

Descriptive Statistics - POIR 613 Final Project

Bijean Ghafouri

11/11/2021

Introduction

Data

First, I describe the data I use and how I prepared it. 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 data, I sample all tweets posted in June 2020. In total, this represents 67,181,551 tweets related to the election (in English).

Project

In this report, I show preliminary results on my project related to the spread of political extremism online. In this project, I describe the efficiency of extremist persuasion on Twitter. To describe the persuasion of online extremism, I characterize the context of *adopting* extremist behavior. What does it take to adopt extremist political behavior? How does it compare to moderate behavior? Do online users need to be more or less persuaded in order to express extremism?

Complex contagion

Modeling political behavior with complex contagion is suitable to answer these questions. Complex contagion describes a learning process in which an agent requires *social reinforcement* to adopt a 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 hearing, where an agent requires other agents to approve of the behavior in order to legitimize and adopt it.

An important assumption required under complex contagion models is that the adoption of behavior is *risky* (Fink et al. 2016). Given the risk associated to adoption, the agent requires social reinforcement to adopt. 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, I expect online users to require less exposure to adopt extremist content compared to moderate content.

The assumption of risk in complex contagion thus presents a puzzle when applied in my theory. My empirical expectation that extremist adoption is *less* risky than adopting moderate behavior is indeed counterintuitive. Theoretically, we should expect extremist behavior to be considered as more *risky* to adopt relative to moderate behavior. I argue that extremist behavior is more persuasive than moderate behavior, reducing the risk associated to adoption. Moreover, the process of adopting political behavior is done incrementally. A moderate user will not suddenly adopt extremist behavior. Moderates will progressively adopt more extreme behaviors through time. This progression towards the extreme, I argue, is more likely than the

progression towards moderate politics. With this logic, the lower risk associated to extremist behavior adoption is theoretically sound.

Empirically, I demonstrate using complex contagion that extremist behavior requires less exposure to adopt relative to other types of political behavior. My hypothesis is that holding exposure constant, users on Twitter are more likely to adopt extremist behavior than moderate behavior.

Exposure and 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 tweets as adopting extremist political behavior. Furthermore, I consider users exposed to #qanon when they retweet, quote tweet or reply to a tweet using #qanon.

Past studies model complex contagion differently (Fink et al. 2016). 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 exposure is assumed. Prior exposure is not necessarily internalized by the user, since they do not necessarily see their followers' behavior. For example, a user might follow 3 users who used #qanon before using the hashtag himself, but that user never saw these individuals use that hashtag.

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 using #qanon, we consider this to be an explicit exposure.

To model complex contagion, I count how many times a user was exposed to #qanon behavior 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 adopted our measure of extremist behavior.

Out of these 4024 users who authored at least one tweet with #qanon, 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.

