hypotheses about the presence of community. The first is community detection, which takes as input the network G, and produces a set of likely communities S_i as output. Alternately, we could consider the problem as community *testing*, which takes as input the network G and a potential community S, then produces as output the "quality" of S. The testing problem can be seen as a sub-problem of detection: community detection inherently is about finding communities that are high-quality. In this work, we are ultimately interested only in measuring the quality of a single community defined by our hypotheses as a means of understanding the relationship between specific users. Thus, we take the community *testing* approach.

However, the relationship between detection and testing has led to the development of effective metrics for community quality. Fortunato provides a historical overview of the subject [14]. In this section, we will focus on two widely-used methods in particular: modularity and conductance.

Definition 1 (Modularity). Suppose each node $i \in V$ of a graph $G = (V, E)$ is assigned a label $C_i \in \mathcal{C}$. Then the modularity of this assignment is [14]:

$$Q = \frac{1}{2|E|} \sum_{i,j \in V} \left(A_{ij} - \frac{k_i k_j}{2|E|} \right) \delta(C_i, C_j)$$

where $m = |E|$, A is the adjacency matrix of G, and $\delta(C_i, C_j)$ is 1 iff $C_i = C_j$ and 0 otherwise.

Definition 2 (Conductance). The conductance of set of nodes $S \subset V, S \neq \emptyset$ on a positively-weighted graph G is

$$\psi(S) = \frac{w(\delta S)}{\min(\mathrm{vol}(S), \mathrm{vol}(V \setminus S))}$$

where $w(\delta S) = \sum_{u \in S, v \notin S} w_{uv}$ and $\mathrm{vol}(S) = \sum_{u \in S} \sum_{(u,v) \in E} w_{uv}$ is the total outgoing weight of S.

Modularity is of interest to us due to its widespread, successful use. Researchers over the past two decades have applied modularity to detect communities on a wide variety of networks [15]. However, modularity has a natural mismatch with our problem: it is framed as a measure for evaluating the quality of a system of communities, not an individual community like the potential #GamerGate community we wish to study. While it is easy enough to take a one-vs-rest approach to community labeling, it is unclear if modularity will remain effective under this modification.

Conductance, on the other hand, has seen frequent use in measuring the quality of real-world communities [16] despite being less useful for global community *detection* than modularity [15]. This is partly due to its definition in terms of a single community S—which happens to align with the testing problem we wish to solve.

Each of these metrics operates on a network, which in our case will be the Twitter mention network that is constructed as follows: On Twitter, a mention takes the form `@screen_name` in the body of a tweet. If the screen name belongs to a real user, Twitter parses it and stores this user's ID in the tweet's metadata; this metadata is present in the archive we use. We treat a user u mentioning a user v as a sign of unidirectional interaction and create an edge (u, v) with weight equal to the count of such mentions.

Before applying these metrics to our Twitter data, we first apply them to a synthetic graph (generated via [17]) containing 1000 nodes with power-law degree (with exponent $\tau_1 = 3$) and community-size (exponent $\tau_2 = 1.5$) distributions, along with labelled communities. To mimic the mention network, we weight each edge with a "mention count" $w_e \in [1, 100]$ according to a power-law distribution.[1]

Our 1% sub-samples $G' \sim G$ are then constructed by sampling an *observed* edge weight for each edge e from a binomial distribution $w'_e \sim B(w_e, 0.01)$; any edges with $w'_e = 0$ are then removed. Finally, nodes with no in- or out-bound edges are removed. We then sample 1000 such sets, denoted $\mathcal{G} = \{G'\}$. These graphs preserve, on average, 25% of all nodes in V and 3.6% of all edges in E. Thus despite working with a 1% sample, we observe a substantially larger fraction of the synthetic network G.

A good metric will produce substantially different values the synthetic communities and random sets of nodes, even under sub-sampling. Fig. 5 shows that while conductance accomplishes this, the "one-vs-rest" approach to modularity does not.[2] Fig. 5a shows a notable shift as we move from ground-truth synthetic communities (leftmost within each column) to completely random sets (rightmost). On the other hand, while the distribution of modularity clearly indicates that *sometimes* modularity identifies individual community structure under sub-sampling, on average it fails to separate it from a random set.

Based on this, we will use conductance as our measure of community structure. We remark that conductance on a directed network counts the rate of out-mentions provided $\mathrm{vol}(S) \leq \mathrm{vol}(V \setminus S)$. We can get an equivalent characterization by looking at the rate of in-mentions (which is $1-\psi(S)$), and will occasionally speak in these terms.

TABLE 2: Comparing the mention behavior of storm participants to a set of random users via the χ^2 test, with the null hypothesis being that mention behavior is independent of set membership. ψ shows the conductance of the storm.

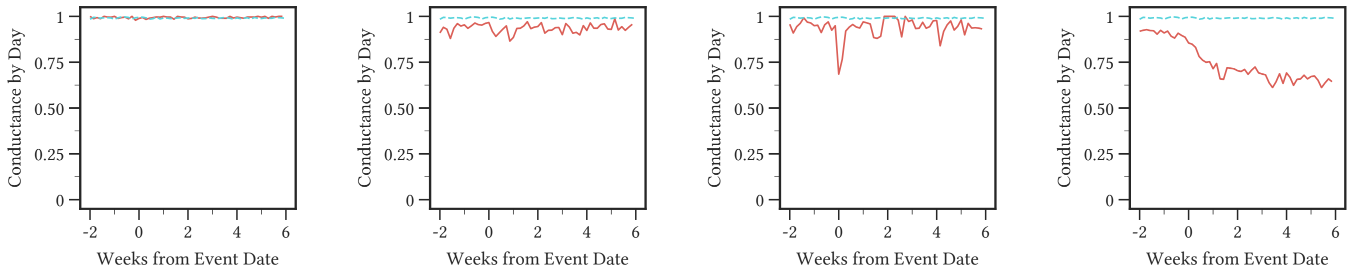
Storm	p	χ^2	ϕ	ψ
@KLM	**0.00**	2,450.46	0.19	0.86
#CancelColbert	$\mathbf{1.42 \times 10^{-275}}$	1,258.13	0.13	0.90
@celebboutique	$\mathbf{3.40 \times 10^{-230}}$	1,049.33	0.12	0.90
@David_Cameron	$\mathbf{3.60 \times 10^{-249}}$	1,136.63	0.12	0.93
#myNYPD	$\mathbf{2.11 \times 10^{-107}}$	484.63	0.08	0.94
#muslimrage	$\mathbf{1.15 \times 10^{-48}}$	214.94	0.06	0.95
#whyimvotingukip	$\mathbf{3.05 \times 10^{-29}}$	126.01	0.04	0.96
#AskThicke	$\mathbf{1.60 \times 10^{-13}}$	54.44	0.03	0.96
@fafsa	$\mathbf{2.43 \times 10^{-7}}$	26.66	0.02	0.96
@TheOnion	$\mathbf{2.45 \times 10^{-7}}$	26.64	0.02	0.97
#AskJPM	1.13×10^{-3}	10.60	0.01	0.97
#VogueArticles	1.27×10^{-2}	6.21	0.01	0.97
#AskBG	2.19×10^{-1}	1.51	0.01	0.97
#McDStories	9.18×10^{-1}	0.01	0.00	0.98
@SpaghettiOs	2.61×10^{-1}	1.26	−0.00	0.98
@GaelGarciaB	2.59×10^{-4}	13.35	−0.01	0.99
#GamerGate	**0.00**	20,934.74	0.40	0.67

