



# Misinformation in the Fog of Tear Gas

A Play by Play of How the #DCblackout  
Hashtag Spread, and Spread Fear

Ben Nimmo and C. Shawn Eib, Graphika

**07.2020**

**Misinformation**

# Misinformation in the Fog of Tear Gas

## A Play by Play of How the #DCblackout Hashtag Spread, and Spread Fear

*Ben Nimmo and C. Shawn Eib, Graphika*

In the early hours of Monday, June 1, 2020, rumors began spreading on social media that electronic communications from Washington, D.C., were being jammed and that a crackdown on the night's [anti-racism protests](#) was imminent. The posts coalesced around the hashtag #DCblackout.

When morning came, telecoms operators and journalists on the ground confirmed that no such "blackout" had occurred. Nevertheless, arguments over the hashtag continued throughout the day. Conflicting reports emerged that "bots" (automated accounts) were amplifying either the [hashtag](#) or the [rebuttals](#). Twitter [confirmed](#) that it had suspended "hundreds of spammy accounts" active on the hashtag. A Twitter account with over 2 million followers associated with the "Anonymous" hacktivist group [argued](#) that it was a "misinformation campaign" designed to "instill panic & fear and deter future protests."

The reality was more mundane. The hashtag appears to have emerged as a spontaneous expression of fear and confusion in the middle of the night. It picked up traction and began to trend because it filled a void in the information space - a void largely created by the prosaic fact that verified sources on the ground had gone to bed. It went viral in the morning because the Anonymous account tweeted about it - first amplifying claims of the blackout, then setting the record straight. The fog of tear gas gave way to a fog of confusion that was only dispelled with the day.

The story of #DCblackout is a lesson in the importance of verified information from accountable sources, especially in times of fast-moving events. It also highlights the responsibility that major influencers have to check their facts before they amplify dramatic claims: the Anonymous account appears to have acted in good faith and to have corrected its error to the best of its ability, but it played the key role in making the hashtag trend.

Finally, it is a reminder of the nuances involved in calling out “bots.” Automated and semi-automated accounts do swarm around many hashtags, either because they have been programmed to focus on a key phrase or because they are programmed to capitalize on whatever trending topics arise, but identifying such accounts can be a complex endeavor, as organic behavior can sometimes appear anomalous and automation can take many forms. Many of the accounts initially classed as bots on June 1 were actually human-operated, as far as we have been able to establish. Insufficiently evidenced claims of fake accounts deepened the confusion.

This report attempts to tell the story of #DCblackout based on the available evidence as of June 4, 2020. As such, the data on Twitter traffic represented in the graphics do not include content from accounts suspended and posts deleted before that date. Where relevant in the text, we have provided references to known prior suspensions and deletions.

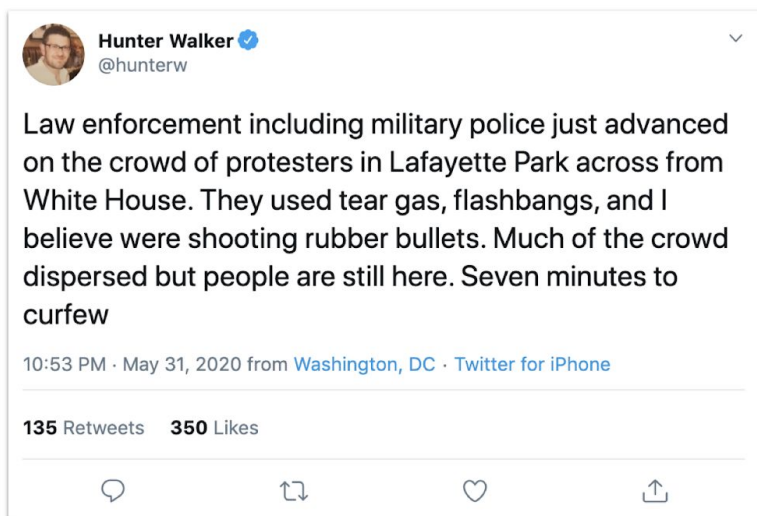
## Part 1: Fear and Confusion

To understand the explosion of #DCblackout on the morning of June 1, it is necessary to understand the series of events that preceded it. These were marked by fast-paced developments on the ground, reported in close to real time by verified journalists.

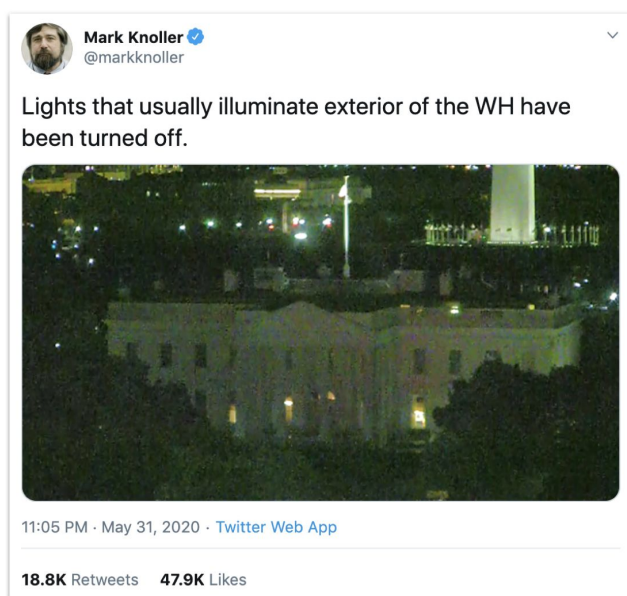
Late in the evening of May 31, protesters were gathered outside the White House. With a curfew due to take effect at 11 pm local time, law enforcement officials dispersed them using [tear gas](#) and [flashbangs](#). Reports began to circulate that the White House’s external lights had been shut off, as if this were a significant event.<sup>1</sup>

---

<sup>1</sup> Fact-checking site factcheck.org later [confirmed](#) that one of the images purporting to show the White House in darkness was a forgery. CNN’s White House correspondent Caitlan Collins subsequently [tweeted](#) that the White House lights always go off around 11pm. The significance given to the images of the White House in darkness indicate the tension and sense of drama in the reporting of the night’s events (a sense which the author of this paper has experienced first hand in covering other demonstrations).



Left, [tweet](#) by Yahoo News White House correspondent Hunter Walker reporting the police advance on protesters in Lafayette Park just before curfew. Right, [tweet](#) by journalist Samantha-Jo Roth showing a photo of a TV report of smoke rising in D.C., just after curfew. The latter post was retweeted some 58,000 times.



[Tweet](#) by Mark Knoller, CBS News White House correspondent, documenting the White House in darkness at 11:05 PM local time.

Following this outbreak, the situation fragmented, with small groups of protesters and rioters scattered through the city. Around 12:30 AM, Yahoo News' Hunter Walker tweeted a short but tense [thread](#) to say that he had spotted a "small group of people attacking a nearby building" and was "hiding in an alley" and [pinned down](#). He then chronicled his return home.

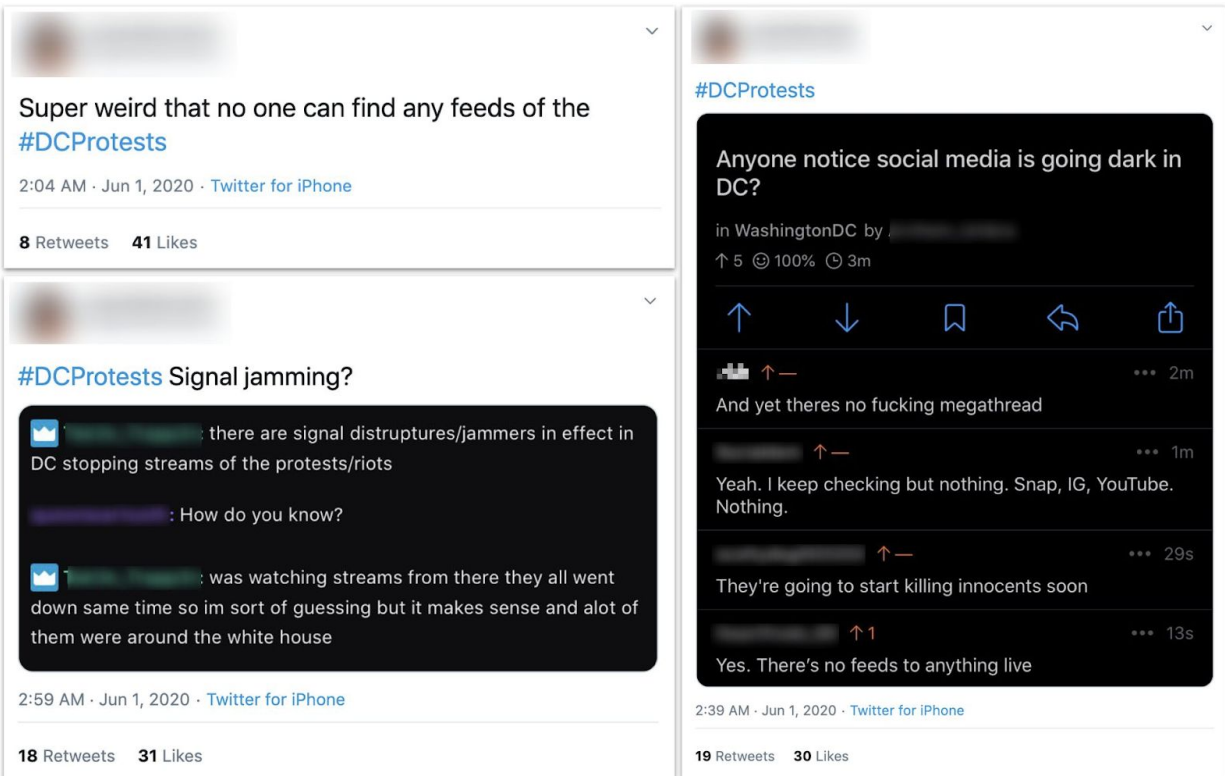


Left, [thread](#) and right, subsequent individual tweets, by @HunterW, chronicling the tension of his walk home. He posted his [final tweet](#) of the night at 1:53 AM. (Tweets read in the direction of the arrow.)

As the reporters—and the majority of protesters—went home, however, internet users in other parts of the country were still trying to see what was happening via live streams, Twitter feeds, and other sources on social media. Without authoritative voices providing updates on the situation on the ground, the conversation quickly became speculative and fearful. The primary tone was one of anxiety and tension, not the gleeful malice of trolling: these appear to have been a variety of users across the United States trying to work out what was going on from any online resources they could find.

The initial rumor was that live feeds of the protests had been cut off. Conversations started on platforms including Twitter, Reddit, and Twitch as users asked whether there were still live feeds available, and where they were.





Three tweets by a Texas-located account asking what was happening with the protests and sharing reports of possible “signal jamming.” Note that the tone is cautious and questioning, asking for a source (“How do you know?”) in the bottom left-hand tweet. The right-hand tweet provides a screenshot of a Reddit post that has since been deleted. The same Texas-based user subsequently tweeted links to a number of live feeds that were still visible as of 3 AM. In this report, we have obscured the names of users who are not verified or otherwise of major significance to this report, to protect their privacy.

