Measuring coordinated vs. spontaneous activity in online social movements

Camille Francois *Vlad Barash* *John Kelly*

August 2017

Abstract

Social media platforms provide people all over the world with an unprecedented ability to organize around social and political causes. However, these same platforms enable institutional and organized actors to engineer fabricated social movements to advance their agenda. These "astroturfing" or "false amplification" phenomenons leverage a variety of of tools and techniques, ranging from fully automated bot activity to accounts manned by extrinsically motivated (e.g. compensated) human operators. These campaigns also range from simple spam operations to sophisticated efforts involving numerous orchestrated accounts, sometimes coordinated across linguistic and cultural clusters. While the former category is straightforward to analyze via data mining methods, sophisticated fabricated campaigns in the latter category are engineered to mask their true nature from the public.

Working from the proposition that a large number of accounts controlled by a small number of coordinated entities will lack the behavioral diversity of a similar number of accounts controlled by uncoordinated individual actors, we propose a framework of signals (metrics) along three dimensions:

- *Network:* how accounts are connected to one another, and the clusters they form within the online conversation,
- *Temporal*: patterns of messaging across time in the online conversation,
- *Semantic:* an observation of the diversity of topics and meaning throughout the online conversation.

We test this framework on three case studies: the online conversation on #ColumbianChemicals in the U.S., the international discussion of the #DopingLeaks event, and an analysis of political discussions in Venezuela. In all three cases, we find what we assess to be anomalous, fabricated behavior on at least one dimension.

Introduction

In the last decade, social media has become a key communication and organization channel for collective action worldwide. Numerous papers (Borge-Holthoefer et al. 2011; Woolley & Howard

2016), essays (Toska, 2014), and books (Castells, 1998) have tackled the subject of groups self-organizing through social media to effect meaningful social change, from raising awareness of ALS through the #ALSIceBucketChallenge to helping topple longstanding autocratic regimes in the Arab Spring uprisings. Though often similar in their tools and tactics, experts and research groups have pointed out that there is a diversity of actors and motivations underlying coordinated social media campaigns (Gu, Kropotov & Yarochkin, 2017; Ferrara, 2017).

Developed originally from an interplay of commercial techniques to achieve marketing objectives and political propaganda techniques to manipulate public opinion (Thomas 2012), sophisticated disinformation operations aim to engineer and manipulate collective action by fabricating campaigns to appear legitimate and spontaneous. Both state and commercial actors have invested in standing capabilities to manipulate the networked public sphere. This phenomenon has been referred to as “astroturfing” or “false amplification” in existing literature (Weedon, Nuland & Stamos, 2017).

For example, in 2015, Adrian Chen (Chen, 2015) exposed a Saint Petersburg based operation, nicknamed a “troll farm”, which provides accounts operating under false identities and manned by employees to fabricate or manipulate conversations on Twitter. The Syrian conflict also offers a recent case study of a large scale international information operation deployed in support of a military engagement (Allbright et al. 2016). While most of the public conversation around these campaigns has focused on Russian actors (Rid 2017), the same techniques are employed by a variety of different actors around the world (Benedictus, 2016). We call this new form of online deception “*fabricated collective action”*.

Fabricated collective action is a form of social media behavior that seeks to create the impression of genuine collective sentiment (waves of panic, support or dissent from a particular agenda, etc.) In this paper, we attempt to characterize fabricated collective action, as distinct from spontaneous or organic online movements, in a formal manner.

The campaigns we study use a number of different tactics to manufacture waves of pseudo-public sentiment online. These include the use of “troll farms” in which operators control seemingly uncoordinated “sock puppet” social media accounts to drive a campaign, and also social media bot networks (Sanovich et al. 2015) in which sock puppet accounts are controlled by software (Ferrara et al. 2016). Despite the diversity of tools and tactics, the aim of fabricated social movement remains constant: to mask behind-the-scenes coordination and make online campaigns appear as organic groundswells of opinion.

Some fabricated social movements are easy to detect: accounts spamming the exact same message over and over to promote a hashtag, for instance, can be easily identified with simple data mining techniques. However the state of the art of fabricated social movements is now far more sophisticated, with a mixture of humans and software controlling accounts that “pass the Turing Test” of appearing like real people, often coordinating activity across linguistic and cultural boundaries.

Roughly, we distinguish between three “generations” of online manipulation via trolls and bots across social media platforms.

The first is comprised of lone trolls on the one hand, and simple bots on the other. These bots typically manifest one or a small few behaviors and lack convincing profile information.

The second generation features large numbers of convincingly human sock-puppet accounts run by a combination of coordinated trolls and more complex automated bots. These networks of hundreds or thousands of accounts typically follow each other and amplify each-other’s activities, appearing to be large neighborhood of like-minded citizens.

The third generation features the same coordinated human troll / bot sock puppet approach, but targeted to embed these assets into the “organic” online network. These accounts infiltrate real communities of interest, and seek to inject their own messages and framing into the discussion. Those accounts may leverage well constructed false identities and operate “feeds” on multiple platforms, therefore crafting the impression of being legitimate, influential users.

Despite the growing sophistication of these campaigns, we hypothesize that coordinated actors will still fail to replicate the diversity of real human behavior in at least some aspects. For instance, several thousand accounts made to look like supporters of a particular political movement and mainly following each other will have a greater network density and similarity of connections than an organic network of real citizens who, even though they share a political ideology, nonetheless follow other accounts of interest and build their networks via an accumulation over time of many discrete decisions about whom to follow and what to like. Similarly, one would expect that a large number of accounts controlled by a small number of entities would not display the same range of topics and sentiments as a similar number of accounts controlled by separate, organic individuals. The coordinated accounts would either appear more similar in their behaviors than a natural community of interest, or would appear less similar, with seemingly random behaviors.

A systematic comparison of organic movements and coordinated fabricated campaigns along key dimensions of behavior should reveal these discrepancies and suggest reliable markers of hidden coordination. We identify three key behavioral dimensions in our framework:

- *Network:* how accounts are connected to one another, and the clusters they form within the online conversation,
- *Temporal*: patterns of messaging across time in the online conversation,
- *Semantic:* an observation of the diversity of topics and meaning throughout the online conversation.