3.2 Community Among the Storms

We now return to the question of whether the set of users that used the #GamerGate hashtag contains community structure. Figure 6 shows how the conductance changes over time on each of several storms. To calculate the conductance of a storm, we define the set S as the set of users that tweeted the relevant hashtag or username at least once in the observation period. We assume that $\mathrm{vol}(S) < \mathrm{vol}(V \setminus S)$, where V is the entire (unobserved) Twitter network and calculate $\psi(S)$ directly. Note that by Def. 2, $w(\delta S)$ only includes edges *leaving* S, and we therefore do not need to scan for all edges entering S from $V \setminus S$. This allows us to directly calculate $\psi(S) = w(\delta S)/\mathrm{vol}(S)$.

While most storms appear similar to what is shown in Figures 6a or 6b, a large dip in conductance centered on the date of the firestorm as seen in Fig. 6c. Contrast this with the behavior shown for #GamerGate in Fig. 6d, which reaches nearly the level of the dip for @KLM but *sustains it for weeks*. A reference line corresponding to a random sample of 10,000 users that tweeted between August and September 2014 is shown in each plot. Seven weeks after the first use of the #GamerGate hashtag, the conductance of this set has dropped to 0.55. We remark that this is well

1. This distribution was chosen because it produced a similar count distribution to what we see in our #GamerGate data after sub-sampling.
2. We have preliminary results indicating that if every community is labeled, modularity is robust to the sub-sampling. However, labeling every community on Twitter is beyond the scope of this work.



(a) @TheOnion (b) #CancelColbert (c) @KLM (d) #GamerGate

Fig. 6: Conductance evaluated on 1-day bins. The value on a random sample during Aug–Sep 2014 is shown as the dashed line. Most storms in our data look similar to (a) @TheOnion or (b) #CancelColbert.

TABLE 3: Comparing the mention behavior of #GamerGate participants to storm participants via the χ^2 test, with the null hypothesis that mention behavior is independent of set membership. ψ shows the conductance of the storm.

Storm	p	χ^2	ϕ	ψ
@KLM	$\mathbf{2.14 \times 10^{-236}}$	1,077.86	0.12	0.86
#CancelColbert	**0.00**	2,532.43	0.18	0.94
@celebboutique	$\mathbf{5.92 \times 10^{-308}}$	1,407.13	0.14	0.90
@David_Cameron	**0.00**	5,215.90	0.24	0.93
#myNYPD	**0.00**	3,285.67	0.20	0.94
#muslimrage	**0.00**	2,952.19	0.19	0.95
#whyimvotingukip	**0.00**	4,275.49	0.23	0.96
#AskThicke	**0.00**	2,302.70	0.17	0.96
@fafsa	$\mathbf{5.13 \times 10^{-119}}$	538.00	0.09	0.96
@TheOnion	**0.00**	6,025.09	0.26	0.97
#AskJPM	$\mathbf{4.46 \times 10^{-176}}$	800.38	0.11	0.97
#VogueArticles	$\mathbf{1.11 \times 10^{-131}}$	596.23	0.09	0.97
#AskBG	$\mathbf{4.21 \times 10^{-111}}$	501.62	0.08	0.97
#McDStories	$\mathbf{7.89 \times 10^{-109}}$	491.18	0.08	0.98
@SpaghettiOs	$\mathbf{1.54 \times 10^{-155}}$	705.92	0.10	0.99
@GaelGarciaB	$\mathbf{1.80 \times 10^{-242}}$	1,105.81	0.12	0.99

within the inter-quartile range for the "33% random" set on each community shown in Fig. 5.

While conductance gives us a clear connection to community detection, it does not tell us anything about how likely our observation is to be caused by random variation in the sampling. Observe that we can frame the question of "is there a community in set S?" as a hypothesis test with H_0 being "the distribution of in- and out-mentions is independent of membership in S." We therefore augment our measurement of conductance with a set of χ^2 tests that compare the mention behavior of a set S to the mention behavior of a group of random users. Significance on this test indicates that the set S has mention behavior that is unlikely to *appear* structured but actually be random. Table 2 shows the conductance and the results of this test on #GamerGate and each storm we consider.

While we find significance in many cases, effect sizes vary wildly. On @TheOnion, we have the correlation coefficient $\phi = 0.02$, a value that is exceptionally small—and which highlights the fact that our relatively large samples lead to significance even when the correlation with the variable is small. On the other hand, with #CancelColbert and

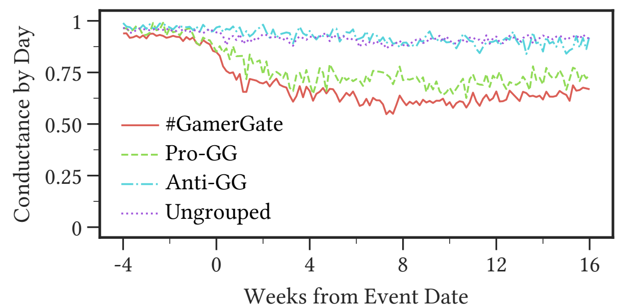


Fig. 7: Conductance of #GamerGate and sub-groups defined by retweet behavior over the 2014 observation period.

