

Descriptive Statistics - POIR 613 Final Project

Bijean Ghafouri

11/11/2021

Introduction

Project

In this report, I show preliminary results for my project on the spread of online political extremism. My goal is to demonstrate the efficiency level of extremist persuasion on Twitter. Specifically, I seek to understand the likelihood of *adopting* extremist behavior. What does it take to adopt extremist political behavior? How does it compare to moderate behavior? How much persuasion does a user need in order to adopt extremist views?

Data

To answer these questions, I use a dataset published by Chen, Deb and Ferrara (2021) which includes all tweets related to the 2020 US Presidential election from early 2020 until the summer of 2021. These tweets were collected through keyword search using Twitter's API. From this dataset, I sample all tweets posted in June 2020. This represents a total of 67,181,551 tweets in English related to the election.

Complex contagion

Modeling political behavior with complex contagion is suitable to understand the conditions under which agents adopt extremist behaviors. Complex contagion describes a learning process in which agents require *social reinforcement* to adopt behavior (Centola and Macy 2007). In other words, an individual adopts a behavior after multiple individuals in their social network already adopted this behavior. This process follows a logic of herding, where an agent requires other agents to approve the behavior to gain legitimacy.

Using complex contagion, I demonstrate that extremist behavior adoption requires less exposure relative to other types of political behavior. I expect that when holding exposure constant, Twitter users are more likely to adopt extremist rather than moderate behavior.

An important assumption required under complex contagion is that behavior adoption is *risky* (Fink et al. 2016). The reason why an agent requires social reinforcement is because of the relative risk. The assumption of risk presents a puzzle for my theory. My theory posits that extremist content is more persuasive than moderate content. To confirm this expectation, online users should require less exposure to adopt extremist behavior compared to moderate behavior.

However, my empirical prior is that extremist adoption is *less* risky than adopting moderate behavior. Theoretically, we should expect extremist behavior to be *more* risky to adopt than moderate behavior. Although this theoretical expectation is sound, it is not necessarily at odds with my argument.

I argue that the process of adopting political behavior is done incrementally, reducing the risk of adopting extremist behavior. A moderate user will not suddenly adopt extremist behavior. Moderates will progressively

adopt slightly more extreme behaviors through time. Following my empirical expectation, this progression towards the extreme should be more likely than the progression towards moderate politics. With this logic, the lower risk associated with extremist behavior adoption is theoretically sound. It is true that suddenly adopting extremist behavior is risky. However, when an agent slowly progresses towards extremism, they are not conscious of the level of risk, thus allowing for persuasion effects.

Exposure and adoption

The literature on contagion effects offer multiple measures of behavior adoption, often limited by the nature of the data. Researchers will measure at the time of adoption how many users followed had also adopted the behavior. We may consider this an implicit measure of complex contagion, where we assume exposure. We cannot be certain that users have witnessed their followers' behavior. Prior exposure is thus not necessarily internalized by the user. For example, a user might follow 3 users who adopted the behavior before adopting himself, but never actually saw them express the behavior.

In this project, I use an explicit measure of complex contagion. As with the implicit measure, we know how many users followed adopted the behavior. Here, however, we also know that the user was directly exposed to the behavior, since they engaged with that behavior online. For example, if a user retweets, replies or quotes a tweet expressing the behavior, we consider this to be an explicit exposure.

Following Fink et al. (2016), I use hashtags as a measure of adoption. My goal is to measure how many exposures to #qanon is required *before* a user authors a tweet using #qanon. This hashtag is meant to conceptualize extremist behavior. I consider users who use #qanon in a tweet as adopting extremist political behavior. Furthermore, I consider users exposed to #qanon when they retweet, quote tweet or reply to a tweet using #qanon.

To model complex contagion, I count how many times a user was exposed to #qanon before using the hashtag. Out of the 67,181,551 tweets posted in June 2020 related to the US Presidential election, 351,108 tweets contained #qanon, authored from 83,867 unique users. Out of these tweets, only 4024 tweets were original, meaning that 4024 users used #qanon in an original tweet and adopted extremist behavior. I filter these tweets for users with ideal points larger than 0, which denotes Republicans (Barbera 2015).