From this framework we derived 25 individual candidate measures of phenomena at the campaign, cluster (set of accounts), and individual account level. These signals were tested for two case studies of known coordinated (inorganic) and spontaneous (organic) campaigns on Twitter, revealing three promising *diagnostic* signals at the campaign level and a few promising *investigatory* signals at the cluster and actor levels. Some of the remaining signals showed little or no value, while others merit further refinement and testing.

Essential to our analysis is a networked model of the online conversation space, what Benkler has called the *Networked Public Sphere* (Benkler 2006) and Kelly has called *Cyber-Social Terrain* (Kelly 2014). In our model, social media platforms like Twitter constitute a cyber-social "network terrain" formed by the complex network of relationships (such as following in Twitter) among actors. The structure of this network determines who and what is visible to whom, and thus is the social landscape on which the struggle for influence occurs. We analyze our case study campaigns across specific network terrain "maps" in order to understand the relationships between participants and the patterns of campaign propagation across specific online clusters (discovered using machine learning analysis of network relationships).

A specific socio-political "network terrain" provides the natural habitat for emerging social movements, organic and coordinated alike. For example, the network terrain of US politics serves as the natural habitat for Black Lives Matter, Occupy, and alt-Right movements. The network context is critical for observing and measuring the emergence of these movements and is the basis for our analysis. The connections between actors in networked terrains are the vehicles of their influence, and thus critical for coordinating message content and propagation; by studying these connections, we can trace the coordination efforts.

The rest of the paper is organized as follows: first, we present our analytical framework for studying fabricated social movements. As part of this framework, we describe our mapping technology for identifying clusters around a particular social movement. Then we focus on three specific metrics for analysis in this paper. Next, we include a brief note on data sensitivity and privacy. We then proceed to three case studies: the online conversation on #ColumbianChemicals in the U.S., the international discussion of the #DopingLeaks event, and an analysis of political discussions in Venezuela. Finally, we conclude with discussion of the patterns we discover in the case studies and directions for future development of this research.

Analytical framework

In proposing a framework for analyzing fabricated social movements, we seek not only to distinguish these movements from truly organic ones, but also to create a formal method for studying fabricated patterns of fabricated, pseudo-grassroots (also, "astroturf") collective action.

Any such collective action must give the impression of a large group of people coalescing around a movement that is easy to describe and share with others. If the group is not well-connected enough, then it will be logistically difficult for any actor to organize the group's online behavior. If the group is not acting in temporal lockstep, then its message will not achieve a high frequency. Low-frequency messages do not appear as global trends; for example, Twitter's "trending" algorithm "identifies topics that are popular now, rather than topics that have been popular for a while or on a daily basis, to help you discover the hottest emerging topics of discussion on Twitter." (Twitter Support) Finally, if the group behind a fabricated social movement does not promote it with a coherent message, the movement's impact on the general public will be blunted by conflicting information.

These constraints suggest a natural set of three dimensions for analyzing fabricated social movements: the semantic dimension (how messages are formulated), the network dimension (how accounts within the campaigns are connected to one another), and the temporal dimension (when messages spread throughout the campaign). These dimensions, and their intersections, can yield discrete signals that can be used to scrutinize social media operations and assess if they display a suspicious degree of hidden coordination.

Our framework operates on three levels: **Event,** the level of an event within an entire social campaign (ex. all accounts and associated messages related to a specific #hashtag across a determined period of time) ; **Cluster,** the level of a community of actors participating in a social media campaign (e.g. A cluster of interconnected supporters of the same political party), and **Actor,** the level of an individual user participating in a social media campaign (e.g. an account identified by their social media handle, ex. @thisisanexampleaccount on Twitter).

Table 1 below shows our three dimensional analysis framework in more detail- specifically, the signals relevant for particular combinations of level and dimension. Not every combination of level and dimension has corresponding relevant signals.

	Network	Temporal	Semantic
Network	**Event**: how concentrated is online participation in the movement? Does it cover a broad range of politically / socially / culturally distinct clusters, or is it contained in a homogeneous "echo chamber"? **Cluster:** do clusters of actors who participate in the movement pay disproportionate attention to each other? **Actor**: do actors who	**Cluster**: How does participation in the movement vary between different clusters and over time? Are particular clusters always lagging behind the rest in participation (taking time to formulate a response)? **Actor:** how long does the average actor participate in the movement?	**Cluster**: How topically diverse is all discourse among clusters participating in the movement?

	participate in the movement do so in conjunction with their clusters, or independently of them?		
Temporal		**Event**: Does participation in the movement follow an unusually temporally regular pattern, when compared to spontaneous human posting behavior? **Cluster:** do clusters of actors coordinate their activities, even across time zones? **Actor**: Do some actors behave similarly to pre-identified troll or spambot accounts with regard to their temporal posting patterns?	**Event/Cluster/Actor**: How does the diversity of the discourse among all participants / specific clusters / individual actors participating in the movement vary over time?
Semantic			**Event/Actor:** How topically diverse is the discourse around the movement among all actors / individual actors?

Table 1. Three-Dimensional Analysis Framework

This framework is a helpful methodological tool, but it would not be useful without operational definitions, which we capture via mathematical metrics of campaign activity. We map each signal in Table 1 above to a discrete metric in Table 2. We provide key definitions for understanding these metrics, and explain any non-obvious activity metrics, in more detail below.

Dimension #1	**Dimension #2**	**Analysis Level**	**Resulting Metric**

Network	Network	Event	Concentration: Entropy *E*
Network	Network	Cluster	Inter-cluster homophily *H*
Network	Network	Actor	% of actor's cluster participating in campaign, by number individuals or number posts *CP*
Network	Temporal	Cluster	Time delta between peak date of campaign participation by accounts in a cluster and peak date of campaign participation by all actors *TD*
Network	Temporal	Actor	Commitment by actor *M*
Network	Semantic	Cluster	Semantic Diversity by Cluster Omega Ω_S
Temporal	Temporal	Event	Peakedness *P*
Temporal	Temporal	Cluster	Dynamic Time Warp alignment between Clusters *DS*
Temporal	Temporal	Actor	Dynamic Time Warp alignment between actors *DU*
Temporal	Semantic	Event/Cluster/Actor	Semantic Diversity over time by Event / Cluster / Actor $\Omega_{TE,}$ $\Omega_{TS,}$ Ω_{TA}
Semantic	Semantic	Event/Cluster/Actor	Semantic Diversity by Event / Actor Ω_E,Ω_A

Table 2. Mapping of Signals to Metrics

Key Definitions

Network
The network dimension assumes that actors participating in a campaign are connected to each other in a directed network *G* (i.e. a connection from actor *a* to actor *b* does not imply the reverse). Twitter following networks are an example of directed networks: many people follow Twitter celebrities, but those celebrities do not follow their fans back as a general rule.