@KLM we have $\phi = 0.13$ and $\phi = 0.19$, respectively. These correlations remain small, but reflect the greater difference seen in Fig. 6. However, these are scarcely half the value seen for #GamerGate at $\phi = 0.40$–an effect that is more than ten times larger than what is seen for twelve of the storms, and more than twice as large as the remainder.

Further, applying this test to #GamerGate with each of the storms replacing the random set in turn, we find that each storm has significantly lower in-group communication than #GamerGate. Table 3 lists the results. The closest storm is #McDStories ($\boldsymbol{p \approx 0}$, $\phi = 0.08$), which is likely due to a difference in scale. Compare the 69,690 mentions by #GamerGate users to the 1,148 by #McDStories users within each eight-week window. Thus, we have:

Result 3. #GamerGate users had a significantly higher rate of mentioning other #GamerGate users than both a random set and other storms. Further, it has a low conductance once the event begins and this conductance is sustained over time. As a result, it seems highly likely that one or more communities exist amongst #GamerGate users and that this distinguishes it from other storms.

To answer RQ 4, we compare the weeks before the event to the weeks after to see if there is a significant difference. We expect this to be the case based on Fig. 6d, which shows a large gap between the conductance before (0.87) and after (0.54) day zero of #GamerGate. Unsurprisingly, adapting the above χ^2 test to compare the fifth and sixth weeks after the first use of the hashtag to the two weeks prior to its use reveals that the later period had a much higher rate

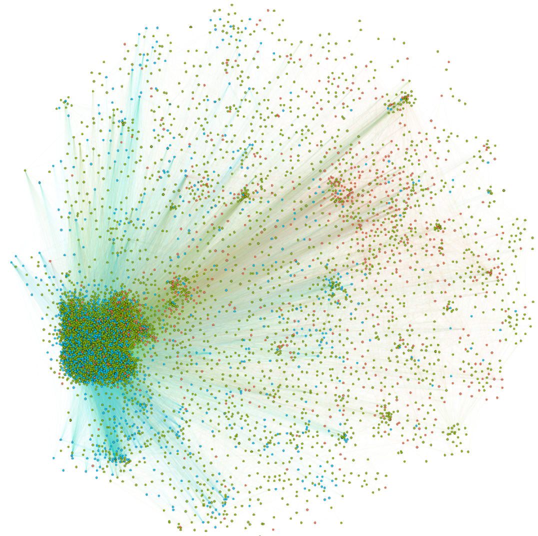


Fig. 8: The #GamerGate mention network. Isolated users are not shown. The central cluster contains primarily Pro-#GamerGate users (blue) and Ungrouped users (green). Anti-#GamerGate users (red) are spread throughout the network, primarily in the periphery. Edges are colored according to their source node. Generated using Gephi [18] with the OpenOrd layout algorithm [19].

of in-mentions ($\boldsymbol{p} \approx \mathbf{0}$, $\phi = 0.29$). Interestingly, we find that this gap in conductance remains when we compare to the last two weeks of December 2014 instead of weeks 5 & 6 ($\boldsymbol{p} \approx \mathbf{0}$, $\phi = 0.29$). Further, there is *not* a significant difference between weeks 5 & 6 and late December in the rate of in-group mentions ($p = 0.98$, $\phi = 0.00$)—as indicated by Fig. 7. Taken together, we reach the following conclusion:

Result 4. The community structure among #GamerGate users was newly formed, formed in the first few weeks, and then remained over the following months.

Next, we consider the question of whether key influencers are related to this community. To this end, we collect the top thirty most-mentioned by #GamerGate users in Aug–Dec 2014. After excluding @YouTube due to the way it is used (as an automated mention via YouTube's share function), we are left with a list of 29 users. Manual inspection of collected tweets reveals that most of these users can be placed into two groups: Pro-#GamerGate (e.g. Adam Baldwin, Milo Yiannapoulos) and Anti-#GamerGate (e.g. Zoë Quinn, Brianna Wu). Seven of the top thirty do not fall into either category. We construct sub-groups of the overall GG group based on their retweet behavior: if a user retweeted a notable Pro-GG user, they are placed in the Pro-GG group, etc. Note that this method allows groups to overlap. This method uses retweets, which broadcast another user's tweets on one's own timeline, as indicators of agreement. Prior work supports this interpretation, with Metaxas *et al.* [20] remarking (emphasis ours):

> Our findings show that retweeting indicates, not only interest in a message, but also trust in the message and the originator, and *agreement with the message contents.*

TABLE 4: Fraction of mentions from Source to Target Group. "Other" is the set of users not observed using #GamerGate.

Target / Source	Anti-GG	Pro-GG	Ungrouped	Other
Anti-GG	0.12	0.10	0.13	**0.65**
Pro-GG	0.06	**0.40**	0.21	0.32
Ungrouped	0.05	0.14	0.13	**0.69**

Figure 7 shows the conductance of the Pro- and Anti-GG groups, along with ungrouped #GamerGate users. The conductance of the overall #GamerGate is shown as a reference. It is immediately clear that only the Pro-GG group, which comprises 21.7% of all #GamerGate users, rivals the overall GG group in in-mention behavior. Indeed, by the same test used in Table 3 we find that this sub-group is more strongly correlated with in-group mentions than the #GamerGate group ($\boldsymbol{p} \approx \mathbf{0}$, $\phi = 0.46$), although the conductance of the overall group is slightly higher. On the other hand, the Anti-GG group, with a paltry-in-comparison 8.7% of #GamerGate users, is in-line with results on other storms–though still significantly different from random ($\boldsymbol{p} \approx \mathbf{0}$, $\phi = 0.15$). We find it interesting that the un-grouped set of users (i.e. those that neither retweeted a pro- or anti-GG user) maintains both a larger correlation coefficient on this test ($\boldsymbol{p} \approx \mathbf{0}$, $\phi = 0.22$) and a lower conductance.