It was at this time that a single Twitter account tweeted the first publicly available post on the hashtag #DCblackout. This account only had three followers at the time (it had five, including two journalists, as this report was being written), had been registered in May 2019, and largely posted content that criticized U.S. Democrats and the mainstream media, at an average rate of 2-3 tweets per day. Despite some [suggestions](#) in the aftermath of the events in Washington that it was a “fake” account, a review of its content suggests that it was run by an individual user who typically posted unique, authored tweets. It is also important to note that, although this was the first use of the hashtag, it barely received attention: it only received three retweets, and the next known use of the hashtag did not come for another hour. Thus, while this account used the hashtag first, it does not appear to have triggered the trend that ultimately followed.



*The first known tweet to use #DCblackout. The exact context of the tweet is unclear.*

A much more authoritative voice during this period was the Anonymous-affiliated account @YourAnonCentral, which [at that time](#) had some 2.4 million followers. The account had spent much of May 31 retweeting leaks from another anonymous account, @OpDeathEaters, but during the night, it also posted content about the protests that used mainstream hashtags, such as #Icantbreathe, #GeorgeFloyd, and #BlackLivesMatter.

Shortly before 3 AM Eastern time, the account tweeted two claims that had already circulated - one of the White House in darkness, and the other of D.C. "in flames." The former was later [exposed](#) as a photoshopped image; the latter was an image of a TV screen similar, but not identical, to the photo published by Samantha-Jo Roth a few hours before. These tweets yielded tens of thousands of retweets each. They used the mainstream hashtags described above, but did not, at this stage, mention #DCblackout.



*Two tweets by @YourAnonCentral shortly before 3 AM on June 1.*

However, the account also continued to retweet historic content from @OpDeathEaters, a pattern of activity that it continued until its final tweet of the night, at 4:56 AM.



*Last tweet of the night from @YourAnonCentral on @OpDeathEaters.*

This content is relevant to our report because it shows that @YourAnonCentral was actively posting at this time about the dramatic events unfolding in the capital but was also posting on unrelated content. The tweets also confirm that, at this time, @YourAnonCentral was using the hashtags associated with the protests (and thus would have been visible to users searching for information about them), and it was influential in terms of retweets.



That became increasingly important from around 3 AM onward, as online observers began to coalesce around the hashtag #DCblackout. The next two uses of the hashtag, from the Texas-based account mentioned earlier, retweeted Spanish-language content without any editorial comment and gained no significant traction, but at 3:26 AM a liberal-leaning account that had focused on the protests tweeted a dramatic warning that “They’re going to start killing and are trying to hide it via jammers,” and this tweet spread much more widely, with 210 retweets.

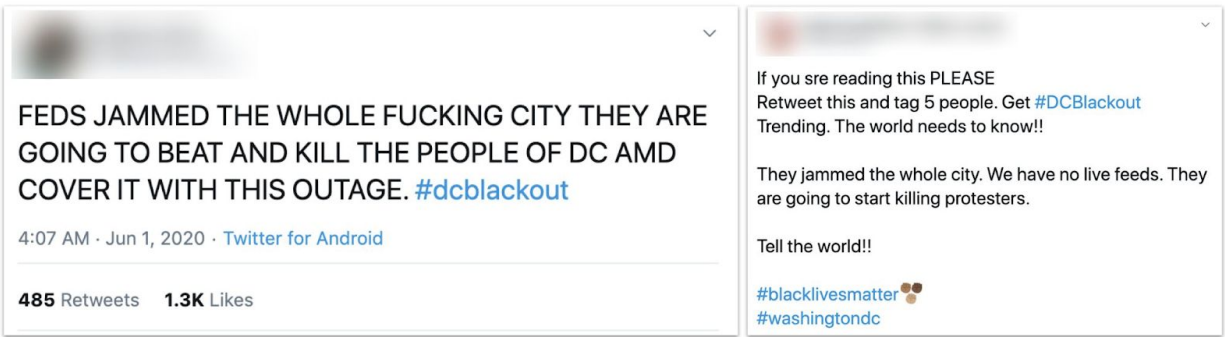


*The first tweet on #DCblackout to achieve significant traction, measured in likes and retweets.*

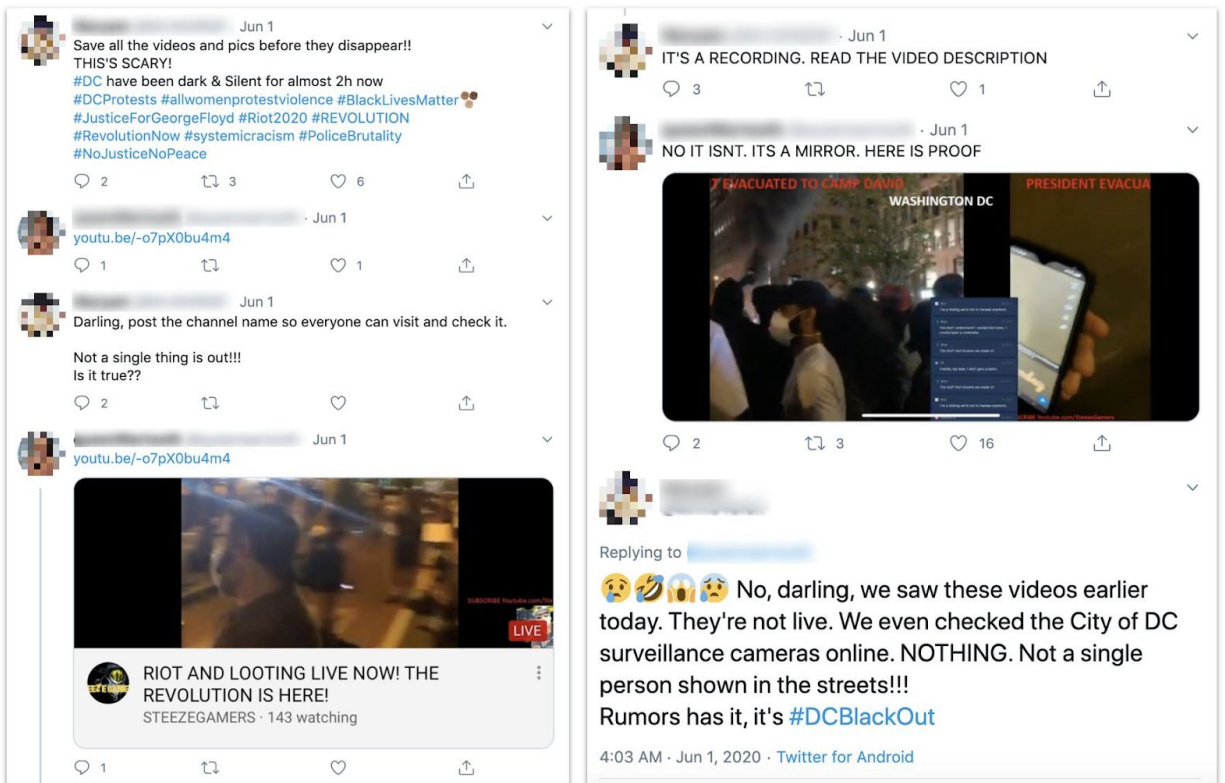
Over the next three hours, other users picked up on the hashtag and added their own impassioned warnings. Their activity did not resemble an influence operation designed to launch a narrative: rather, it was a series of claims, shoutouts, and arguments, as different users made their positions known in increasingly vocal and dramatic ways.



*Call to make the hashtag trend shortly before 4 AM on June 1. Note how the claim has moved from possible jamming, as in the tweets from the Texan account above, to the statement "they are jamming."*

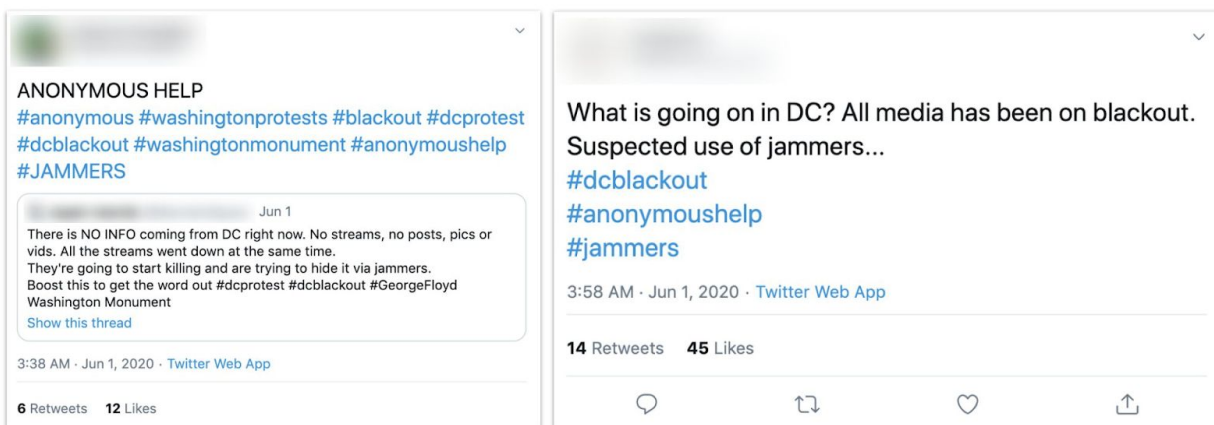


*More warnings of impending violence and calls for help in making the hashtag trend.*

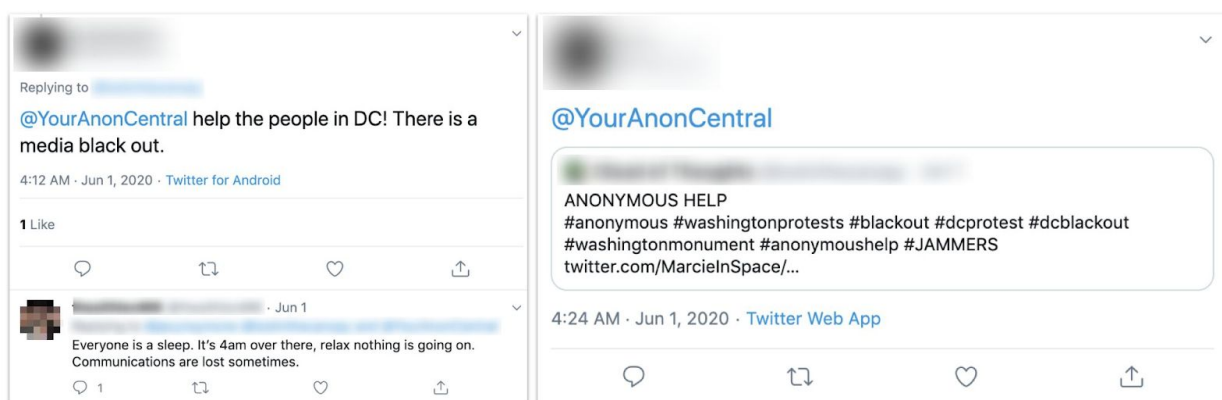


*Argument about the availability of live feeds from D.C. between the Texan account (which posted a YouTube link) and another user. This exchange exemplifies the atmosphere of fear, but also of the search for reliable information and wariness of possibly false information.*

Crucially for what was to follow, some of these users began calling for information from the Anonymous group using the hashtag #AnonymousHelp, and others directly tagged @YourAnonCentral with their requests for help.