Cluster
When calculating metrics at the network level, we assume that each actor participating in a campaign belongs to exactly one cluster *c.* We define a cluster as a set of actors with shared patterns of attention to the outside world, as opposed to a shared pattern of interconnections.

Identifying Networks and Communities
In order to identify relevant networks and clusters within those networks, we leverage Graphika's mapping framework, tested on more than eight hundred different sociocultural contexts with many academic applications (Etling et al. 2008; Kelly et al. 2012). The key unit of analysis in this framework is a "map," which is a collection of key social media accounts around a particular social context. For this analysis, we constructed three maps, one for the #ColumbianChemicals case study, a second for the #DopingLeaks case study, and a third one for the Venezuelan politics case study. A map is composed of "nodes," which are the social media accounts pertinent to the online conversation. Each node is connected to one or more nodes in the map through "edges" - edges represent social relationships embedded in the respective social media platform (e.g., "following" for Twitter). Graphika represents the pattern of connections between nodes via a Fruchterman-Rheingold visualization algorithm (Fruchterman & Reingold, 1991). The algorithm places nodes in a map onto a canvas according to two principles: first, a "centrifugal force" acts upon each node to push it to the edge of the canvas; second, a "cohesive force" acts upon every connected pair of nodes to push them closer together.

Each node in a the map belongs to exactly one "cluster" and one "group." A cluster is a collection of nodes with a shared pattern of interests (e.g., a collection of Twitter accounts who all follow US Tea Party politicians). Each cluster has a label (e.g.,"Tea Party"). A group is a collection of clusters with similar interest profiles (e.g., a collection of "Tea Party," "Constitutional Conservatives," etc. clusters into a "Conservative" group)[1].

Key Metrics Explained
To illustrate metrics in this section, we will use a toy campaign example. The example consists of 100 actors connected in a network *G.* The network *G* further breaks down into exactly two clusters *A* and *B,* each with exactly one half of the total population. The overall number of

[1] The process for generating clusters, groups, labels, and colors for a map is 95% automated, as follows: a proprietary clustering algorithm automatically generates clusters and groups for a map; subsequently, the map-making process uses supervised machine learning to generate labels for clusters and groups from human-labeled examples. At the end of the automated process, a Subject Matter Expert, an individual well-versed in the topic and/or geographical area covered by the map, performs a quality assurance check on the cluster and group labels.

connections from members of *A* to any other actor in the network is 500, while the number of connections from members of *A* to members of *B* is 200. In other words, *A* has 300 intra-group connections and 200 extra-group connections. The campaign proceeds over the course of ten days, and the first of those days features the highest level of campaign activity, with exactly one quarter of all actors participating.

Concentration: Entropy E

This metric is the degree to which a particular campaign is concentrated in one cluster versus diffused among many different clusters. Given a mapping of actors to clusters, which we describe in more detail below, the entropy of a campaign is the information theoretic entropy (Shannon & Weaver 1949) of the distribution of actors active in the campaign among different clusters. In our toy example, the Entropy of the campaign is:

$$E = -\sum_{i=1}^{|c|} p(c(i))log_b(c(i)) = -0.5log_2(0.5) - 0.5log_2(0.5) = 1$$

In general, low values of *E* represent campaigns concentrated in one cluster, while high values of *E* represent campaigns distributed among a wide array of clusters.

Inter-cluster Homophily H

The inter-cluster Homophily *H* (McPherson et al. 2003) is the degree to which clusters active around the campaign are more interconnected than one would expect by random chance. Mathematically, *H* is calculated for an ordered pair of clusters *A, B*. The quantity *H(A,B)* is the ratio of the actual number of connections from members of *A* to members of *B, E(A,B),* to a normalizing factor ρ that assumes that members of *A* make their connections to all other nodes at random. In the random baseline, the number of connections from members of *A* to members of *B* is the number of all connections from members of *A* to any other node in the network *G*, multiplied by the fraction of *G* that *B* represents. In our toy example, the Homophily from cluster *A* to cluster *B* is:

$$H(A,B) = \frac{E(A,B)}{\rho} = \frac{200}{500*0.5} = 0.8$$

Values of *H* below 1.0 represent heterophily, or lower-than-expected interconnectivity between clusters. Values of *H* equal to 1.0 represent the baseline random expectation. Values of *H* above 1.0 represent homophily, or higher-than-expected interconnectivity.

H is superlinear, so a value of *H* = 4.0 is much more than twice as interconnected as *H* = 2.0.

While the random baseline for Homophily is established in the citation above, we recognize that it may be an excessively low baseline for our empirical analyses. Empirical social networks are never randomly interconnected, so using a random baseline would identify many clusters — whether coordinated or not — as having a high degree of homophily. Therefore, when possible, we use H values for cluster pairs where we would expect low / high values (e.g. ideologically separate / ideologically aligned clusters) in the same networked terrain as our case study as a baseline.

Commitment M

Commitment to a particular campaign is measured in two ways:

M_E , the number of subsequent engagements with the campaign by an actor

M_T , the length of time between first and last recorded engagement with the campaign by an actor

Semantic Diversity Ω

Semantic diversity of a particular actor's / clusters / campaign's messaging is based on the assignment of messages to topics. We use LDA, a common method for identifying topics in text data (Bruce, 1963). Once messages have been assigned to topics, we use a method covered in (Hoffman et al. 2013) to calculate a semantic diversity score for the message set. The authors of the referenced work represent their measure of semantic diversity as the probability that two documents chosen from the corpus at random with replacement will be on the same topic. In our case, the corpus is the message set, and the documents are actor Tweet histories, aggregated by actor. We run the LDA algorithm for 15 iterations, with a number of topics no less than 20% of the number of documents and no more than 30. We average semantic diversity over 20 distinct runs of the LDA algorithm on the same corpus to smooth out variations due to the initial conditions for a particular run. For topics that do not co-occur in documents, we assign a topic distance score of 1000.