We also find that the mention behavior of users is correlated with the notable users they retweet. Table 4 shows the fraction of mentions by each group's members that go to target group. Of the three sets (Pro, Anti, and Ungrouped) only Pro-#GamerGate users have a larger fraction of in-group mentions than mentions to other groups. Further, we find that there is a significant relationship between retweeting notable users and mentioning the "opposing" group: retweeting a Pro-#GamerGate notable is correlated with mentioning Anti-#GamerGate users more ($\boldsymbol{p} \approx \mathbf{0}$, $c = 0.38$, $r = 0.20$). This is not the case for the Anti-#GamerGate group ($p = 1.00$, $c = 0.25$, $r = -0.11$)–in fact, the opposite relation is significant ($\boldsymbol{p} \approx \mathbf{0}$, $c = 0.36$, $r = 0.11$). While the effect sizes are relatively small, the inversion of the relationship indicates some degree of *directionality* to communication amongst #GamerGate users. These results align with our understanding of the harassment associated with the event, though since we do not examine tweet content here we cannot say more.

Taken together, these results indicate that there are substantive differences between different sub-groups of users that tweeted with the hashtag. While the conductance we observe for this is not exceptionally low, given the range of values we see in Fig. 5 it is reasonable to believe that there is *some* community structure within the Pro-GG group that is lacking in the Anti-GG group.

Result 5. The Pro-#GamerGate sub-group has strong evidence of community structure: a high correlation with in-group mentions, low conductance. We do not see a similar result for Anti-#GamerGate users.

4 #GamerGate: One Year Later

As shown in Section 2, there are significant differences in the attributes of #GamerGate participants ca. 2014 and ca. 2016.

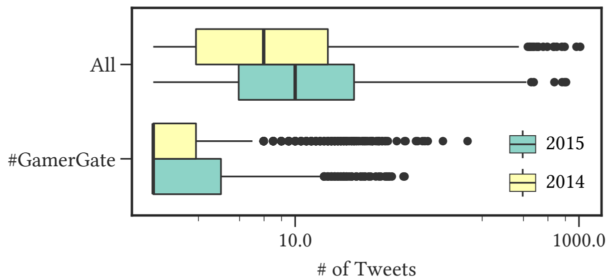
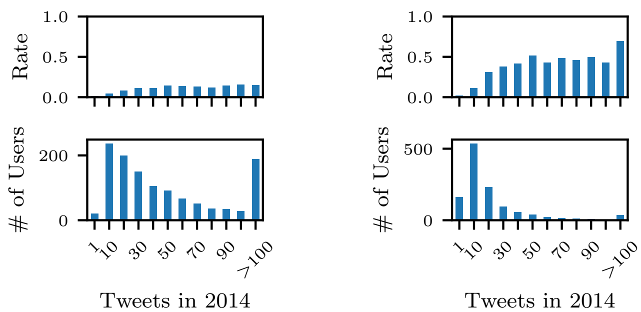


Fig. 9: Distribution of (a) how many times users of #GamerGate tweeted in 2014 (top) and 2015 (bottom), and (b) how many times they tweeted using the hashtag #GamerGate.



(a) All Tweets

(b) #GamerGate Tweets

Fig. 10: The rate and count of users remaining active on #GamerGate in 2015 binned by activity in Aug–Dec 2014.

This leads to a multitude of questions that all boil down to: what changed? In this section, we explore the factors that did (and did not) contribute to continued participation in #GamerGate. To this end, we compare activity in August–September 2015 to August–December 2014.

We begin by establishing the *presence* of continued activity. There are 2615 unique user accounts observed tweeting on #GamerGate in August–September 2015. While this is much lower than the 10,099 observed in the same time frame in 2014, it is still nontrivial. Figure 9 shows the distribution of tweet counts in August–September of both 2014 and 2015. We note the shift in the distribution of overall tweets: users active on the hashtag in 2015 tended to tweet more overall. This trend also appears in the number of tweets using the hashtag, but is less visible as the average continues to be about one hashtagged tweet. Our focus within this section is twofold: determining *which users remained active* and *which users became newly-active*. With this in mind, we pose the following questions:

Research Question 6. What factors are correlated with continuing activity on #GamerGate from 2014 to 2015?

Research Question 7. Is there a significant contingent of newly-active users in 2015?

4.1 Remaining Users

We begin seeking an answer to RQ 6 with a group highlighted by Fig. 9: **highly-active users**. Fig. 10 shows the rate at which users remained active (the "conversion rate") as a function of the number of tweets observed in 2014, and indicates that there is likely a relationship between activity on the hashtag and continued activity in 2015. While having ≥ 10 observed tweets overall is *negatively* correlated with continued activity ($\boldsymbol{p \approx 0}$, $\phi = -0.13$), having ≥ 10 observed tweets using #GamerGate is the opposite ($\boldsymbol{p \approx 0}$, $\phi = 0.18$). Indeed, users in the latter group are more than 2.5 times as likely to remain active their less-active peers.

Relatedly, we find that #GamerGate users in 2015 were more active than in 2014, but still not to the point observed by Chatzakou *et al.* [6]. They tweeted more ($\boldsymbol{p \approx 0}$, $c = 0.68$, $r = 0.37$), had more followers ($\boldsymbol{p \approx 0}$, $c = 0.71$, $r = 0.43$), and followed more users ($\boldsymbol{p \approx 0}$, $c = 0.70$, $r = 0.40$) than users in 2014. It seems likely that this is a result of less active users "dropping out" of the event. Repeating these comparisons with a random sample from Aug–Sep 2015, we find only that #GamerGate users followed more–though the effect is small ($\boldsymbol{p \approx 0}$, $c = 0.55$, $r = 0.10$). The tests on tweets ($p = 1.00$, $c = 0.42$, $r = -0.16$) and followers ($p = 1.00$, $c = 0.48$, $r = -0.03$) are both insignificant. It is thus clear that more changes occur between this period and the observations taken in 2016 by Chatzakou *et. al.*

Stepping beyond mere activity level, we find that users that **interacted frequently** with other #GamerGate users remained active significantly more often. In particular: users that remained active were mentioned by #GamerGate users 4.03 times as much on average ($\boldsymbol{p \approx 0}$, $c = 0.69$, $r = 0.48$) and were retweeted 4.58 times as much on average ($\boldsymbol{p \approx 0}$, $c = 0.63$, $r = 0.45$) as those that dropped out. Further, users that were mentioned at least once by a #GamerGate user remained active 1.62 times as frequently ($\boldsymbol{p \approx 0}$, $\phi = 0.18$) and those that were retweeted at least once remained active 1.99 times as frequently ($\boldsymbol{p \approx 0}$, $\phi = 0.21$).