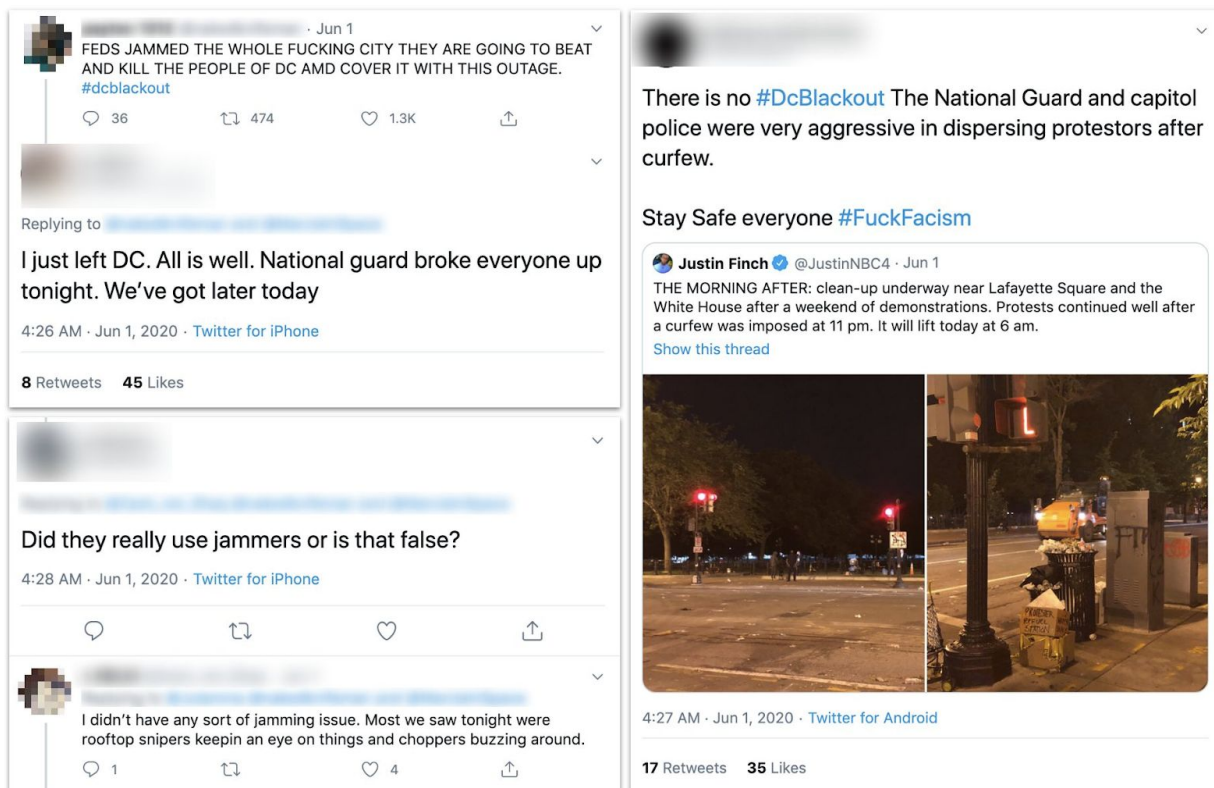


*First tweet and second tweet calling for #anonymousehelp, 3:30 to 4 AM.*



*Further tweets tagging @YourAnonCentral.*

Not all the content was alarmist. A number of replies to the most dramatic posts poured cold water on their fears, citing mainstream reporting and personal experience to point out that nothing sinister was happening in D.C. other than that people had gone to bed. These posts scored some positive reaction, but not enough to calm the sense of urgency and fear overall.



*Left, tweets discussing what happened. Right, reporting on the cleanup.*

Between 6 and 7 AM, new users began to come online. Some of them picked up on the #DCblackout hashtag and the sense of alarm that went with it and began posting their own content, asking what had happened in the night and urging still more users to make the hashtag trend (which suggests that it was not yet trending). The four tweets illustrated below were among them, and were soon to play an important role of their own.





*Tweets between 6 AM and 7:15 AM.*

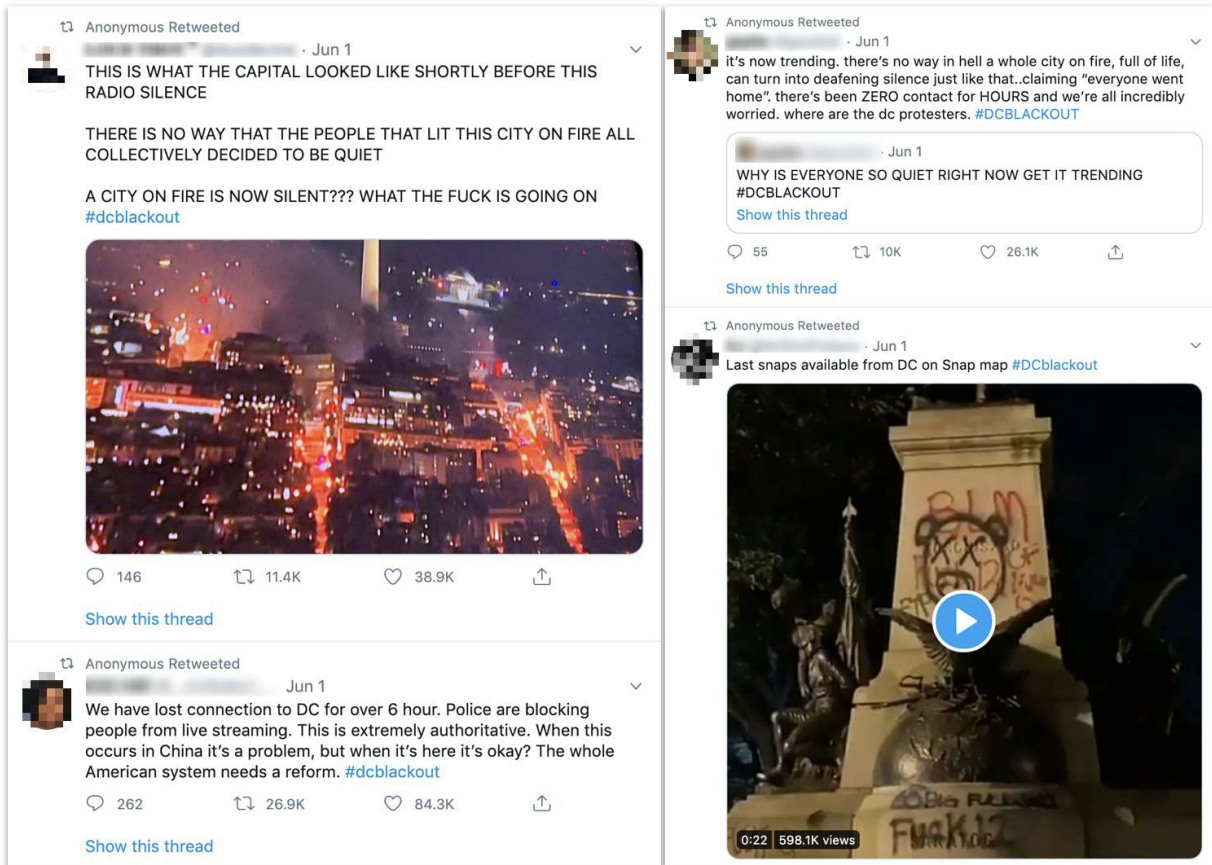
## Part 2: Enter the Influencers

At 7:30 AM, @YourAnonCentral, which had been quiet for two and a half hours, returned to posting with a [tweet](#) aimed at President Trump. In line with its posts from the night before, this account used three hashtags: #ICantBreathe, #GeorgeFloyd, and #BlackLivesMatter. The tweet was massively successful, generating over 100,000 retweets and almost half a million likes.



[Tweet](#) from @YourAnonCentral returning to the fray.

At 7:36 AM, however, @YourAnonCentral began retweeting almost a dozen tweets that featured the #DCblackout hashtag, including the four tweets illustrated above, without additional comment. The effect was instantaneous and remarkable. The first account that @YourAnonCentral retweeted had 126 followers: by the time Graphika viewed the tweet on June 5, it had received almost 27,000 retweets. The second account to be retweeted had 230 followers and received 11,800 retweets on the one post.

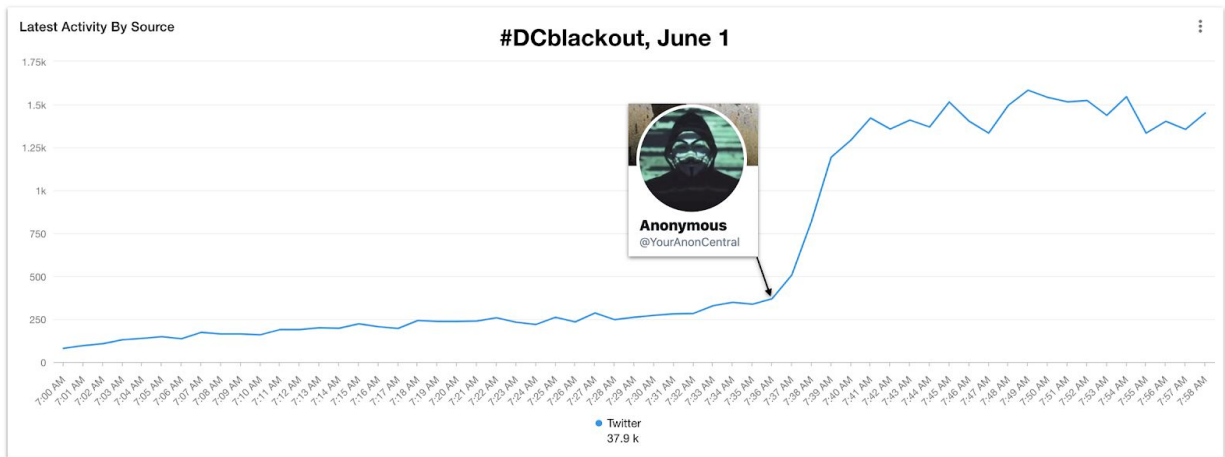


Retweets by @YourAnonCentral, starting at 07:36 local time, of the posts illustrated above.



The amplifier effect: profile of the first account Anonymous retweeted showing the number of followers, and the tweet with the number of retweets highlighted.

The impact of @YourAnonCentral and its 2.4 million followers left a glaring trace in the Twitter traffic. A minute-by-minute scan of tweets on the hashtag #DCblackout from 7 to 8 AM shows a precipitous surge in traffic from 7:36 AM, the minute at which @YourAnonCentral began retweeting posts on the hashtag.



*The immediate effect of @YourAnonCentral. Timeline of tweets from 7 to 8 AM, showing the point at 7:36 AM at which @YourAnonCentral began retweeting posts with the hashtag.*

We do not know whether @YourAnonCentral discovered the #DCblackout hashtag because of the accounts that had tagged it in the night or by searching for associated hashtags, but the data are unmistakable: the primary reason that #DCblackout made such an impact on the Twitter traffic on the morning of June 1, as D.C. was waking up, was that @YourAnonCentral made it trend.