Versions of Ω are run for individual actors (Ω_A), clusters (Ω_C), or events (Ω_E). These metrics can also be run for all messages within a particular time period (Ω_{T*}) to calculate the change in semantic diversity over time.

Due to the complex process of calculating semantic diversity, we do not provide a toy example for this metric. However, we provide the following guidelines: semantic diversity scores of less than one represent actors who exclusively post about the same topic, characteristic of fabricated campaigns. Semantic diversity scores between 1 and 100 represent actors who post on a variety of topics, characteristic of normal human activity. Finally, semantic diversity scores above 100 represent actors who post on an extremely diverse set of topics, characteristic of spambots or actors who bridge different cultural and/or linguistic clusters (e.g. actors who post in different languages, etc.)

Peakedness P

The Peakedness of a campaign is the fraction of all activity that occurs in the day with the most campaign-related activity during some time frame. In our toy example, $P = ¼ = 0.25$. We use this simple definition, acknowledging that it has limitations such as time zones and day boundaries, as a first approximation of the temporal concentration of activity around a coordinated campaign. We will refine it in future work to overcome these limitations.

*Dynamic Time Warp Alignment D**

Dynamic Time Warp is an algorithm (Salvador & Chan, 2004) for comparing two temporal sequences of activity. Here, we use it to compare the activities of individual actors (*DU*) or entire clusters (*DS*). In general, the Dynamic Time Warp between two sequences *S1* and *S2* is the number of warping transformations that are required to change *S1* into *S2*. We follow (Woolley et al.), who use Dynamic Time Warp to identify bots and trolls in a different social media setting. The sequences *S1* and *S2* represent sequences of campaign-related posts, either by two individuals (for *DU*) or aggregated across all members of two different clusters (for *DS*). In order to correct for differences due to raw posting volume, we first normalize each sequence by its sum. We then calculate the Dynamic Time Warp between *S1* and *S2*. In order to correct for post sequence length, we divide the Dynamic Time Warp by the sum of the lengths of *S1* and *S2*. The result gives us a fully normalized Dynamic Time Warp distance measure. Finally, we subtract this distance measure from 1.0 to get our alignment measures *DU* and *DS*.

Metrics for Case Studies

For the purposes of our case studies, we focus on three specific signals from Table 1: network coordination between clusters, measured as inter-cluster Homophily *H*, temporal coordination between clusters, measured as the Dynamic Time Warp between Clusters *DS* , and semantic diversity of the discourse among all actors, measured as Ω_E. We intend to cover all signals and metrics from Tables 1 and 2 in future work.

Data Sensitivity and Privacy

Our analysis leverages publicly available social media data, such as Twitter posts and following relationships between Twitter users. However, even in this public space, we seek to acknowledge that actors may have an expectation of privacy (Madden et al. 2013).

Actors may consider their social media activity to be occurring in a private or semi-private space available only to themselves and friends, colleagues, or loved ones. It is not clear whether actors who are participating in a social campaign intended to have a widespread effect have the same expectations. Since it would be impractical to interview every actor we analyze, we err on the side of masking the identities of all actors in our research. The majority of our analysis focuses on Event-level and Cluster-level aggregated metrics. Whenever we present individual-level metrics below, we conceal the true screen name and other personally identifying information of the actors involved.

Case Study 1: Revealing coordination in the #ColumbiaChemicals campaign

The Columbia Chemicals online campaign took place on Sept. 11, 2014 and was exposed as a sophisticated fabricated social movement, with coordinated activity originating in one country (Russia) and targeting a local US cluster. The campaign spread fabricated news that a chemical plant in St. Mary Parish, Louisiana, Columbia Chemicals, had exploded. News spread under the hashtag #ColumbiaChemicals on Twitter. A fake video of the explosion spread, as well as doctored screenshots of news outlets reporting on the incident (both local and national -

including CNN). The hoax was debunked within hours, but the campaign nevertheless experienced some degree of success, fooling many users online and some journalists. Gilad Lotan (Borthwick et al. 2015) and Adrian Chen (Chen, 2015) have both covered the underpinnings of this campaign.

We generated a map for #ColumbiaChemicals. We identified matching nodes as any Twitter user who used the one of the following hashtags related to Columbian Chemicals between 09/11/2014 and 09/17/2014 (we collected extra days at the end of the campaign to ensure that we tracked both activity during the hoax and in its aftermath): #DeadHorse, #ColumbianChemicalsInNewOrleans, #ChemicalAccidentLouisiana, #LouisianaExplosion and #ColumbianChemicals.

The total number of accounts spreading the hashtag was relatively low, so we used a snowball collection method to add the network neighbors of the hashtag users to the map[2].

For reference, we provide an image of the #ColumbiaChemicals map in Figure 1.

Figure 1. #ColumbiaChemicals map

The map showed a few Russia-based accounts (red and pink, above), many US politically conservative accounts (green, above), a number of celebrity / local news accounts (blue above), and a large number of social media marketing / spam accounts (purple above). The

[2] We used Graphika's regular process for constructing the map: we first reduced the map to a key core of actors. Subsequently, we used Graphika's clusteration algorithm to map each actor to a distinct cluster. We then used machine-learning based cluster labeler to generate a human-readable label for each cluster, which was checked by a Subject Matter Expert.

Russia-based accounts followed state-supported media outlets such as rt.com, and Russian political figures aligned with Vladimir Putin. The Conservative accounts were of a wide variety — the clustering algorithm identified distinct clusters of Tea Party / Gun-interest conservatives, Christian Conservatives, and Libertarians. The celebrity / local news accounts were mostly related to Louisiana, the site of the hoax attack. Spam / Social Media Marketing clusters are focused on advertising and straightforward spam content. We chose to leave them in to identify some spam content as being generated by specific clusters. Spam-focused clusters can sometimes push coordinated campaigns — for example, if a hashtag becomes trending in a specific location, some spam accounts may automatically include it in their Tweets to raise visibility.

Results: Network coordination (Homophily)

Network coordination (Homophily) scores for #ColumbianChemicals: Coordination only between ideologically aligned clusters
6.48: (Christian Conservative / Guns ; Conservative / Tea Party)
2.7: (Conservative / Tea Party ; Constitutional Conservatives)
<0.25: (Russian Paid Twitter Accounts, all Conservative clusters)

Table 3. Network Coordination in #ColumbiaChemicals map

We do not find unusual patterns of network coordination between clusters around #ColumbianChemicals (Table 3). The Conservative clusters pay a lot of attention to each other, but we expected this level of coordination, given that these clusters are ideologically aligned.