Out of these 4024 users, only 1624 users were exposed to #qanon prior to adopting the behavior. This means that 2400 users in the sample used #qanon without prior exposure.

Among users who required prior exposure to adopt extremist behavior (1624 users), we observe an average of 5.467 exposures prior to behavior adoption. Among the entire sample of 4024 users, the average is 2.206 exposures prior to behavior adoption. In other words, Twitter users who authored a tweet containing #qanon were exposed to tweets authored by other users they saw on their timeline about 2.2 times before they authored that tweet.

It is important to mention that I do not measure exposures to #qanon after the first time a user adopts the behavior. I mark a difference between behavior adoption and extremist expression. That is, adoption measures the first time a user tweets with #qanon. Extremist expression measures the number of times a user tweets #qanon. It is possible that a single exposure enables a user to initially adopt the behavior, while they subsequently require far more exposures to express political extremism frequently. On the contrary, it is possible for a user to require multiple exposures to initially adopt the behavior, but subsequently use extremist hashtag frequently without exposures. Although this may be interesting to measure, I do not capture these possibilities in this report.

Results

My hypothesis is that extremist hashtags are more likely to adopt than moderate hashtags. To confirm my expectation, users should require less exposures to extremist hashtags before adopting extremist behavior relative to moderate hashtags.

First, let's take a look at some descriptive statistics about the data on Republican hashtags I collected. This table presents different information. The first column represents the number tweets used the hashtag. Surprisingly, the extremist hashtags #qanon, #obamagate and #wwglwga were used by Republican Twitter users significantly more than the moderate hashtag #trump2020.

Let's look at the #trump2020 more in detail. Out of the 12,506 users who engaged in the hashtag (original, reply, retweet or quoted), 3849 users authored a tweet with the hashtag (original and reply). This means that 30.8% of all users who engaged with #trump2020 adopted the behavior. In the fifth column, we see the number of users who required prior exposure to #trump2020 before adopting the behavior. Out of all users who authored a tweet with #trump2020, this represents only 3.95%. In other words, 96.1% of users who used #trump2020 were never previously exposed to the tweet.

To understand persuasion effects, we need to look into the behavior of these 3.95% of users who required exposure. These users will tell us more on what is required for them to use #trump2020. The 'mean exposures' column indicates the average number of exposures users needed before tweeting #trump2020. This excludes sole adopters. We see that users require on average 1 exposure before adopting the behavior. In other words, users will author a tweet, either in the form of an original tweet or reply, after retweeting or quoting a tweet with #trump2020 only once.

The mean number of exposures before using #trump2020, a behavior I consider to be moderately Republican, is low. To confirm my hypothesis, the number of exposures before adoption for the other extremist hashtags should be less than 1.

This is not the case. In fact, the mean number of exposures for extremist hashtags varies between 3 and 5. I exclude #stopthesteal before of the low sample size. I show that extremist hashtags require significantly more persuasion attempts to adopt the behavior than a moderate hashtag.

We also see that the mean ideal point for users who engaged with #trump2020 is lower relative to extremist hashtags. In others, users who engage with #trump2020 are more moderate than users using extremist hashtags. This confirms my expectation of the nature of these hashtags: authoring a tweet with #trump2020 is an indication of moderate behavior, while authoring a tweet with what I identify as extremist hashtags is an indication of extremist behavior. We observe a similar tendency with users who required exposure before adopting and users who did not require exposure before adopting. Users who authored a tweet with #trump2020, who authored a tweet with #trump2020 after being exposed and who authored a tweet with #trump2020 without prior exposure are more moderate than other users.

```
xtable(republican_descriptives)
```

```
## % latex table generated in R 4.0.3 by xtable 1.8-4 package
## % Fri Nov 26 13:43:02 2021
## \begin{table}[ht]
## \centering
## \begin{tabular}{rllrrrrrrrrrr}
## \hline
## & hashtag & n & unique users & adopters & adopters exposed & sole adopters & mean exposures & mean
## \hline
## 1 & trump2020 & 20490.00 & 12506.00 & 3849.00 & 152.00 & 3697.00 & 1.00 & 2.69 & 2.40 & 2.59 & 2.39
## 2 & qanon & 223754.00 & 59489.00 & 5100.00 & 1558.00 & 3542.00 & 4.00 & 2.90 & 2.64 & 2.78 & 2.57
## 3 & obamagate & 190058.00 & 60834.00 & 6901.00 & 1511.00 & 5390.00 & 3.00 & 2.93 & 2.75 & 2.87 & 2.7
## 4 & wwglwga & 359369.00 & 79835.00 & 10934.00 & 3725.00 & 7209.00 & 5.00 & 2.92 & 2.78 & 2.87 & 2.7
## 5 & whitelivesmatter & 9268.00 & 396.00 & 108.00 & 4.00 & 104.00 & 5.00 & 1.76 & 2.21 & 2.55 & 2.2
## 6 & stopthesteal & 1168.00 & 743.00 & 29.00 & 5.00 & 24.00 & 1.00 & 2.87 & 2.76 & 2.93 & 2.73 \\
## \hline
## \end{tabular}
## \end{table}
```