TABLE 5: Correlation of continued activity with membership in one of two disjoint retweet (RT) groups in our 2014 sample. ϕ is the correlation coefficient. FR is the ratio of frequency of continued activity.

Group X RTs	Group Y RTs	p	ϕ	FR
Any Notable	No Notable	$\mathbf{7.61 \times 10^{-134}}$	0.20	3.63
Any Pro	No Pro	$\mathbf{1.36 \times 10^{-225}}$	0.26	4.95
Any Anti	No Anti	$\mathbf{2.35 \times 10^{-11}}$	−0.05	0.38

Returning to the idea of "notable users" introduced at the end of the previous section, we test for correlation between retweeting of notable users and continued activity on #GamerGate. Table 5 shows the results. Retweeting any notable is correlated with remaining active, with those who do so sticking around more than 3.5 times as often. Similarly, those who retweeted a Pro-GG user remain active about 5 times as often as those who do not. However, these results are tempered by the relatively low correlation coefficients: $\phi \approx 0.26$ indicates that there is likely a sizeable correlation, but that the magnitude of these frequency ratios may be overstated. We also note that few notable users remain active on the hashtag in our data: only one of eight notable Anti-GG users remained active, and only six of fourteen notable Pro-GG users tweeted with the hashtag in Aug–Sep 2015. These outcomes lead us to the following conclusion:

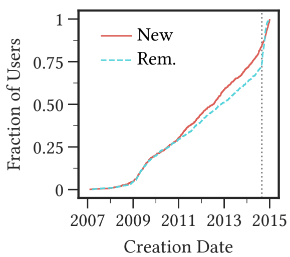


Fig. 11: CDF of Account Creations, 2007–2014.

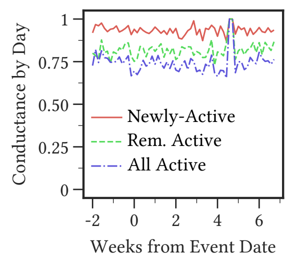


Fig. 12: Conductance of Newly-Active, Remaining, and all active users in 2015.

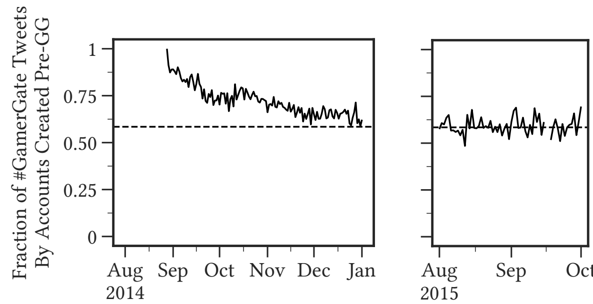


Fig. 13: Fraction of #GamerGate tweets by accounts created prior to it, converging toward the median in Aug–Sep 2015.

Result 6. We find several factors are correlated with continued activity on #GamerGate: level of activity on the hashtag, level of interaction with other #GamerGate users, and retweeting of Pro-GG notable users. Given these results, it seems likely that participation in the Pro-GG sub-"community" led to continued activity.

4.2 Newly-Active Users

Next, we turn our attention to the complementary group: users active in 2015 but not 2014, which we refer to as "newly-active users." Of the 2615 active users in 2015, 1396 (53.4%) fall into this category. We find that newly-active users interact less with #GamerGate users. They tend to retweet about 80% as often as remaining users ($\boldsymbol{p \approx 0}$, $c = 0.56$, $r = 0.18$), with a similar pattern shown in their mentions ($\boldsymbol{p \approx 0}$, $c = 0.52$, $r = 0.13$). Further, these users also retweeted #GamerGate users 44% as frequently as remaining users ($\boldsymbol{p \approx 0}$, $\phi = -0.18$). The average newly-active user retweeted #GamerGate users 2.5 times—less than half of the 5.7 times a remaining user retweeted the same group ($\boldsymbol{p \approx 0}$, $c = 0.60$, $r = 0.34$). A similar pattern holds for mentions: newly-active users mention GG users 3.7 times on average, significantly less than the 6.7 times remaining users do so ($\boldsymbol{p \approx 0}$, $c = 0.52$, $r = 0.25$). Coupled with the result that newly-active users retweet notables marginally less ($\boldsymbol{p \approx 0}$, $c = 0.36$, $r = 0.19$) and in particular Pro-GG notables marginally less ($\boldsymbol{p \approx 0}$, $c = 0.27$, $r = 0.13$), it seems likely that the newly-active group is simply less active overall. The lack of interaction with other #GamerGate users stands out, but at the same time newly-active users are not interacting with outside users more than #GamerGate users. As a result, we see from Fig. 12 that the conductance of the complete set of active users is *lower* than the set of just remaining users. It therefore seems likely that these users, while not of central importance, play a role in the community structure we see.

New Accounts. We notice that of the newly-active users in our data, 640 (45.8%; 24.5% of all active users) were also created after #GamerGate's first use. We refer to this group simply as *new accounts*. Figure 11 shows the distribution of creation dates among newly-active (including new accounts) and remaining users until the end of 2014. Note the sharp up-tick in account creations corresponding to #GamerGate's first use (28 August, 2014; marked by the dotted line). A further 35.74% of newly-active accounts were created in 2015. Figure 13 shows the proportion of #GamerGate tweets attributable to accounts that pre-date this storm. We see a drop down to 90% nearly immediately after the first tweet, and then a slow downward trend towards the median of 58.4% seen in our 2015 data.