Results: Temporal Coordination (Dynamic Time Warp)

We observed temporal coordination between only a small subset of the clusters. Notably, there does not seem to be any temporal coordination between Russian Paid Twitter Accounts and other clusters. The low overall level of activity around the #ColumbianChemicals campaign makes it challenging to observe strong temporal coordination. Because of the very low level of temporal coordination, we chose to not provide an accompanying figure or table.

Results: Semantic Diversity (Ω_E)

Figure 2. Semantic Diversity among #ColumbianChemicals actors

Over one third of the actors tweeting about #ColumbianChemicals and related hashtags had a semantic diversity score under 1.0, indicating that they tweeted exclusively on one topic (Figure 2). We investigated these actors more closely and found clear examples of fabricated social activity on behalf of these actors:

One actor posted 1, 2, 3, etc. numbers going up to 100 with tweets about their daily life, but with #ColumbianChemicals prepended. At tweet #100, the actor posted to the effect of "YES I DID IT!", suggesting that this actor had an assignment to post 100 tweets with the hashtag. This is the text of another one of their posts, translated from Russian to English:

("Right now we're pushing the #ColumbianChemicals hashtag to trend worldwide. "Go" is with us, as well as #RT.")

It is not clear what “Go” is referring to here, but we hypothesize #RT is a reference to rt.com, “the first Russian 24/7 English-language news channel which brings the Russian view on global news” (RT.com).

Another actor wrote several tweets clearly indicating their goal of making #DeadHorse (a hashtag related to the hoax) into a Twitter trending hashtag. These are the English translations of their Tweets:
"Go to sleep this is an operation #deadhorse",
"#deadhorse continues to gain traction so let's go to sleep",
"let's wake up a dormant story #deadhorse".

Case Study 2: Revealing coordination in the #DopingLeaks campaign

Figure 3. Map of international conversation around the #Doping Leaks incident

The #DopingLeaks campaign was an organized response to widespread allegations of doping by Russian athletes. The World Anti Doping Agency (WADA), the agency that issued the allegations, reported that the Russian State-sponsored hacking group known as *Fancy Bear* accessed a database of confidential medical data and released the drug regimens of top US Olympians[3]. Following the hack, the #DopingLeaks social media campaign propagated information, extracted from WADA's database, that US Olympic athletes tested positive for controlled substances. The campaign was focused on shaming US Olympic Athletes and WADA for presumed hypocrisy and did not indicate that, for example, four US athletes who tested positive for controlled substances were given medical exemptions.

Because of the social media's campaign close temporal proximity to the state-sponsored hacking effort, #DopingLeaks is an interesting case to study. While the campaign surely entails organic activity and discussions, we hypothetized that a coordinated social media campaign was crafted as a companion to the hacking effort, to ensure the material hacked could be widely circulated and discussed.

We generated a Twitter map of actors who participated in the #DopingLeaks campaign between 09/10/2016 and 09/17/2016, as well as the Twitter followers of these actors (Figure 3). In this map, we observed a) accounts who identified as supporters of Donald Trump in the US Presidential election; b) accounts who identified as supporters of Vladimir Putin and Russians

[3]https://arstechnica.com/security/2016/09/anti-doping-agency-pins-leak-of-us-gold-medalists-data-on-russian-hackers/

expressing anti-West / nationalist sentiment; c) right-wing nationalist German accounts; and d) accounts who identified as supporters of Wikileaks. The supporters of Donald Trump were mostly on the the extreme end of the US political conservative spectrum, and included members who identified as being part of the "Alt-Right" nationalist movement. These actors followed @realDonaldTrump, along with media accounts such as Breitbart news. Supporters of Vladimir Putin generally followed accounts and linked to media outlets with an anti-Western, pro-Putin bent (e.g. https://ria.ru/). Right-wing nationalist German accounts followed and linked to anti-Muslim, anti-refugee accounts and media outlets. Supporters of Wikileaks generally followed Wikileaks, and Wikileaks-aligned accounts.

Results: Network coordination (Homophily)

Typical organic ranges (Control: GER political map used as baseline)
Expected low: not ideologically aligned 0.05: (German Progressives|Hackers, German AFD Party|Anti-Islam)
Expected high: ideologically aligned 0.40: (German Progressives|Hackers, FDP|Young Liberals)

Table 4: baselines for network coordination in Doping Leaks map

Surprising high scores for #DopingLeaks: Suggest coordination between clusters
2.0: (German Nationalist Anti Muslims, Pro Putin Russian Abroad)
2.7: (Wikileaks, Pro Putin Russian Abroad)
1.2: (German Nationalist Anti Muslims, Alt Right)

Table 5: Network Coordination in Doping Leaks map

We observed high levels of network coordination between clusters in the Doping Leaks map. To our surprise, we observed high levels of network coordination even between clusters that are separated by cultural and linguistic barriers, e.g. US, Russian, and German accounts (see Table 5). As a baseline (Table 4), we show that these levels of coordination are much higher than both ideologically aligned and ideologically non-aligned clusters in a German political map crafted on the same network terrain.

Results: Temporal Coordination (Dynamic Time Warp)

Alt Right	Pro Putin Russian accounts	**0.988**
Alt Right	German Nationalist Anti-Muslim	**0.941**
Alt Right Pro Trump	Pro Putin Daily Life accounts	**0.987**
Alt Right Pro Trump	German Nationalist Anti-Muslim	**0.942**
Pro Putin Daily Life Trolls	German Nationalist Anti-Muslim	**0.940**

Table 6: Top Temporal Alignment Community Pairs in Doping Map

Figure 4. Heatmap of Temporal Alignment in Doping Leaks Map

As with network coordination, we observed high levels of temporal alignment between clusters in the Doping Leaks map. This level of coordination corresponds to posting within a few hours of each other — unusual, considering that these clusters are located across three different continents, with possibly the same general agenda but not necessarily specific common causes, tweet in different languages, and from different timezones. Table 6 shows the pairs of clusters

A, B with highest levels of dynamic time warp *DS(A,B),* while Figure 4 shows a heatmap of temporal alignment between all clusters in the map (darker colors -- higher levels of temporal alignment).