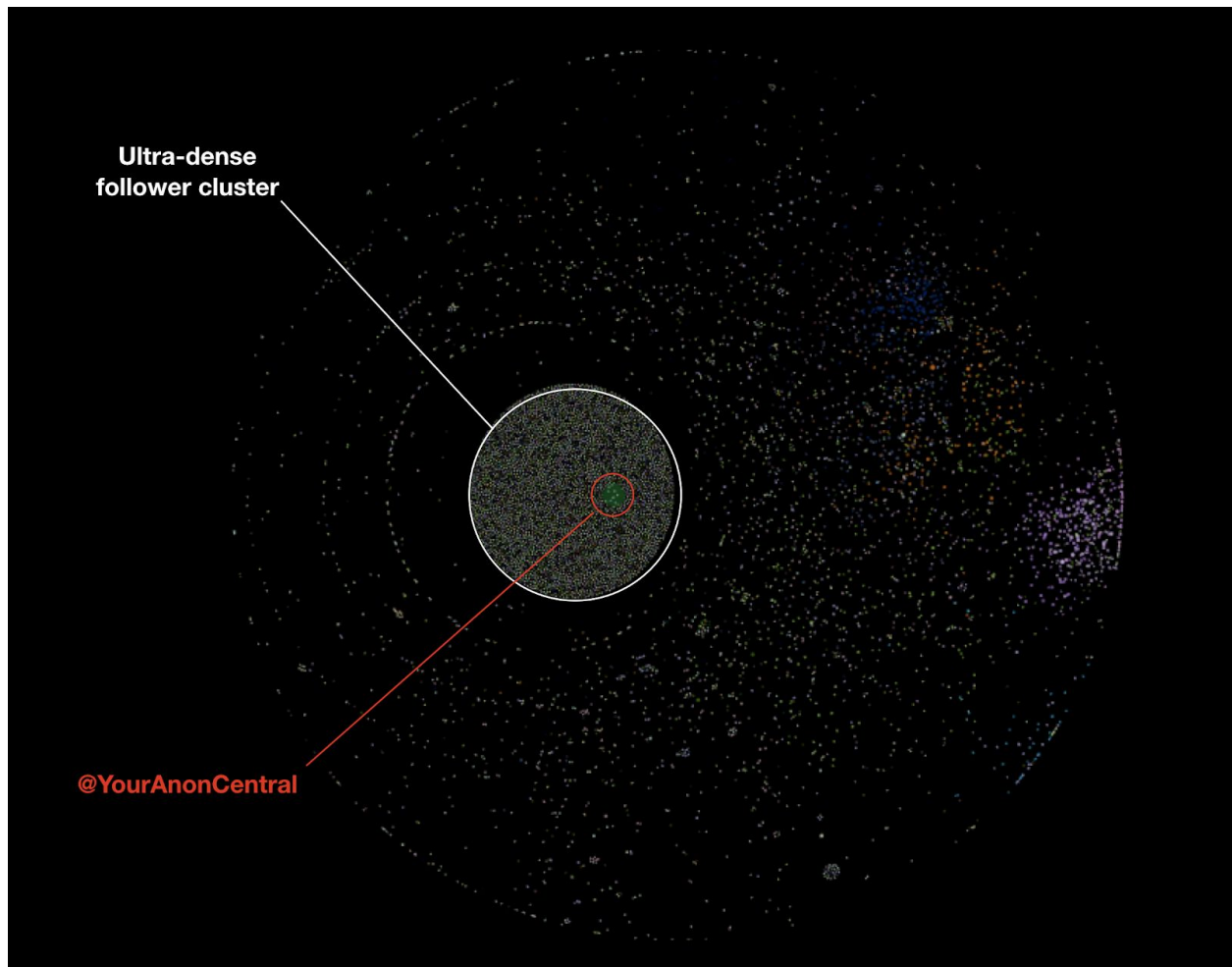
## Network Effect

There are several reasons @YourAnonCentral had such a cardinal impact on the hashtag. The first was simply the size of its following: at 2.4 million followers, it is a major influencer. The second was its preceding interest in the protests: this was a high-influence account that was heavily politically engaged, and therefore likely to draw the attention of engaged users who were equally focused on the protests.

A third reason is the unique structure of the Twitter conversation around #DCblackout. Graphika mapped the accounts that used the #DCblackout hashtag on June 1, and the result was a structurally striking map. Near the center of the map is a bright point that represents @YourAnonCentral; it is larger than the rest because it is followed by almost 6,000 other accounts in the map. Around it is a very dense cluster of accounts that follow @YourAnonCentral but do not follow each other or any of the other accounts in the map. Scattered on the right-hand side of the map are a diverse range of accounts from a number of different communities, including Black Lives Matter, Kpop and U.S. pop fandom, as well as journalists. These accounts follow one another to varying degrees, thus forming small separate clusters. The visualization of the map has a close resemblance to the Death Star from Star Wars.

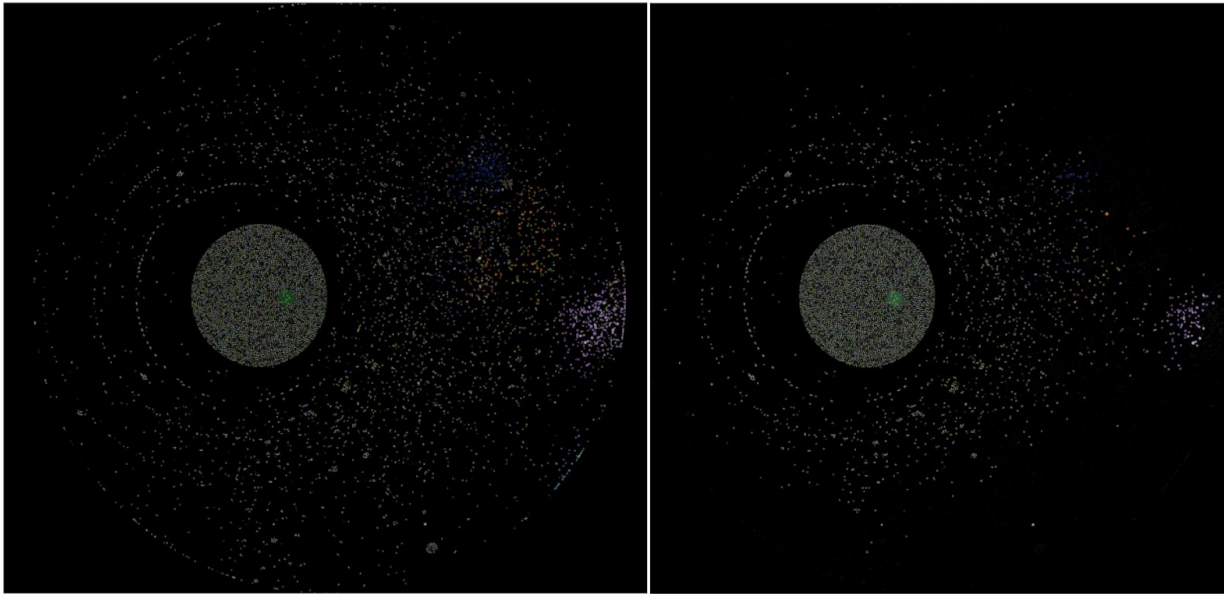
This structure illustrates the disproportionate influence that @YourAnonCentral had on the conversation around #DCblackout. Roughly 70 percent of the accounts in the map (that is, accounts that tweeted about #DCblackout) followed it. Many of them were not followed by any other account in the map, but they were followed by other accounts on Twitter. That meant that

each account was positioned to amplify tweets from @YourAnonCentral to a distinct Twitter community. In turn, that meant that any post from @YourAnonCentral, if retweeted by the core, was able to spread through many different communities at once, giving it even more influence on the conversation than its follower count (already substantial) might suggest.

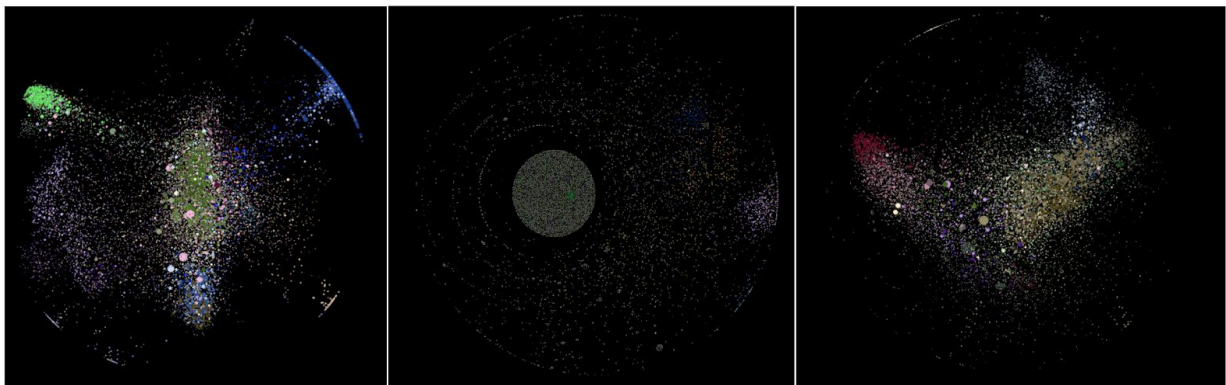


*That's no moon, it's a space station. Map of the Twitter accounts that posted about #DCblackout. The central ball represents a very large number of accounts that follow @YourAnonCentral, but do not follow, and are not followed by, other accounts in the map.*





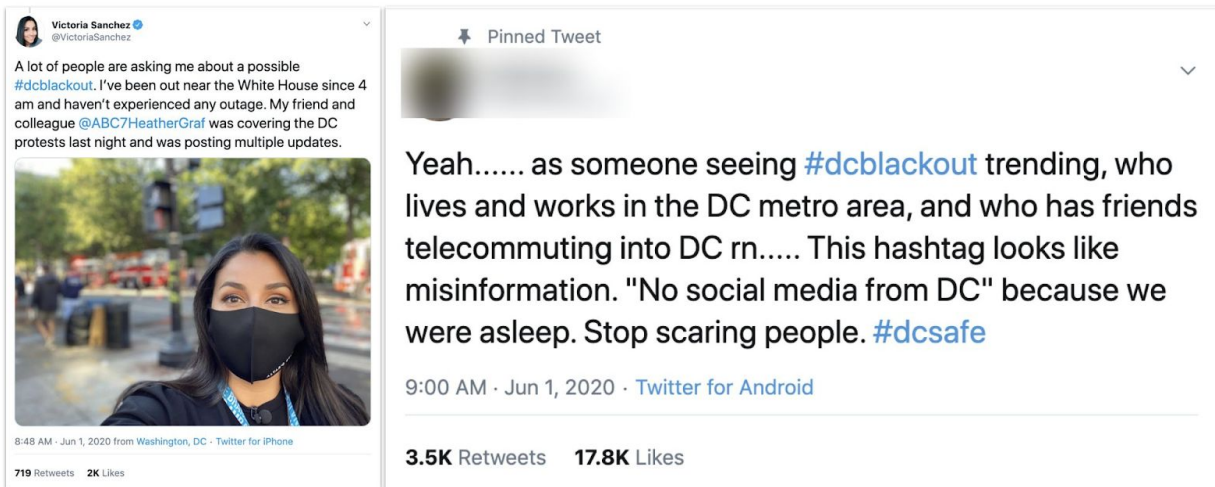
*Left, map of accounts that tweeted on #DCblackout. Right, the same map filtered to show only those accounts that tweeted the hashtag and followed @YourAnonCentral. This further illustrates how central the Anonymous account was to the conversation around the hashtag. In a map containing 8,479 accounts, 5,960 followed @YourAnonCentral.*



*To clarify the unusual nature of the Death Star map, this image compares it with contemporaneous Graphika maps of the conversations around the George Floyd protests in Denver (left) and Minneapolis (right). The flanking maps are much more typical in that they show distinct groups of accounts with differing affiliations, clustering in a number of different locations on the map.*

## Part 3: Clearing Up

@YourAnonCentral was not the only account to recommence posting at around that time. Journalists and ordinary users were also waking up and logging on, and some of these clearly saw the #DCblackout hashtag and decided to refute its central claim.