	hashtag	n	unique users	adopters	adopters exposed	sole adopters	mean exposures	mean
1	trump2020	20490.00	12506.00	3849.00	152.00	3697.00	1.00	
2	qanon	223754.00	59489.00	5100.00	1558.00	3542.00	4.00	
3	obamagate	190058.00	60834.00	6901.00	1511.00	5390.00	3.00	
4	wwg1wga	359369.00	79835.00	10934.00	3725.00	7209.00	5.00	
5	whitelivesmatter	9268.00	396.00	108.00	4.00	104.00	5.00	
6	stopthesteal	1168.00	743.00	29.00	5.00	24.00	1.00	

Next, let's look at the Democrat hashtags.

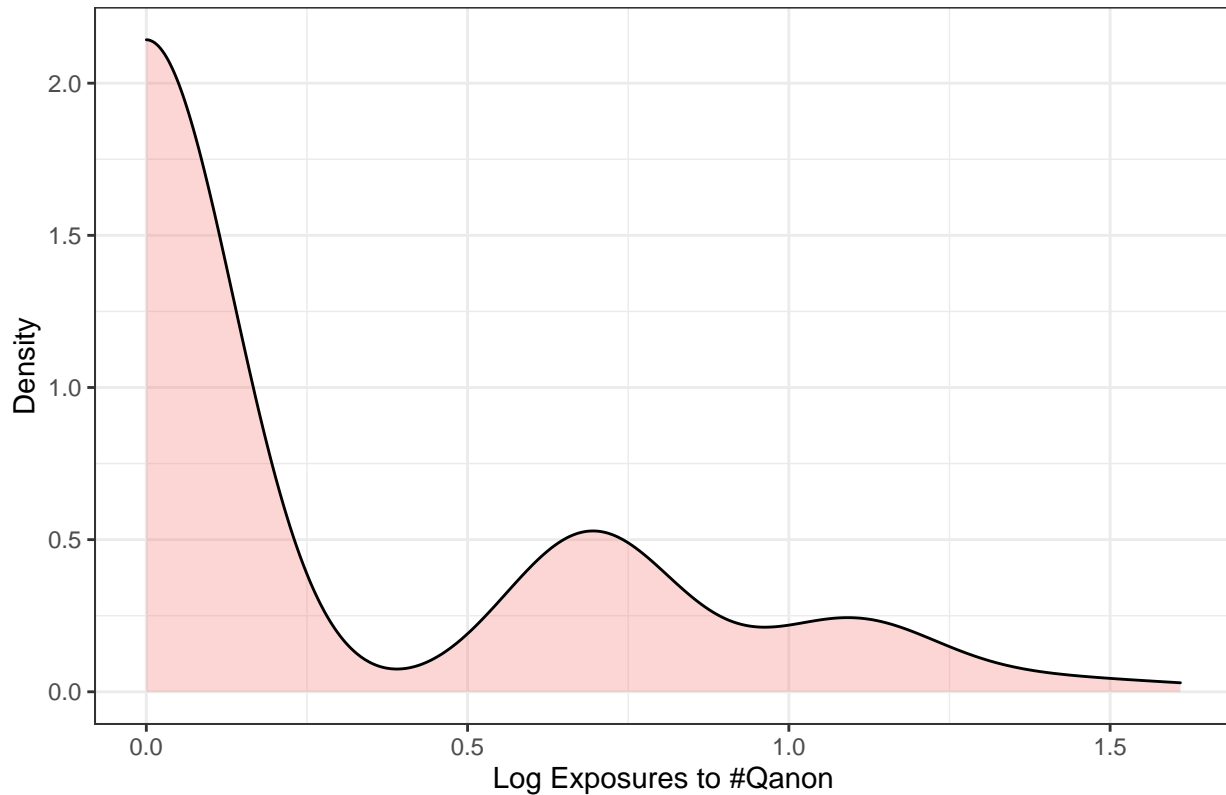
The number of exposed adopters is the key statistic to appreciate the persuasiveness of extremist behavior. This statistic computes the number of users who need to be persuaded *and* got persuaded. I assume they needed persuasion because they adopted only after being exposed to the behavior. The assumptions lies in that users needed to see the behavior on Twitter before adopting. Otherwise they would have adopted without requiring exposure.