Results: Semantic Diversity (Ω_E)

Figure 5. Semantic Diversity among participants in Doping Leaks campaign

We observed even lower semantic diversity in the Doping Leaks map (Figure 5). Nearly 72% of accounts had a semantic diversity score of less than one, suggesting a focus on one topic exclusively. We investigated these accounts more closely and found that they did not, as in the #ColumbiaChemicals case, explicitly identify their purpose as making #DopingLeaks trend or fueling a social media campaign. Nor did these accounts include #DopingLeaks in unrelated tweets about their daily lives. To the contrary, they tweeted about #DopingLeaks in alignment with the campaign's message — shaming US Olympics and WADA for their presumed hypocrisy, with the effect of distracting from the doping scandal around Russian Olympic athletes.

Case Study 3: Comparing Coordinated and Spontaneous Activity in the Venezuelan Political Environment, 2017

Context

Venezuela has been in political turmoil for well over a year (GlobalVoices), suffering extreme food and medicine shortages because of poor governmental policy and a fall in global oil prices (Casey). Tensions erupted in April 2017, when Venezuela's President, Nicolás Maduro, effectively dissolved the country's parliament and called for a new constitution (Prieto). Waves of protests, resulting in violence, ensued.

The tension between these two sides - Maduro's government and a public with 80% of its citizens calling for Maduro to step down (Zuñiga and Miroff) - has played out online as well. This tension prominently came to head when Twitter unexpectedly suspended several Venezuelan propaganda accounts (Cosimi, 2017). Maduro promptly claimed that "thousands" of accounts had been taken down and that Twitter's intervention was an "expression of fascism". Maduro's Minister of Information, Ernesto Villegas, estimated that 180 accounts had been hit (Daniel, 2017).

The Venezuela case study presents an opportunity for studying both spontaneous, organic online movements against the Venezuelan government and coordinated, pro-regime campaigns emanating from organized teams within the government, in the same network terrain.

Data Collection

Terrain Map

In our previous case studies, we examined individual coordinated movements within the context of their own participants. This approach is inherently biased, since it does not allow for comparison of coordinated behavior to non-coordinated behavior in the same context. For the Venezuela case study, we wanted to directly compare spontaneous and coordinated movements. Therefore, we first constructed a “network terrain” map of Venezuelan politics in 2017 to serve as a sociocultural context for both kinds of movements.

Figure 5. Map of coordinated and spontaneous Venezuelan messaging campaigns on Twitter from 04/01 to 05/21/2017.

The terrain map includes all actors who followed key Venezuelan political accounts, including main political figures and news outlets (as determined by subject matter experts). The visualization of the map in Figure 5 shows a clear bi-partite structure, with Venezuela Socialist (pro-regime) clusters on the right and Venezuela Resistance (anti-regime) on the left. The Venezuelan Socialist clusters include clusters aligned with the PSUV (the Venezuelan Socialist party), “Chavistas” who believe in Hugo Chavez’s vision for the country, and an anti-US cluster that sees anti-regime opposition as a sign of US imperialist meddling in Venezuelan affairs. The Venezuelan Resistance clusters include clusters aligned with Voluntad Popular, a centrist political party in Venezuela that was formed in reaction to alleged infringements of individual freedom and human rights by Maduro (Wikipedia). Beyond Voluntad Popular, these clusters include a wide assortment of interest groups opposed to Maduro: from actors and performers to Libertarian Democrats to student groups.

As would be expected, the map includes some connections to Latin American politics beyond Venezuela, specifically, Colombia and Cuba. Latin American Politics accounts heavily overlap Venezuela Resistance accounts in the visualization, which suggests that these two groups are well connected on Twitter.

Movements

We tracked the spread of all eight hashtags against our network terrain map: two coordinated hashtags, and six spontaneous protest hashtags. Overall, we analyzed data on 185,325 tweets and 53,647 actors.

Two coordinated propaganda hashtags: these hashtags were promoted by Telegram channels that are suspected to be run by the Venezuelan Ministry of Communications, or to be run by politicians close to Maduro, such as Mario Silva[4].

#VenezuelaEsChavista - promoting “Chavista” unity and support of the current government
#RichardMardoTerrorista - an attack hashtag against Richard Mardo, a leader of one of the main opposition parties in Venezuela, First Justice (Primero Justicia, PJ). Mardo has faced harsh criticism from his Government as one of the most vocal opponents of some of key members of the government, including former congressional deputy Diosdado Cabello and Vice President Tareck el Aissami.

[4] http://www.economist.com/blogs/americasview/2011/08/venezuelan-television

Figure 6: Screenshots from a Telegram channel promoting the coordinated hashtag #VenezuelaEsChavista in support of the Venezuelan government.

Six spontaneous civil protest hashtags: We selected civil protest and human rights related hashtags by surveying civil society groups' activities in Venezuela, looking for grassroots Twitter campaigns who leveraged popular hashtags and hashtags used by the anti-regime opposition to bring attention to their messages[5].

Results: Network coordination (Homophily)

Network coordination (Homophily) scores for Venezuela pro-regime clusters: Coordination only between ideologically aligned clusters	**Network coordination (Homophily) scores for Venezuela anti-regime clusters:** Coordination only between ideologically aligned clusters
24.71 (VE Anti-US, PSUV Activists)	**1.61** (Free Venezuela, VE Student Movement)
10.49 (Venezuela Socialist, VE Anti-US)	**1.27** (VE United Resistance, Free Venezuela)
7.44 (Chavista, VE Anti-US)	**0.90:** (VE Student Movement, VE United

[5] Selected hashtags: *#PlantónNacional, #ArteSinMordaza, #EstoEsTuyo, #NoEstasSolo, #NoMásRepresión, #NoMásDictadura.*

	Resistance)

Table 6. Network Coordination in Venezuela terrain map

We find much higher levels of network coordination between pro-regime clusters than between anti-regime clusters. By comparing network coordination between two different sets of ideologically aligned clusters on the same network terrain, we can draw a clear comparison: pro-regime clusters, which we would expect to be generating coordinated movements, are an order of magnitude more aligned than anti-regime clusters, which we would expect to be generating spontaneous movements.