*Tweets by journalist [Victoria Sanchez](#) (left) and another user (right) pushing back on the hashtag as possible misinformation.*

One leading voice among them was Yahoo News' Hunter Walker, who tweeted advice to users in multiple languages asking them not to spread the #dcblackout hashtag, as it was false.



Multilingual [tweets](#) from Hunter Walker urging users not to share the #dcblackout hashtag.

Just before 8 AM, @YourAnonCentral tweeted out a call to protesters in D.C., asking them to use the hashtag #DCsafe and to share "if you have more information about what happened last night." At least one user replied within three minutes to say that "literally nothing happened." This was certainly a more understated attitude to the events of the night than many others managed, but the user also pointed out that they had had no communications issues since 5 AM. Crucially, this would have inserted a rebuttal into @YourAnonCentral's notifications just under half an hour after it launched the #DCblackout retweets.



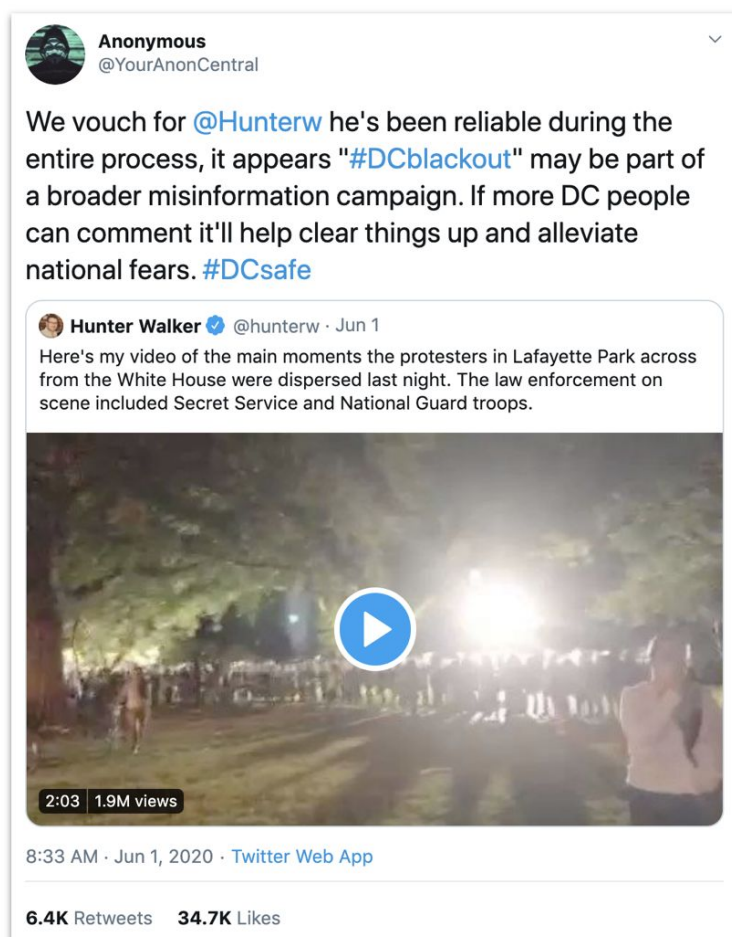
*[Call](#) and response from @YourAnonCentral.*

@YourAnonCentral's response was to reach out to Hunter Walker and ask him for a direct message conversation. Walker replied positively within less than ten minutes.



*Conversation between [@YourAnonCentral](#) and [Hunter Walker](#).*

@YourAnonCentral quickly followed up the conversation by quoting an earlier tweet from Hunter and acknowledging that #DCblackout “may be part of a broader misinformation campaign.” The account then proceeded to retweet a number of posts that demonstrated that there had been no blackout, including one from the verified account of network data analysis site netblocks.org that showed that data and cellular connectivity had remained stable throughout the night.



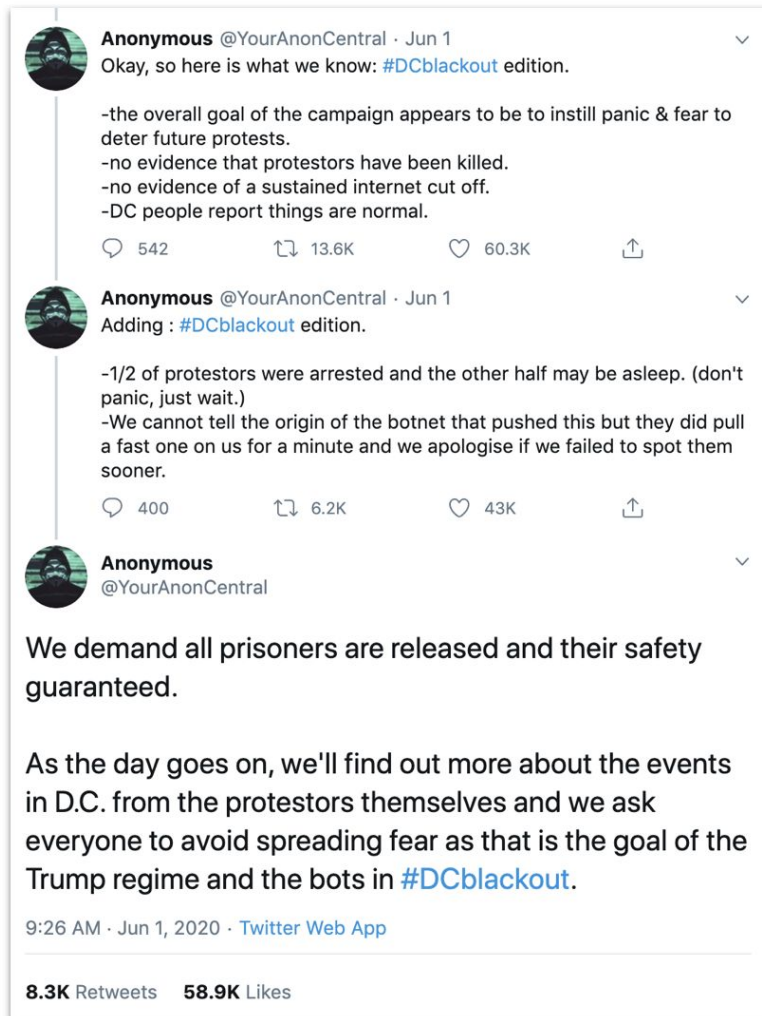
The first [acknowledgment](#) that #DCblackout may have been an error.





*Retweets by @YourAnonCentral rebutting the earlier claims.*

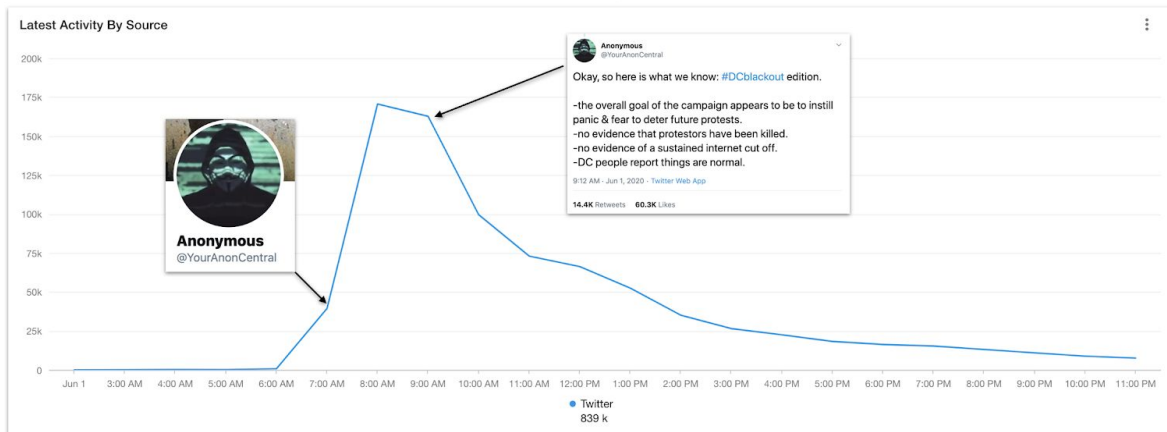
Although @YourAnonCentral played the key role in sending the hashtag viral, it played an important role in debunking it. Shortly after 9 AM, an hour and a half after it had first used the hashtag, the account [tweeted](#) a clear summary of its rebuttal, underlining that there was no evidence that protesters had been killed and “no evidence of a sustained internet cut off.” The account further advised users to “stop spreading fear” and apologised (using the UK spelling) “if we failed to spot” the false information sooner.



The key [tweets](#) from @YourAnonCentral correcting the record on its own story, and apologizing.

The calming effect was significant. Between 9 and 10 AM, traffic on the hashtag fell off almost as steeply as it had climbed and continued a downward trend thereafter. It is, of course, more difficult to ascribe this to a single factor than the explosive impact of @YourAnonCentral's earlier retweets, but it is likely that such a significant influencer's decision to correct the record and advise its users not to amplify the message was one of the major factors involved.

## #DCblackout, June 1

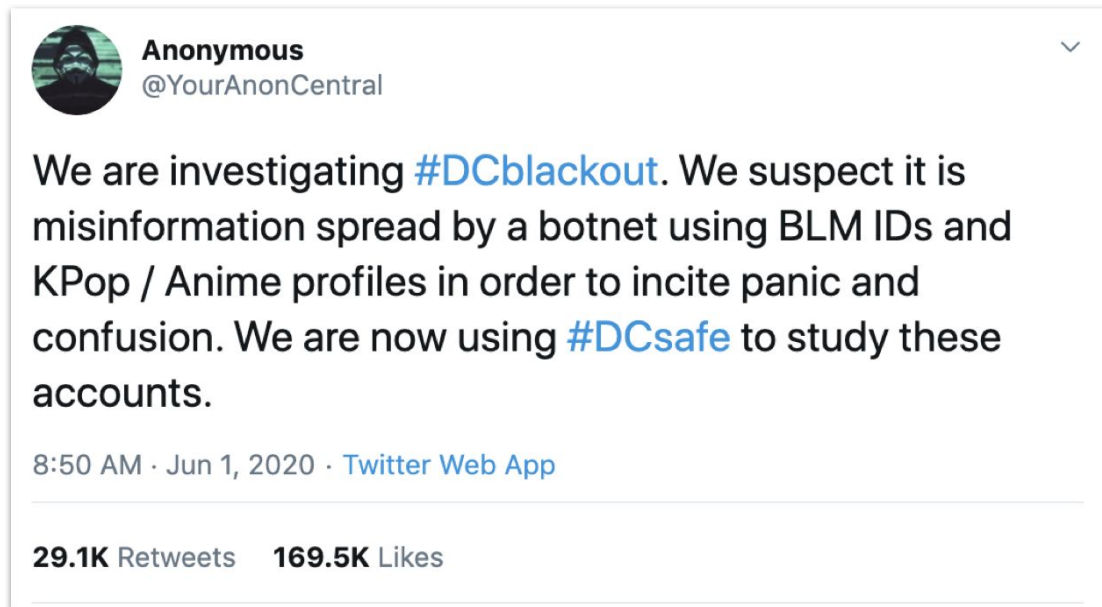


Timeline of traffic on #DCblackout on June 1, showing the moment when @YourAnonCentral first used the hashtag (left) and when it refuted it (right).