Data visualization

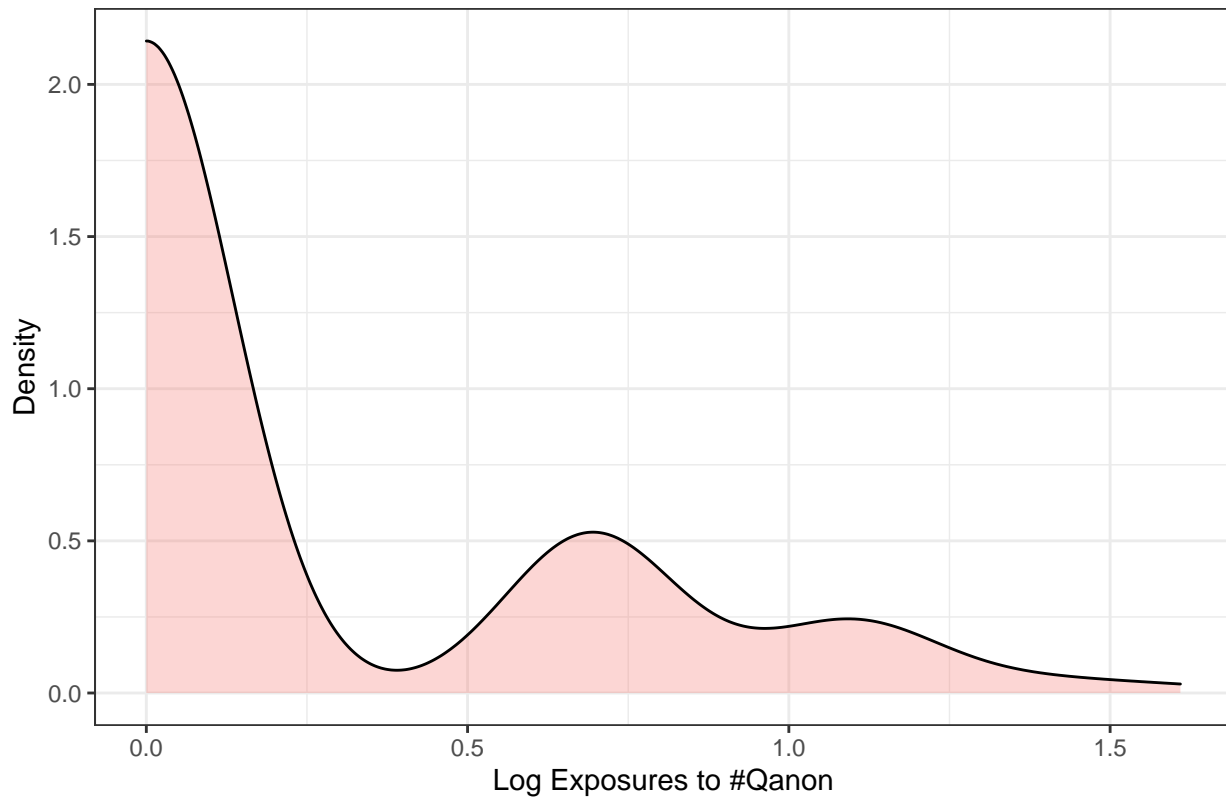
Here, I show some density plots that demonstrate my preliminary results.