Results: Temporal Coordination (Dynamic Time Warp)

Among participants in coordinated campaigns			**Among participants in spontaneous campaigns**		
PSUV	Chavista.PSUV	**0.993**	VE Anti-US	PSUV Activists	**0.997**
Venezuela Socialist	PSUV	**0.993**	Venezuela Socialist	PSUV Activists	**0.994**
Venezuela Anti-US	PSUV|Ecuador	**0.993**	PSUV|Ecuador	PSUV Activists	**0.993**
Venezuela Socialist	VE Anti-US	**0.992**	PSUV Activists	Chavista	**0.992**
VE Anti-US	PSUV	**0.992**	PSUV|Ecuador	VE Anti-US	**0.991**

Table 6: Top Temporal Alignment Community Pairs in Venezuela Map

Among participants in coordinated campaigns		
VE Student Movement	VE United Resistance	**0.967**
VE Student Movement	Free Venezuela	**0.965**
VE United Resistance	Free Venezuela	**0.965**

Table 7: Sample Temporal Alignment Pairs for Anti-regime Movements in Venezuela map

Figure 5. Heatmap of Temporal Alignment for Participants in Coordinated Campaigns in Venezuela map

Figure 5. Heatmap of Temporal Alignment for Participants in Spontaneous Campaigns in Venezuela map

We found very high levels of temporal alignment among clusters in the Venezuela map. Table 6 demonstrates that, irrespective of whether we look at participants in coordinated or spontaneous movements in the map, pro-regime clusters have the highest levels of coordination. In contrast, Table 7 shows that the temporal coordination for anti-regime clusters around the spontaneous movements is several percent lower. Our results also demonstrate that our coordination metric inflates levels of coordination — the vast majority of clusters had temporal coordination of over 95%. We believe that this inflation is due to calculating temporal coordination at the day level instead of the hour level, and will address this issue in a future iteration of this analysis.

As Figure 5 demonstrates, activity around the coordinated campaigns occurs predominantly in the pro-regime clusters we identified in the previous section. Indeed, there is not enough coordinated campaign activity by the anti-regime clusters to demonstrate any level of temporal coordination. In Figure 6, we see that the network terrain as a whole engages in the spontaneous campaigns, and that the temporal coordination of pro-regime clusters is higher than the temporal coordination of anti-regime clusters, in line with Tables 6 and 7.

Results: Semantic Diversity (Ω_E)

Figure 7. Semantic Diversity Among Participants in Coordinated Campaigns in Venezuela

Figure 8. Semantic Diversity Among Participants in Spontaneous Campaigns in Venezuela

We found that a very small percentage of participants in any campaigns in the Venezuelan network terrain — spontaneous or coordinated — had a semantic diversity score less than one. However, more participants in coordinated campaigns had such a low score vs. participants in spontaneous campaigns. Furthermore, nearly half of the participants in coordinated campaigns had a diversity score above 100, indicating an abnormal lack of topical consistency within their Tweets. As a result, only about half of the participants in coordinated campaigns fell into the semantic diversity range between 2 and 100, which we associate with non-coordinated human activity; in contrast, almost all of the participants in spontaneous campaigns fell into that range.

Discussion

Observations of the three case studies suggest the following pattern at this point: #ColumbiaChemicals was an early effort at a fabricated social campaign, with posts by campaign members openly discussing their intentions, little network connectivity and temporal coordination. The Doping Leaks and coordinated Venezuelan campaigns were much more sophisticated, with extensive network and temporal coordination among clusters that have not been ideologically aligned in the past. The Doping Leaks campaign also featured a much more narrow discourse, with nearly three-quarters of the participants posting about only one topic. Our case studies suggest that efforts at online coordination were refined after the #ColumbiaChemicals hoax and used more artfully in the Doping Leaks campaign, although not necessarily by the same actors.

Analysis of the three case studies points to a natural evolution for our metrics framework. The #ColumbiaChemicals and Doping Leaks case studies were based on network maps of the campaign participants, and thus do not allow for comparison to spontaneous campaigns occurring in a similar network terrain. We focused on campaign participants exclusively for these cases, because they do not fit neatly into a particular sociocultural context, but instead span linguistic and cultural boundaries. In contrast, for the Venezuela case study, we witnessed spontaneous and coordinated campaigns playing out in the same sociocultural context, and thus were able to construct a network terrain as a backdrop for comparison of campaign activity patterns. As a result, our Venezuela case study yields clearer results than either #Columbia Chemicals or Russian Doping Leaks: we can directly observe that coordinated, pro-regime campaigns feature much higher levels of anomalous network, temporal, and semantic activity than spontaneous, anti-regiment campaigns.

References

Allbright, M., Bildt, C., Burns, N., Hunstman, J. The Atlantic Council's Breaking Aleppo. URL: http://www.publications.atlanticcouncil.org/breakingaleppo/. Retrieved on 04/15/2017.

Benedictus, L. (2016). Invasion of the troll armies: from Russian Trump supporters to Turkish state stooges. The Guardian, Sunday November 6, 2016. URL: https://www.theguardian.com/media/2016/nov/06/troll-armies-social-media-trump-russian. Retrieved on 04/15/2017.

Benkler, Y. (2006). The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press, New Haven, USA.

Borge-Holthoefer, J., Rivero, A., García, I., Cauhé, E., Ferrer, A., Ferrer, D., Francos, D., Iñiguez, D., Pérez, M., Ruiz, G., Sanz, F., Serrano, F., Viñas, C., Tarancón, A. & Moreno Y. (2011)
Structural and dynamical patterns on online social networks: the Spanish May 15th movement as a case
study. PloS One, 6, e23883.

Borthwick, J. & Lotan, G. Media hacking. URL: https://render.betaworks.com/media-hacking-3b1e350d619c. Retrieved on 04/15/2017.

Bradshaw, S. & Howard, P.N. (2017). "Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation." Samuel Woolley and Philip N. Howard, Eds. Working Paper 2017.12. Oxford, UK: Project on Computational Propaganda. comprop.oii.ox.ac.uk. 37 pp. Retrieved on 07/17/2017.

Casey, N. (2016) Venezuelans Ransack Stores as Hunger Grips the Nation. New York Times: https://www.nytimes.com/2016/06/20/world/americas/venezuelans-ransack-stores-as-hunger-stalks-crumbling-nation.html. Retrieved on 07/13/2017

Castells, M.The Information Age: Economy, Society and Culture, Vols. I-III. The Information Society 14, no. 4: (1998), http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/c/Castells:Manuel.html.