These new accounts are–in general–not significantly different from other newly-active users. They retweet ($p = 0.30$, $c = 0.47$, $r = 0.02$) and mention ($p = 0.21$, $c = 0.46$, $r = 0.02$) a similar amount to other newly-active users, and while they do retweet GG users at a different rate than other newly-active users ($\boldsymbol{p \approx 0}$, $\phi = -0.03$) this translates to an insignificant difference in the average number of retweets of GG users ($p = 1.26 \times 10^{-3}$, $c = 0.43$, $r = 0.09$). Removing the new accounts from the overall GG group has a small but significant negative impact on the conductance ($\boldsymbol{p = 7.32 \times 10^{-7}}$, $\phi = -0.02$), indicating that these users are mentioned by other active users enough to outweigh the amount of out-group mentioning done by new accounts.

Lastly, we ask whether these new accounts were likely to be replacing older (e.g. banned) accounts. Comparing the screen names of newly-active users to users that were active in 2014 but not 2015 reveals that 123 newly-active accounts have at least one inactive account within a Levenshtein distance of 3 when capitalization is normalized. However, most of these are duds: the old name has little-to-nothing to do with the new name (consider, as an example @CasemanXP and @CalemAnnk, which are within a distance of 3 yet seem unrelated). On the other hand, there are some number of accounts that do appear to replace the old: @_trolljackoutis seems to replace @TrollJackOutis and @TrollJackOutis1, while @sleekit003 replaces @sleekit001. We find 17 using this heuristic, with another 11 that appear to replace with no changes aside from capitalization. While this test is obviously imperfect—there are no rules stating that the same users should always pick similar usernames—if the replacement of old accounts was a common application of new accounts, we'd expect to see much more than a mere 1.96% of new accounts being apparent replacements. The result below summarizes our conclusion:

Result 7. There is a sizable contingent of newly-active user accounts on #GamerGate in Aug–Sep 2015. These users tend to have newer accounts, to interact less with others in general, and to interact with other #GamerGate users less in particular. Despite this, their inclusion appears to support the community structure.

5 Related Works

#GamerGate itself has seen little examination in the computing literature. Much of the existing work uses it merely as a source for labelled data rather than an object of study. Our primary point of comparison is the recent work by Chatzakou *et al.* [6], [21] on the properties of #GamerGate users ca. 2016. This study differs from their work in two key ways. First, we use historical data to study #GamerGate's growth rather than current data to study the present. Second, we focus on the structure of the interaction network through our analysis of mentions and retweets.

Beyond computing literature, we find further study of #GamerGate. Mortensen [22] gives a detailed accounting of the event and its impact, while Massanari [23] provides a sociological perspective on the role of Reddit's structure in supporting toxic cultures–with #GamerGate as an exemplar. This highlights one key aspect of further study: cross-platform analysis. While we focus our analysis on Twitter, #GamerGate also had a significant presence on Reddit. In particular, `r/KotakuInAction` was a hotbed of Pro-GG activity, while `r/GamerGhazi` became associated with Anti-GG sentiment. Works exist studying diffusion across platforms (e.g. [24]). It would be worthwhile to augment our results with an analysis incorporating Reddit's impact.

Meanwhile, Salter [25] argues that the abuse seen in #GamerGate should not be divorced from the platform on which it takes place. In particular: that platform design and admistration have an inherent impact on the way it is used—including usage for abuse. We find this interesting given that the initial weeks of #GamerGate were met with the development of new tools (e.g. [26]) to deal with the influx of harassment. This points to a weakness in Twitter: the built-in tools addressing harassment simply don't scale.

Lastly, Salter remarks that #GamerGate was subsumed into the "alt-right." We note that several alt-right individuals, including Milo Yiannapoulos and Mike Cernovich, appear in our set of notable Pro-GG users. Salter claims that these figures were not instrumental in organizing #GamerGate, which our results support. However, the presence of these individuals so early in the hashtag's lifespan—along with their impact on continued activity—illuminates the potential for polarization via campaigns like #GamerGate.

Firestorms as a whole have seen greater study. Pfeffer *et al.* [27] examine the causes of firestorms in the context of marketing. They identify a number of factors contributing to their explosive growth, including the speed of communication, lack of diversity leading to a "filter bubble," and weak ties leading to unrestrained information spread. However, we observe that #GamerGate lacks this explosiveness despite other similarities. While our work examines the cause of it's longevity, future work examining this substantially different growth may reveal valuable insight.

Pfeffer *et. al.*'s work was followed shortly by Lamba *et al.* [1], which focused on establishing the common structure of firestorms. Of particular relevance to our work, they challenge the notion that firestorms in general have a lasting effect on network structure. While our results confirm their observations in the case of the storms we compare to, #GamerGate belies this understanding.

Interestingly, Lamba *et. al.* further relate the study of firestorms to that of news cycles. We find it particularly notable that the temporal dynamics of the firestorms studied by Lamba *et al.* [1] are–at least at a high level–quite similar to the news cycles studied by Leskovec *et al.* [28]. While we have established that #GamerGate has large macro-level differences from other storms, it may be worthwhile to examine the role of the news cycle on this event.

Lastly, we remark that #GamerGate is often thought of in terms of harassment. Matias *et al.* [29] saw a sizeable amount of #GamerGate-related reports in their study of the harassment data from *Women, Action, and the Media*'s experiment with an alternate Twitter report process. While a long line of bullying detection work exists (see [30] and references therein), we have not seen such work applied to #GamerGate. This line of work is largely orthogonal to this paper and is only mentioned for the sake of completeness.

Community detection is one of the most widely-studied problems in network science. While we differentiate *detection*—the task of locating communities on a network—from *testing*—the task of evaluating the whether a partition "looks like" a community—there remains substantial related work. Fortunato [14] surveys the study of community detection up through 2010 in great detail. The inferential approaches discussed within (e.g. [31]) are particularly interesting due to the ease with which they may incorporate a prior reflecting the Twitter sampling mechanism.

More recently, Chakraborty *et al.* [15] completed an updated survey of community detection metrics. Much of the literature has been focused on either resolving problems with modularity (e.g. its resolution limitations [14] or preference for large, monolithic communities [32]) or adapting modularity to problem variants (e.g. [33]). While they find that conductance performs worse at detecting communities on synthetic networks, they also found it had one of the highest correlations with their validation metrics on a sparse, real-world coauthorship network. We note that the 1% sample of Twitter is extremely sparse.