We can plot the density distribution of the number of exposures before adoption. This includes users who require 0 exposures, which represent 60% of the sample.

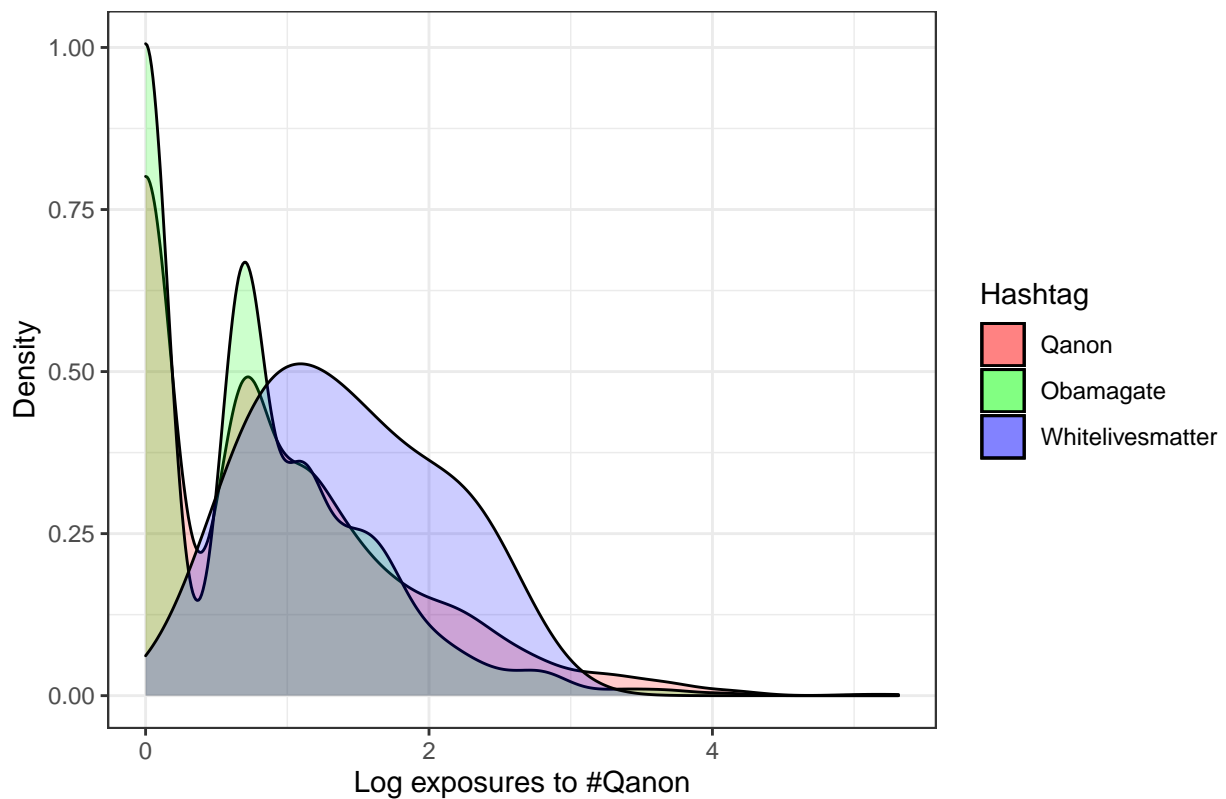
Density plot of Tweet exposures before adopting #qanon



Density plot of Tweet exposures before adopting #qanon



Density plot of Tweet exposures before adopting hashtag



Next, let's consider users who required at least one exposure before using #qanon.

We notice that only a few users require more than 100 exposures to adopt the behavior. Let's exclude these users from our sample, considering users who required less than 100 exposures.

Finally, let us consider the same sample, but with users who again required at least one exposure before adoption.