Chen, A. The Agency(2015). New York Times. URL: https://www.nytimes.com/2015/06/07/magazine/the-agency.html. Retrieved on 04/15/2017.

Cosini, S. (2017). “È giusto che in Venezuela Twitter sospenda gli account governativi?”. WIRED Italy. https://www.wired.it/attualita/politica/2017/06/19/venezuela-twitter-maduro/. Retrieved on 07/14/2017.

Daniel, F. (2017). Venezuela's Maduro slams Twitter after accounts blocked. Reuters. https://www.reuters.com/article/us-venezuela-politics-twitter-idUSKBN1980U1. Retrieved 07/14/2017.

Etling, B., Internet and Democracy, Kelly, J.(2008). Mapping Iran’s Online Public: Politics and Culture in the Persian Blogosphere. Berkman Klein Center for Internet & Society at Harvard University. URL: https://cyber.harvard.edu/publications/2008/Mapping_Irans_Online_Public. Retrieved on 04/15/2017.

Etling, B., Faris, R., Palfrey, J., Gasser, U., Kelly, J., Alexanyan, K., Barash, V. (2012). Mapping Russian Twitter. Berkman Klein Center for Internet & Society at Harvard University. URL: https://cyber.harvard.edu/publications/2012/mapping_russian_twitter. Retrieved on 04/15/2017.

Ferrara, E. Varol, O. Davis, C. Menczer, C. Flammini, A. (2016). The Rise of Social Bots. Communications of the ACM, Vol. 59 No. 7, Pages 96-104

Ferrara, E (2017). Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election. Page 1. URL: https://ssrn.com/abstract=2995809. Retrieved on 07/19/2017.

Fruchterman, T.M.J. & Reingold, E.M.(1991). Graph drawing by force-directed placement Authors. Journal of Software: Practice and Experience. First Published: November 1991. DOI: 10.1002/spe.4380211102

GlobalVoices(2017): What is happening in Venezuela: https://globalvoices.org/specialcoverage/what-is-happening-in-venezuela/. Retrieved on 07/13/2017

Gu, L., Kropotov, V., & Yarochkin, F. (2017). The Fake News Machine How Propagandists Abuse the Internet and Manipulate the Public. Trend Micro. Pages 52-56. URL: https://documents.trendmicro.com/assets/white_papers/wp-fake-news-machine-how-propagandists-abuse-the-internet.pdf. Retrieved on 07/19/2017.

Hoffman P., Lambon R. M.A., Rogers T.T. (2013). Semantic diversity: a measure of semantic ambiguity based on variability in the contextual usage of words. Behav Res Methods. 2013 Sep;45(3):718-30. doi: 10.3758/s13428-012-0278-x.

Kelly, J. Toward Cybersocial Geography: Meeting the Analytic Challenge of Social Media (2014). In: Human Geography: Socio-cultural Dynamics and Challenges to Global Security, Tomes R. R., Tucker, C. K. (eds.). United States Geospatial Intelligence Foundation, USA.

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., Beaton, M.(2013). Teens, Social Media, and Privacy. Pew Research Center: Internet, Science & Tech. URL: http://www.pewinternet.org/2013/05/21/teens-social-media-and-privacy/. Retrieved on 04/15/2017.

McPherson, M., Smith-Lovin, L. Cook, J. M. (2003). Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology. 27 (1): 415–444. doi:10.1146/annurev.soc.27.1.415

Prieto, H. (2017). Chaos Looms Over Venezuela. New York Times: https://www.nytimes.com/2017/05/03/opinion/chaos-in-venezuela-nicols-maduro.html. Retrieved on: 07/13/2017

Rid, T. Disinformation: a Primer in Russian Active Measures and Influence Campaigns. (2017) Hearings before the Select Committee on Intelligence, United States Senate, One Hundred Fifteenth Congress. 30 March, 2017. URL: https://www.intelligence.senate.gov/sites/default/files/documents/os-trid-033017.pdf. Retrieved on 04/15/2017.

Salvador, S. & Chan, P. (2004). FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space. KDD Workshop on Mining Temporal and Sequential Data, pp. 70-80, 2004

Sanovich, S., Stukal, D., Penfold-Brown, D., Tucker, J.A. (2015). Turning the Virtual Tables: Government Strategies for Addressing Online Opposition with an Application to Russia. In Proceedings of 9th Annual Conference of The International Society for New Institutional Economics, Cambridge, MA, USA, 2015.

Schneier, B. (2015). Data And Goliath : the Hidden Battles to Collect Your Data and Control Your World. New York, N.Y. :W.W. Norton & Company, 2015. Print.

Shannon, C.E., Weaver, W. (1949) The Mathematical Theory of Communication, Univ of Illinois Press.

Thomas, K., Grier, C. & Paxson, V. (2012). Adapting social spam infrastructure for political censorship. In LEET'12 Proceedings of the 5th USENIX conference on Large-Scale Exploits and Emergent Threats.

Toska, S.(2014) Protest Mobilization in the Age of Twitter, in The Arab Uprisings: Catalysts, Dynamics, and Trajectories, edited by Al-Sumait, F., Lenze, N., Hudson, M. C. Rowman & Littlefield Publishers.

Trending Topics. Twitter Support. (2017). URL: https://support.twitter.com/articles/101125. Retrieved on 04/15/2017.

Weedon, J., Nuland W. & Stamos A. (2017). Information Operations and Facebook. Facebook. https://m.facebook.com/notes/facebook-security/making-facebook-safe-and-secure-for-authentic-communication/10154362152760766/. Retrieved on 07/17/2017.

Wikipedia: Popular Will (2017). https://en.wikipedia.org/wiki/Popular_Will. Retrieved on 07/14/2017.

Woolley, S. Pre-print.

Woolley, S., & Howard, P. (2016). Automation, Algorithms, and Politics| Political Communication, Computational Propaganda, and Autonomous Agents — Introduction. International Journal Of Communication, 10, 9. Retrieved from http://ijoc.org/index.php/ijoc/article/view/6298/1809

Zuñiga, M. & Miroff, N.

The authors would also like to thank the participants of the Data & Society Propaganda & Media Manipulation workshop, who offered unvaluable feedback and perspectives on an earlier draft of this paper.