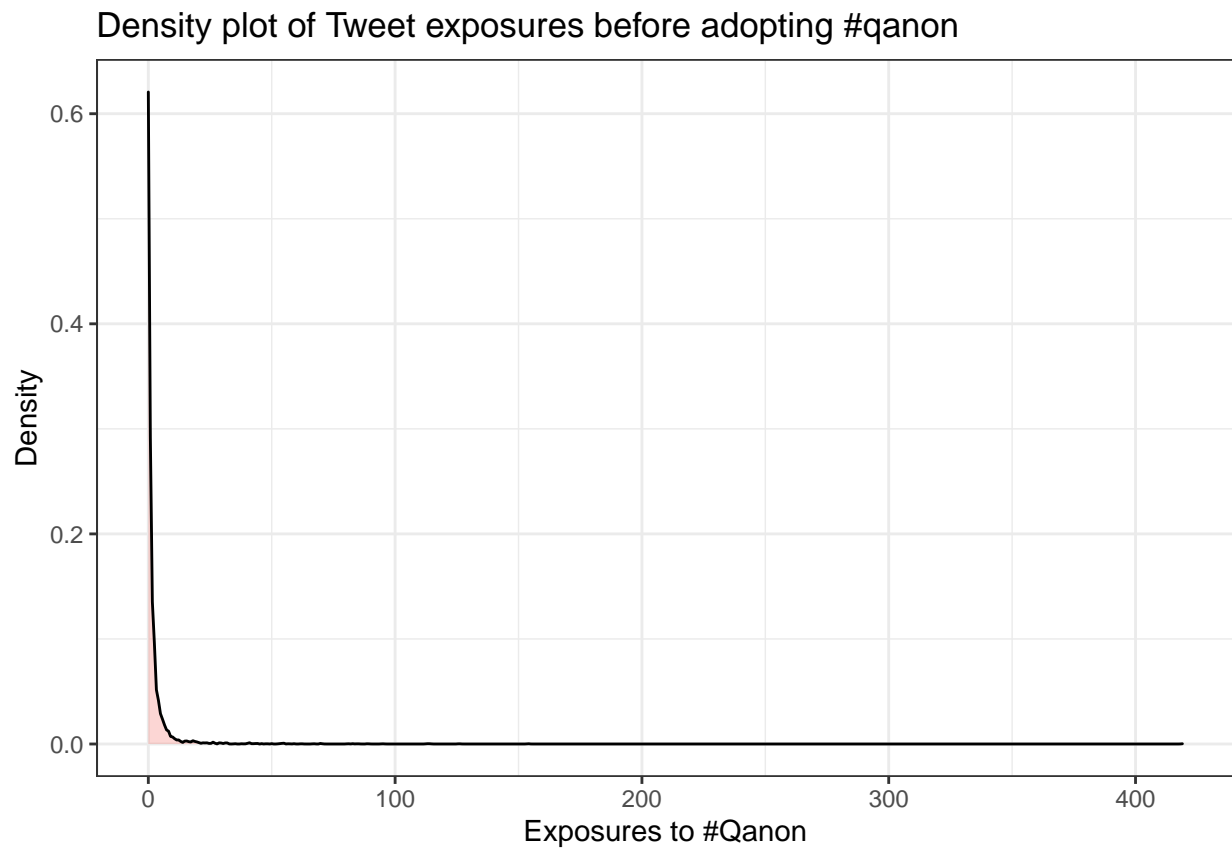
Among users who required prior exposure to adopt extremist behavior (1624 users), we observe an average of 5.467 exposures prior to behavior adoption. Among the entire sample of 4024 users, the average is 2.206 exposures prior to behavior adoption. In other words, Twitter users who authored a tweet containing #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.