## Part 4: Confusion Compounded

### “Kpop bots”

*The rebuttals by @YourAnonCentral helped calm the main hashtag traffic, but they set off a separate chain of speculation by claiming that there had been bots involved, specifically “a botnet using BLM IDs and KPop / Anime profiles in order to incite panic and confusion.”*



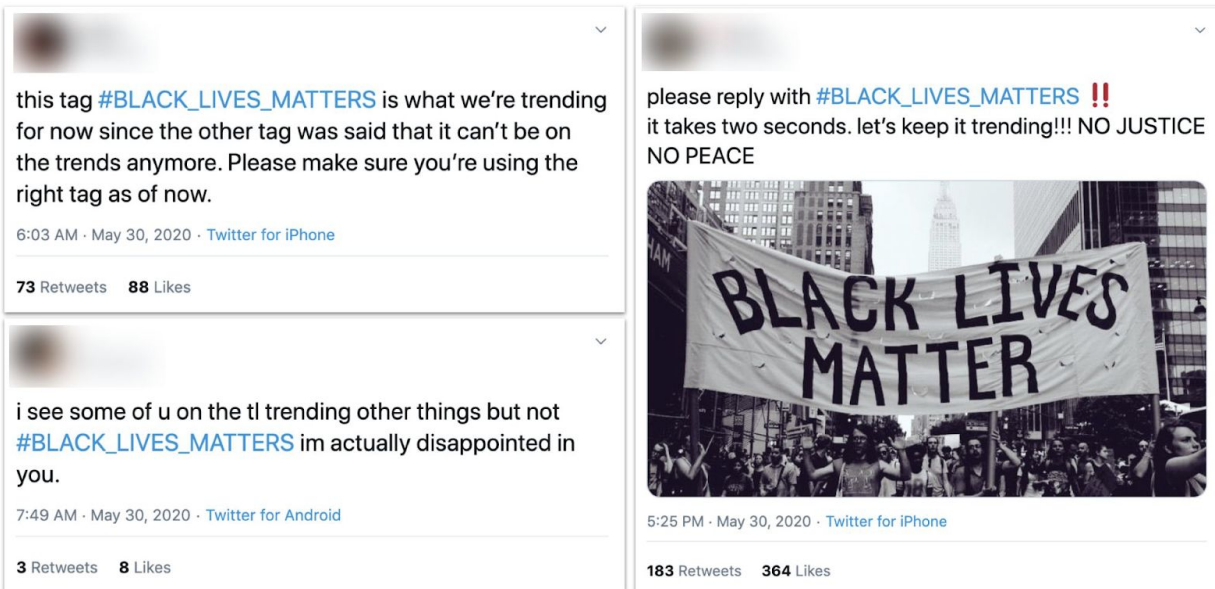
[Tweet](#) by @YourAnonCentral at 08:50 on June 1, an hour and a quarter after the account began retweeting #DCblackout content.

[Subsequent reporting confirmed](#) that Twitter had taken down “hundreds of spammy accounts” on the hashtag, although it is unclear at what stage of the process these spammy accounts had begun to boost it. The claim of bot intervention is therefore accurate as far as it goes.

However, Graphika’s analysis of traffic data four days later, after the bots had been removed from the platform, suggests that bot intervention was only one part of the equation. At least equally important was that genuine accounts from the K-pop and Anonymous communities had coalesced around the U.S. protest movement and begun amplifying each other.

@YourAnonCentral itself is emblematic of this: as of June 9, a scan of its 3,163 most recent posts (both tweets and retweets) spanning the period from April 13 onward showed that it had used the hashtag #BlackLivesMatter 181 times, #ICantBreathe 175 times, and #OpFanCam (an online movement that saw K-pop fans swamp American [police servers](#) with pop footage to prevent others being able to upload images of protesters) 21 times. In effect, one of Twitter’s most influential Anonymous accounts had joined the global BLM campaign.

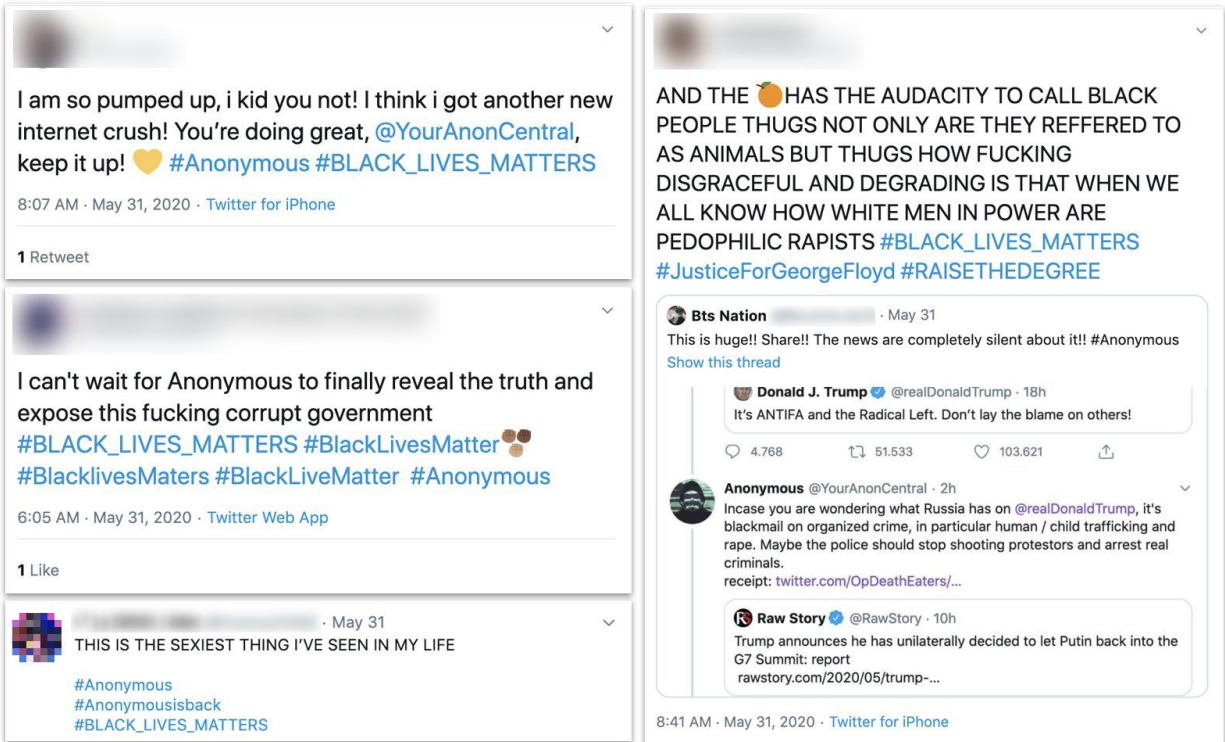
As has been [widely reported](#), K-pop fans had followed an equivalent trajectory. One of their particular practices was coordinating on agreed hashtags, whether to swamp perceived “[enemy](#)” trends or to launch hashtags of their own. On May 30, K-pop fans coalesced around the hashtag #Black\_Lives\_Matters (as opposed to the typical #BlackLivesMatter) and put out a call to their communities to make it trend. On that day alone (from midnight to midnight UTC), the hashtag was tweeted over 28,000 times.



*Tweets posted by K-pop fan accounts with respectively 36,200 followers, 22,300 follower and 27,300 followers, calling for #Black\_Lives\_Matters to trend on May 30.*

On May 31, @YourAnonCentral [tweeted](#) a video attributed to the Anonymous movement threatening to expose the “many crimes” of U.S. law enforcement and elites. It was explicitly targeted at the protests, using the hashtags #Anonymous, #JusticeForGeorgeFloyd, #ICantBreathe, and #BlackLivesMatter. The account followed up by [accusing](#) President Trump of “child trafficking and rape,” based on a set of documents [leaked](#) in 2019. Within a few hours, accounts that appeared authentic and were largely themed to K-pop, but which had started to post comments about the U.S. protests, began combining the hashtags #Black\_Lives\_Matters and #Anonymous in an outburst of cross-cultural fandom. By day’s end, the combination had featured in tweets from 35,000 accounts.





*Tweets by Kpop fan accounts combining #BlackLivesMatters and #Anonymous. Note that the top left account tagged @YourAnonCentral, while the top right retweeted a post from a BTS-themed account that retweeted @YourAnonCentral's attack on Trump. The "sexiest thing" that the bottom left account referenced was the Anonymous video.*

Some of the accounts involved do appear to have been spam-bots that hijacked the hashtags for their own purposes (such as [selling Windows software](#) or [advertising bot services](#)),<sup>2</sup> but they were outweighed by accounts that looked authentic and were largely themed to K-pop, Anonymous, or Black Lives Matter.

The impact of Anonymous' posts on May 31 was massive. In that one day, the hashtag #Anonymous featured in 4.8 million tweets from 3.3 million users. Just the @YourAnonCentral account's tweets amassed over 1.25 million retweets (not counting a [tweet](#) that said "Please stop sending us nudes or we will close our DMs" and scored 136,000 more).

Its increase in followers was correspondingly huge: between [11:35](#) GMT and [14:40](#) GMT on May 31, its following soared from 835,000 to 1.26 million. (As of June 15, it had 6.5 million followers.) Such an increase will always raise questions over how many of those new followers were automated, but Graphika's maps confirm that the followers came from a number of distinct communities, including K-pop.

<sup>2</sup> The profile picture of the latter account was copied from [UNICEF Macedonia](#).

Based on that evidence, which was gathered *after* Twitter had taken down the bots reported on June 1, it appears likely that at least some of the members of the “botnet using BLM IDs and KPop / Anime profiles” which @YourAnonCentral called out were, in fact, users from the K-pop and Black Lives Matter communities who had made common cause with the Anonymous movement to support the protests.