However, due to the nature of our study much of this literature is not directly applicable. Our work is closer to that of Leskovec *et al.* [16], who evaluated the properties of real-world communities and elected to use conductance in a similar manner to our own use. Indeed, they find a "ubiquitous" nested core-periphery structure. While we did not examine nested structure, we remark that the structure shown in Fig. 8 clearly appears to have a core-periphery structure at the macroscopic level—a structure that modularity is "not well suited" to handle [34].

6 Limitations

We see three main limitations with this work. First: the dataset itself. The 1% sub-sample of the overall Twitter stream imposes constraints on the kinds of analysis that are practical. As noted in Sec. 3.1, the sub-sampling process removes most edges. Although we showed that conductance is relatively robust under this sub-sampling, it is possible (though, we believe, unlikely) that this caused false positives or negatives in our analysis of conductance among various groups. Further, our use of retweets for grouping is also fundamentally based on edges. This ultimately means that we must lean heavily on statistical tests to differentiate between these heavily sub-sampled populations.

We remark that while works on network reconstruction (e.g. [35] and references) and inferring latent network structure [36] exist, the sparsity of this sample combined with assumptions embedded in prior works precluded direct use. However, this is an intriguing avenue for further study.

Second, our use of only Twitter data. As noted in Sec. 3 and related work, there was substantial #GamerGate-related activity on Reddit (see e.g. [23]). Collecting and comparing this data to our Twitter sample may shed additional light on the structure of this event. In particular, an analysis of `r/KotakuInAction` and `r/GamerGhazi` á la Kumar *et al.* [37] may provide an important perspective on communication between Pro- and Anti-#GamerGate groups.

Lastly, we did not take advantage of the text data present in our sample. Each object in our dataset contains the tweet text in addition to the other properties we used. While we believe that our choice to focus on the graphical aspects provides interesting insight, large-scale study of the text associated with this event may yield further discoveries.

7 Discussion

So: is #GamerGate different? The answer is clearly yes—and to such a degree that perhaps a better question is "does comparing to Lamba *et al.* [1] make sense?" As a point of reference, we find prior storms valuable. Key differences, like #GamerGate's slow growth or apparent community structure, are clarified by the comparison even if #GamerGate was ultimately generated by different mechanics.

One may find closer comparisons in recent storms like #BlackLivesMatter and #MeToo. These events—at least on the surface—appear to share the enormous scale and longevity of #GamerGate, along with political undertones. Comparative analysis of these events' structures seems promising as future work.

Looking to other work, we find other lenses through which to understand this event. Mortensen [22] connects #GamerGate to football hooliganism. Like hooligans, she found that #GamerGaters were self-organizing, often assumed victim status, exhibited hypermasculine behavior, and were willing to attack the other "team." The first and last points are visible in our analysis. We note—based both on our own and prior work [25]—that it does not appear that the notable users in our sample led the organization process, and thus it is reasonable to call this *self-organization.*

We could also view #GamerGate as a large-scale example of a *negative mobilization* (or, more colloquially, *raids*) between communities. Recent work found that raids between sub-reddits exhibited strong homophily [37]: attackers communicated almost exclusively with attackers and defenders with defenders. As with the firestorm lens, this is not a perfect fit. While we do observe this behavior on the Pro-#GamerGate side, this pattern is notably absent on the Anti-#GamerGate side. This may be the result of the situation noted by technologist Caroline Sinders ([38], quoted in [25]):

> Using the hashtag in a tweet became akin to saying "Bloody Mary" three times in a mirror, except Bloody Mary actually showed up and she brought a bunch of friends. People, particularly women in games, couldn't talk about Gamergate publicly without getting harassed, so they just stopped talking about it on Twitter.

That is: we may not observe communication between Anti-#GamerGate users because they avoided the hashtag.

Taken together, our results support the findings of prior work both in computing and sociology. We found evidence for the existance of community structure amongst #GamerGate users, and found both that this community and interaction with notable Pro-#GamerGate users were highly correlated with continued activity. This leads us to conclude that this community structure played a fundamental role in the growth and sustenance of #GamerGate.

Acknowledgments

This work is supported in part by NSF grant CSF-1443905 and DTRA grant HDTRA1-14-1-0055.

References

[1] H. Lamba, M. M. Malik, and J. Pfeffer, "A Tempest in a Teacup? Analyzing firestorms on Twitter," in *IEEE/ACM ASONAM 2015*, Aug. 2015, pp. 17–24.
[2] A. Hern, "Felicia Day's public details put online after she described Gamergate fears," *The Guardian*, 2014.
[3] N. Wingfield, "Feminist Critics of Video Games Facing Threats in 'GamerGate' Campaign," *The New York Times*, Oct. 2014.
[4] Anonymous, *"i dated zoe quinn"*, 2014. [Online]. Available: https://archive.is/qrS5Q.
[5] ——, *Zoe Quinn, prominent SJW and indie developer is a liar and a slut*, 2014. [Online]. Available: https://archive.is/QIjm3.
[6] D. Chatzakou, N. Kourtellis, J. Blackburn, E. De Cristofaro, G. Stringhini, and A. Vakali, "Measuring #GamerGate: A Tale of Hate, Sexism, and Bullying," in *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017, pp. 1285–1290.
[7] A. Team, *Archive Team: The Twitter Stream Grab : Free Web : Download & Streaming : Internet Archive*, https://archive.org/details/twitterstream, 2018.
[8] F. Morstatter, J. Pfeffer, and H. Liu, "When is It Biased?: Assessing the Representativeness of Twitter's Streaming API," in *Proceedings of the 23rd International Conference on World Wide Web*, 2014, pp. 555–556.
[9] H. B. Mann and D. R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *The Annals of Mathematical Statistics*, vol. 18, Mar. 1947.
[10] A. Aron and E. N. Aron, *Statistics for psychology.* Prentice-Hall, Inc, 1994.
[11] E. Jones, T. Oliphant, P. Peterson, *et al.*, *SciPy: Open source scientific tools for Python*, 2001. [Online]. Available: http://www.scipy.org/.
[12] D. S. Kerby, "The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation," *Comprehensive Psychology*, vol. 3, Jan. 2014.
[13] S. Chess and A. Shaw, "A Conspiracy of Fishes, or, How We Learned to Stop Worrying About #GamerGate and Embrace Hegemonic Masculinity," *Journal of Broadcasting & Electronic Media*, vol. 59, Jan. 2015.