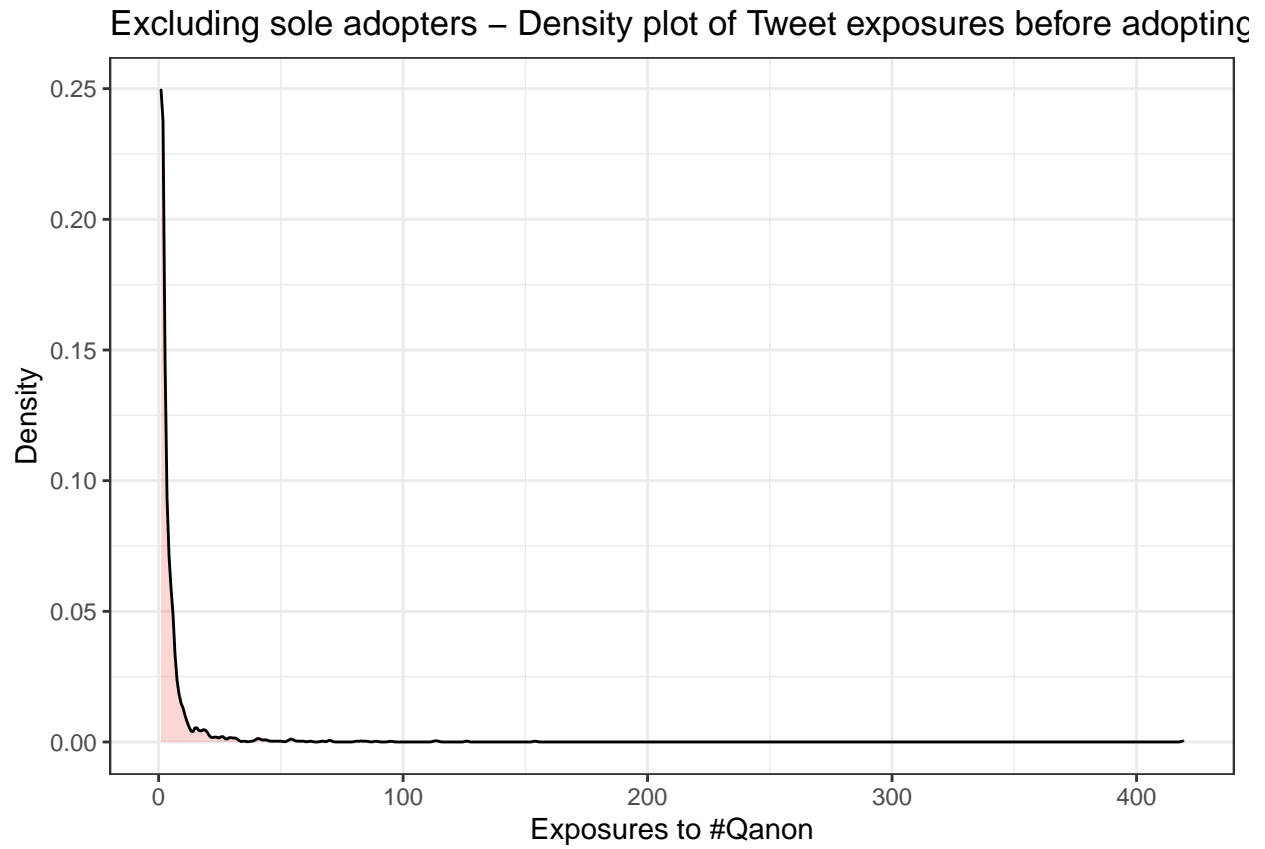
Data visualization

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

First, 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.

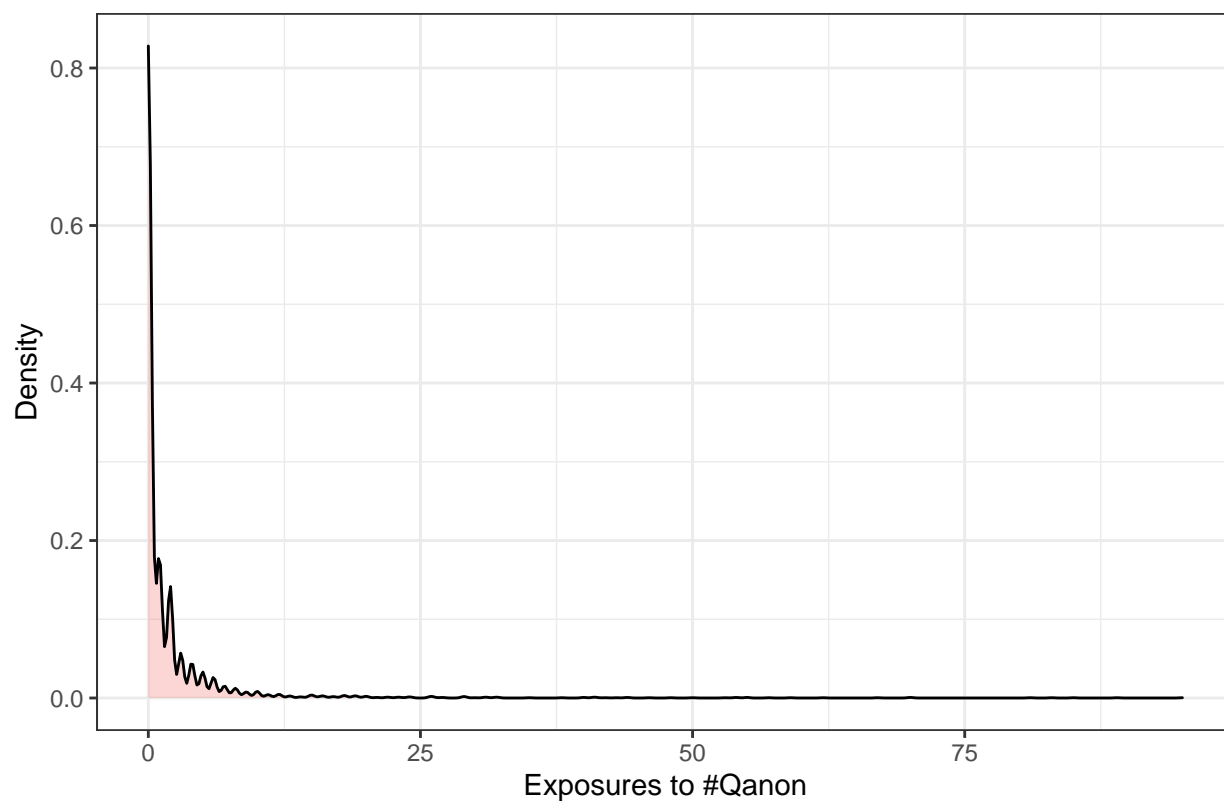


Next, let's consider users who required at least one exposure before using #qanon.



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.

Reduced: Density plot of Tweet exposures before adopting #qanon



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

Reduced: Excluding sole adopters – Density plot of Tweet exposures before