## Hacked Accounts

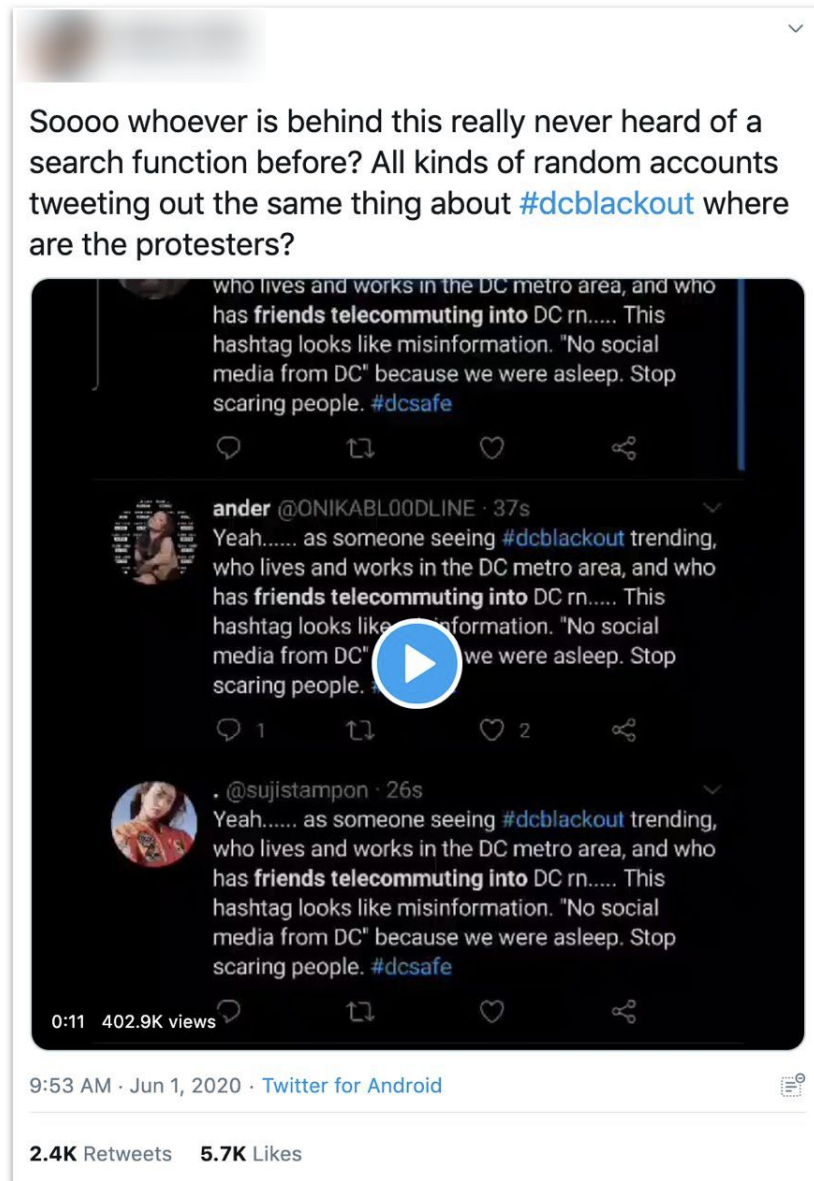
One of the tweets that Anonymous retweeted to calm users’ nerves ended up having the opposite effect, because it was hijacked by unknown users. The original post came from a user who claimed to live and work in the D.C. metro area. It warned (correctly) that the hashtag “looks like misinformation” and advised users to “stop scaring people.” The account holder used the hashtag #DCsafe, which @YourAnonCentral had suggested; @YourAnonCentral retweeted it one minute later.

There is no indication that the original account is in any way inauthentic. A scan of its last 3,170 tweets since December 2019 showed that it posted an average of 17 times a day and authored a high proportion of unique posts. Posting a high proportion of comments that do not appear anywhere else online is typically an indicator of a human operator.



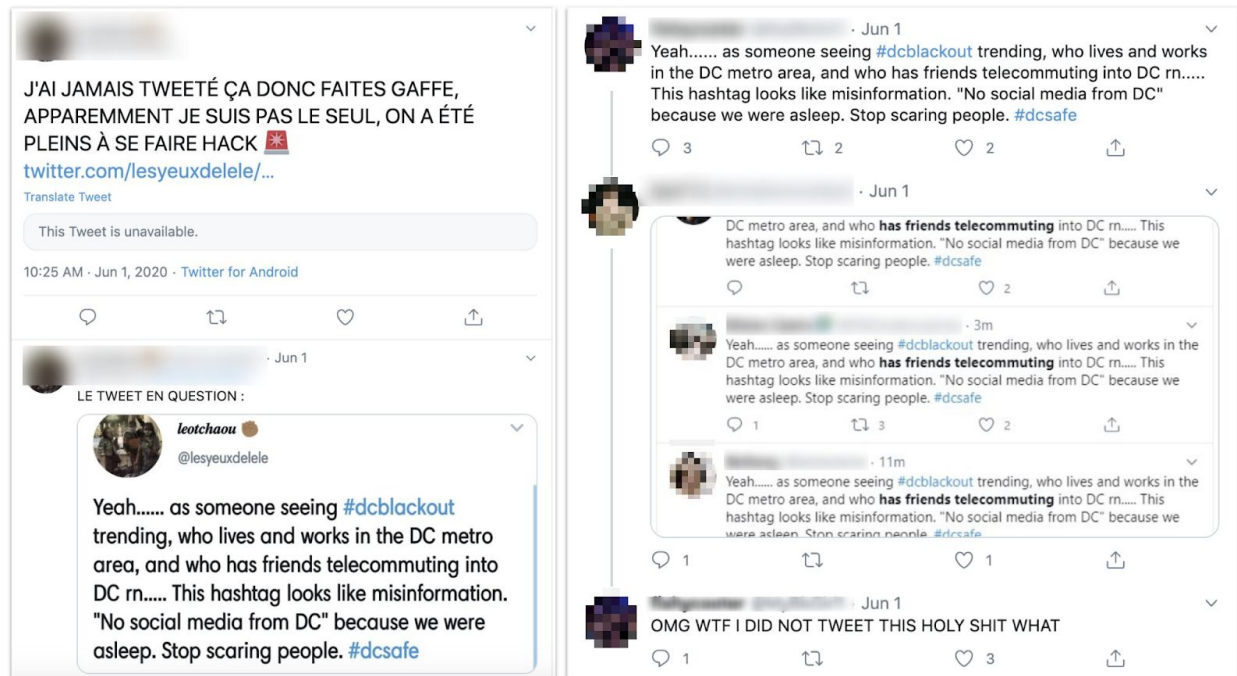
*The tweet to “stop scaring people,” retweeted by @YourAnonCentral.*

However, very shortly credible reports and screenshots began circulating of the same tweet being posted by numerous other accounts. A number of users shared photos and videos of searches for the exact phrase, returning a significant number of identical hits. Again, this led to the conclusion that the “stop scaring people” message was part of a bot-driven disinformation campaign.

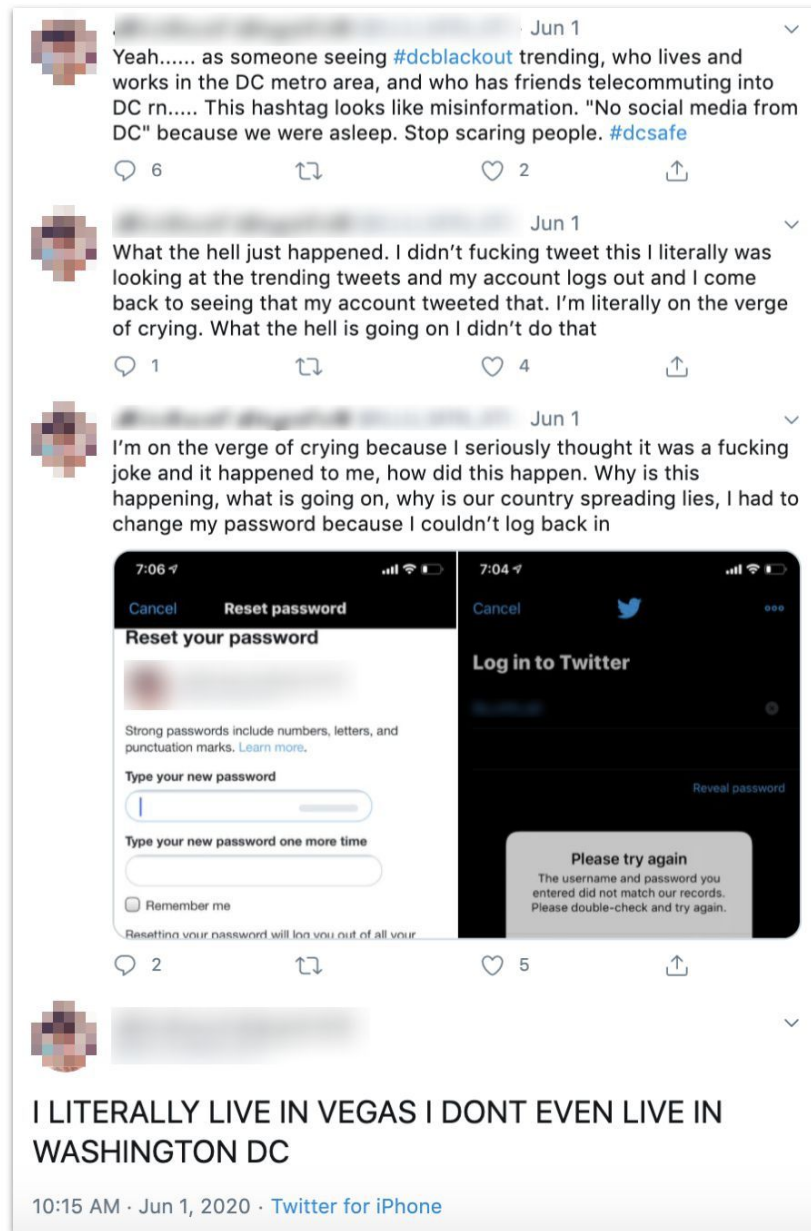


*Tweet from a Seattle-located user of a video of the number of different accounts that posted the same message.*

Some of the tweets were definitely part of a malicious campaign. Soon afterward, Twitter users around the world began complaining that their accounts had been hacked and the tweet had been posted without their consent.



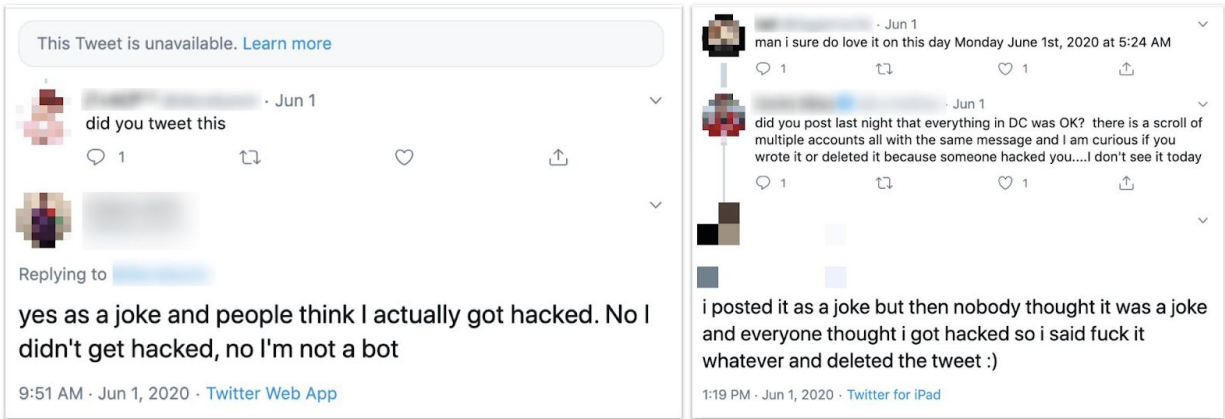
Left: "I NEVER TWEETED THAT SO WATCH OUT, APPARENTLY I'M NOT THE ONLY ONE, THERE ARE LOADS OF US WHO'VE BEEN HACKED." Right, shocked response from another account that posted the tweet.



*A thread from another hacked user, including a screenshot to demonstrate that they had been locked out of their own account in the process.*

However, there was more to the multiple copy-pasting than simply a set of hacked accounts all amplifying it. Some users admitted to manually copying and pasting the text for a joke, a salutary reminder that crisis disinformation can spread for many reasons, including mischief. At least one user admitted to copying and pasting the text because he thought it would be helpful.