[14] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, Feb. 1, 2010.

[15] T. Chakraborty, A. Dalmia, A. Mukherjee, and N. Ganguly, "Metrics for community analysis: A survey," *ACM Comput. Surv.*, vol. 50, Aug. 2017.

[16] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters," *Internet Mathematics*, Jan. 1, 2009.

[17] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," *Physical Review E*, vol. 78, Oct. 2008.

[18] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," in *ICWSM*, 2009.

[19] S. Martin, W. M. Brown, R. Klavans, and K. W. Boyack, "Openord: An open-source toolbox for large graph layout," in *Visualization and Data Analysis 2011*, vol. 7868, 2011, p. 786 806.

[20] P. T. Metaxas, E. Mustafaraj, K. Wong, L. Zeng, M. O'Keefe, and S. Finn, "What Do Retweets Indicate? Results from User Survey and Meta-Review of Research.," in *ICWSM*, 2015, pp. 658–661.

[21] D. Chatzakou, N. Kourtellis, J. Blackburn, E. De Cristofaro, G. Stringhini, and A. Vakali, "Hate is Not Binary: Studying Abusive Behavior of #GamerGate on Twitter," in *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, 2017, pp. 65–74.

[22] T. E. Mortensen, "Anger, Fear, and Games: The Long Event of #GamerGate," *Games and Culture*, Apr. 2016.

[23] A. Massanari, "#Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures," *New Media & Society*, Mar. 2017.

[24] S. Zannettou, T. Caulfield, E. De Cristofaro, N. Kourtelris, I. Leontiadis, M. Sirivianos, G. Stringhini, and J. Blackburn, "The Web Centipede: Understanding How Web Communities Influence Each Other Through the Lens of Mainstream and Alternative News Sources," in *Proceedings of the 2017 Internet Measurement Conference*, 2017, pp. 405–417.

[25] M. Salter, "From geek masculinity to Gamergate: The technological rationality of online abuse," *Crime, Media, Culture*, vol. 14, Aug. 2018.

[26] R. Harper, *Good Game Auto Blocker*, Jul. 2018. [Online]. Available: https://github.com/freebsdgirl/ggautoblocker.

[27] J. Pfeffer, T. Zorbach, and K. M. Carley, "Understanding online firestorms: Negative word-of-mouth dynamics in social media networks," *Journal of Marketing Communications*, vol. 20, Mar. 2014.

[28] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the Dynamics of the News Cycle," in *Proceedings of the 15th ACM SIGKDD International Conference*, 2009, pp. 497–506.

[29] J. Matias, A. Johnson, W. E. Boesel, B. Keegan, J. Friedman, and C. DeTar, "Reporting, Reviewing, and Responding to Harassment on Twitter," Social Science Research Network, SSRN Scholarly Paper ID 2602018, May 2015.

[30] S. Salawu, Y. He, and J. Lumsden, "Approaches to Automated Detection of Cyberbullying: A Survey," *IEEE Transactions on Affective Computing*, vol. PP, 2017.

[31] E. Ziv, M. Middendorf, and C. H. Wiggins, "Information-theoretic approach to network modularity," *Physical Review E*, vol. 71, Apr. 14, 2005.

[32] C. P. Massen and J. P. K. Doye, "Identifying communities within energy landscapes," *Physical Review E*, vol. 71, Apr. 1, 2005.

[33] Y. Kim, S.-W. Son, and H. Jeong, "Finding communities in directed networks," *Physical Review E*, Jan. 2010.

[34] M. R. da Silva, H. Ma, and A. Zeng, "Centrality, Network Capacity, and Modularity as Parameters to Analyze the Core-Periphery Structure in Metabolic Networks," *Proceedings of the IEEE*, vol. 96, Aug. 2008.

[35] H. Boche, R. Calderbank, G. Kutyniok, and J. Vybíral, "A survey of compressed sensing," in *Compressed Sensing and Its Applications*, 2015, pp. 1–39.

[36] S. Linderman and R. Adams, "Discovering latent network structure in point process data," in *International Conference on Machine Learning*, 2014, pp. 1413–1421.

[37] S. Kumar, W. L. Hamilton, J. Leskovec, and D. Jurafsky, "Community Interaction and Conflict on the Web," *arXiv:1803.03697 [cs]*, Mar. 2018.

[38] C. Sinders, *That time the Internet sent a SWAT team to my mom's house / Boing Boing*, https://boingboing.net/2015/07/24/that-time-the-internet-sent-a.html, Jul. 2015.

J. David Smith received his B.Sc. degree from the University of Kentucky. He is currently working towards his PhD in Computer Science at the University of Florida under the supervision of Dr. My T. Thai. His is involved in research on discrete approximation algorithms, and on analysis and algorithms for large, sub-sampled networks—with a special focus on social networks. He is a recipient of the University of Florida Graduate School Fellowship.

My T. Thai (M'06) is a UF Research Foundation Professor at the Computer and Information Science and Engineering department, University of Florida. Her current research interests are on scalable algorithms, big data analysis, cybersecurity, and optimization in network science and engineering, including communication networks, smart grids, social networks, and their interdependency. The results of her work have led to 6 books and 140+ articles published in leading journals and conferences, including IEEE MSN 2014 Best Paper Award and 2017 IEEE ICDM Best Papers Award.

Prof. Thai has engaged in many professional activities. She has been a TPC-chair for many IEEE conferences, has served as an associate editor for Journal of Combinatorial Optimization (JOCO), Journal of Discrete Mathematics, IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Network Science and Engineering, and a series editor of Springer Briefs in Optimization. She is a founding EiC of the Computational Social Networks journal. She has received many research awards including a UF Provosts Excellence Award for Assistant Professors, UFRF Professorship Award, a Department of Defense (DoD) Young Investigator Award, and an NSF CAREER Award.