*Two tweets from users who re-posted the “stop scaring people” tweet as a joke.*



*Tweets from a Colombian-located account, admitting that he copied the text to “help to spread the information.”*

The repeated copying and pasting of the original, apparently genuine tweet deepened the confusion around the #DCblackout story. At the same time, it was significantly more nuanced than a simple case of “bad bots interfering.” Many of the accounts involved had been hijacked, but the very first to post the text appeared authentic. The focus on the hijacked accounts undermined the genuine attempt to calm things down - the truth lost in a fog of falsehoods.

### “Where are the protesters?”

One final phrase that trended that morning was “where are the protesters,” and this, too, met with accusations of bot amplification. According to [Evan Greer](#), deputy director of the privacy and

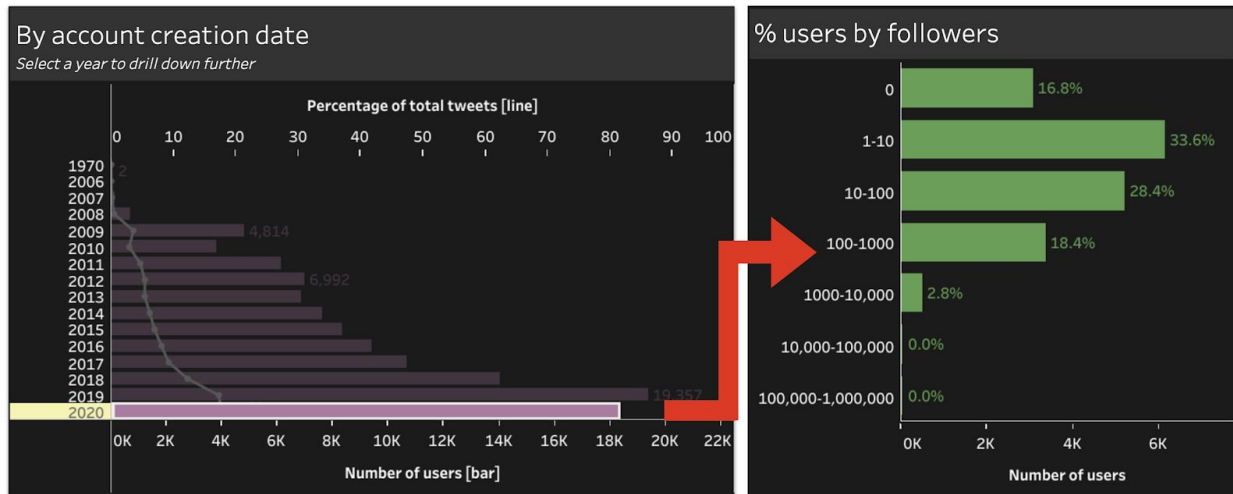


freedom advocacy group Fight for the Future, many of the accounts that tweeted that exact phrase had no followers - traditionally a sign of possible bot involvement.



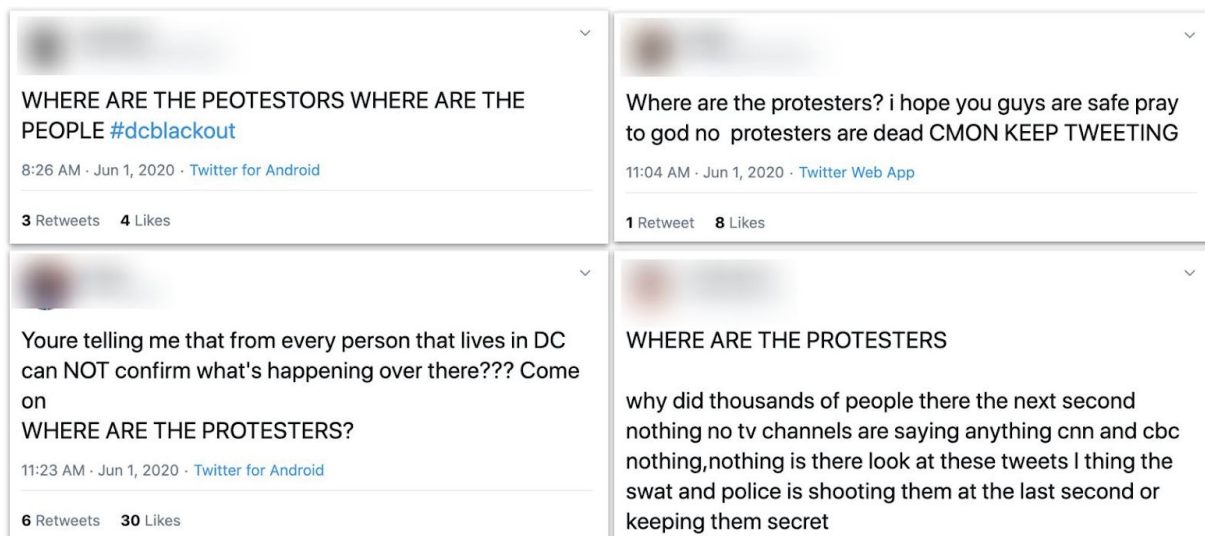
*[Tweet](#) by Evan Greer, pointing out the anomalous presence of "lots of accounts with no followers."*

Graphika also noted the presence of an unusually high proportion of accounts with no followers in the traffic around the #DCblackout hashtag that day. This was particularly true of accounts created in 2020: according to a separate scan of some 190,000 tweets from 117,000 users who posted on the hashtag later in the day on June 1, over 18,000 accounts had been created since January 1, 2020, and of those, one-sixth had no followers at all. Their presence is anomalous and deserves more study.



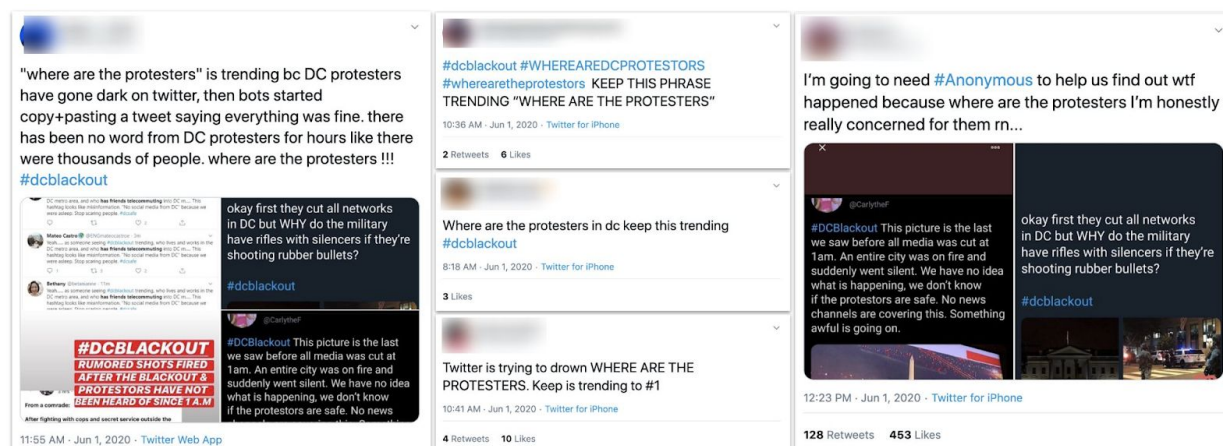
Readout of accounts that posted #DCblackout in the afternoon of June 1, showing (left) the breakdown of account creation per year, and (right) a breakdown of the follower count for accounts created in 2020. The equivalent percentage for earlier years was no more than 3.

However, the number of followers is only a single indicator. Graphika reviewed a number of the accounts without followers that tweeted the phrase “where are the protesters” and found that they did not betray any other indicators of automation. They were recent creations, but their posts included a significant proportion of unique content and they posted at relatively low rates, with variations in language and even in spelling, suggesting that (unless an unusually sophisticated natural language processing algorithm had been deployed) there were human users behind them.



“Where are the protesters” (or protestors or peotestors) from four accounts that, as of June 9, had no followers. All four posted a significant proportion of unique content, including typos.

The phrase “where are the protesters” also featured in a wide range of different linguistic contexts. It was not simply that hundreds of accounts pasted the same comment; they incorporated the phrase into different posts, including a significant number that urged users to make the phrase trend.



*Posts from five separate accounts on June 1 urging users to “keep this trending.”*

## Part 5: Lessons - Learned?

The story of #DCBlackout is one of chaos and confusion. The hashtag emerged in the fog of tear gas in the early hours of June 1, spread through the information vacuum that was left as reliable sources shut down for the night, exploded in the morning after @YourAnonCentral amplified it, and then began to die away as the same account set the record straight. Throughout the morning, conflicting rumors of “bots” added to the confusion as they were used to discredit both the original hashtag and the rebuttal. Some users clearly attempted to subvert the hashtag out of malice or mischief, but the overall tone was one of organic bewilderment, not a disinformation campaign.

Several key lessons emerge from this tale of confusion. The first is the importance of eyewitness testimony from identifiable and accountable users. The journalists who live-tweeted about the night’s events on the ground provided crucial coverage as long as they were still on duty; it was after they shut down that the fear and rumors began to spread, as users turned to unverified and ultimately unreliable sources. In the morning, it was the same voices, especially Yahoo News Hunter Walker, who played the key role in clarifying the situation.

Second, the incident underscores the responsibility of major influencers to check their facts before they share unverified claims. The single most important reason that #DCBlackout trended so high on the morning of June 1 was that @YourAnonCentral retweeted it to over 2 million followers. By the same token, one of the reasons the traffic then declined so steeply was that @YourAnonCentral set the record straight clearly and at some length. In such tense times, any

user with large follower counts has a particular responsibility to check their facts before they post dramatic, but possibly deceptive, claims.

Third, identifying possible bot activity to a high degree of confidence is significantly more complex, challenging, and nuanced than is often thought. Some of the accounts that were labeled as “fake” or possible “bots” during the spread of #DCblackout were indeed automated, and Twitter suspended them. But some of the accounts with no followers looked as though they were run by human users; some of the accounts that copy-pasted the identical phrase did so as a joke or to be helpful; some of the accounts that were flagged because they had K-pop profile pictures but were posting about Black Lives Matter were, in fact, K-pop fan activists. Reliably identifying bots requires [multiple pieces of overlapping evidence](#) and a systematic analysis of an account’s overall behavior.

Beyond that, it is also important to consider how much impact bots - automated amplifiers - really have on the overall conversation, compared with authentic users. Even after Twitter deleted the “hundreds of spammy accounts” on the hashtag #DCblackout, the data flow showed that more than 400,000 users tweeted it over 800,000 times. That would be more than enough to make the hashtag trend without any automated amplification. It is important to expose and take down automated spam accounts, but it is also important not to focus too much on them.

Finally, we are reminded that deceptive behavior has many forms and many causes. Some of the apparently automated accounts that tweeted on the hashtag were commercial spammers, using it to draw attention to their products. Some of the real users who copied the “stop scaring people” tweet did it out of mischief; others did it in an attempt to spread reassuring information. There is no sign that the hashtag was the work of a coordinated disinformation campaign, organized and capable. Rather, it drew many users of many different sorts, who amplified it for many different reasons. It was a conversation, not a campaign. That is one reason it became such a trend.



Graphika

© Graphika All Rights Reserved 2020