PROGRAMME ON
DEMOCRACY
& TECHNOLOGY

China's Inauthentic UK Twitter Diplomacy

A Coordinated Network Amplifying PRC Diplomats

Marcel Schliebs • University of Oxford
Hannah Bailey • University of Oxford
Jonathan Bright • University of Oxford
Philip N. Howard • University of Oxford

OXFORD
INTERNET
INSTITUTE

UNIVERSITY OF
OXFORD

China's Inauthentic UK Twitter Diplomacy

A Coordinated Network Amplifying PRC Diplomats

Marcel Schliebs, Hannah Bailey, Jonathan Bright, Philip N. Howard

EXECUTIVE SUMMARY

Twitter accounts operated by diplomats, including that of the People's Republic of China (PRC), play an important role in the public diplomacy efforts of many governments. We audit all Twitter engagement with PRC diplomats stationed in the United Kingdom over an eight month period, from June 2020 to January 2021.

- We identify a large network of Twitter accounts that demonstrate multiple forms of coordinated inauthentic activity. The network consists of 62 accounts in total, 29 of which were recently active until we flagged their activity for Twitter. Many accounts impersonate UK citizens, with biographies such as "political affairs commentator from London" and usernames such as @JenniferatUK, @UKJenniferin, or @GraceUK5.
- This network has features and behaviors that demonstrate a coordinated information operation:
  - Account creation appears coordinated. Nearly a third of the accounts were created within minutes of each other and the vast majority only amplify and engage with the PRC's diplomats to the UK.
  - Account use appears coordinated. Many accounts sit dormant for extended periods and are activated together at chosen moments. Most parts of the network tend to be active for the morning and early evening hours when social media use in the UK is highest.
  - Account interaction appears coordinated. Many accounts focus exclusively on amplifying UK-based PRC diplomats, and do not engage with other PRC diplomats. Accounts in the network frequently amplify PRC diplomats within sixty seconds of a message from another account in the network. It appears that human operators manage some accounts.
  - Account messages use consistent phrases. Accounts in the network often replicate segments of speeches or commentary from the three most prominent Twitter accounts of the PRC representation in London.
- This coordinated information operation drives a significant proportion of the engagement with the PRC's UK public diplomacy on Twitter. Over the eight month period, 44% of the ambassador's retweets and 20% of his replies came from the coordinated network. At several critical moments, as much as three-quarters of the engagement with the PRC's top diplomat in London came from this inauthentic public diplomacy network.

CONTENTS

1. DETECTING COORDINATED INAUTHENTIC BEHAVIOR IN THE CONTEXT OF PRC-LINKED INFORMATION OPERATIONS

In our global report we find evidence to suggest that the People's Republic of China (PRC) is engaging in a large online public diplomacy campaign, which is supported by suspected inauthentic social media accounts. In recent years, more than 189 Twitter accounts have been created for PRC diplomats and embassies, and these official accounts receive significant amounts of engagement from clusters of other social media users.[1] In this focused case study, we investigate the deployment of this Twitter diplomacy campaign in the United Kingdom, and in so doing we examine potential inauthentic social media behavior more closely.

This report builds on the literature surveyed in our global study, but here we specifically examine the ways by which other researchers have sought to measure coordinated inauthentic social media engagement, and the difficulties associated with this measurement.

An important first step is to clarify the relevant concepts. For "inauthentic social media engagement", we follow Twitter in defining this as "attempt[s] to make accounts or content appear more popular or active than they are".[2] Again, following Twitter, we define inauthentic coordination as "the use of multiple coordinating accounts to inflate the prominence of a particular account or tweet [⋯] or posting identical tweets from multiple accounts operated by a single user".[2] Finally, we refer to Facebook's definition of an information operation as any action "taken by organized actors (governments or non-state actors) to distort domestic or foreign political sentiment" through the use of methods such as, in this instance, constructing "networks of fake accounts aimed at manipulating public opinion".[3]

Coordinated efforts to amplify certain content on social media artificially, also referred to as astroturfing, imitate the organic expression of genuine social movement through inauthentic means.[4], [5] For researchers, this is both an opportunity and challenge. On the one hand, a coordinated campaign is centrally orchestrated by definition, making it hard to fully obfuscate these organizational structures in the digital traces left by an information operation. On the other hand, however, this very goal of imitating genuine grassroots movements often makes the boundaries between inauthentic astroturfing campaigns and genuine movements blurry, as the imitated behaviors employed by information operations are very similar to the traces one would observe within a genuine digital movement.

Some of the evidence about inauthentic coordinated networks supporting PRC diplomats is contested. In May 2020, the US State Department's Global Engagement Center (GEC) accused Beijing of inauthentically amplifying its diplomats on Twitter in an operation with "highly probable links to the Chinese Communist Party".[6] In response to these allegations from the State Department, Twitter disputed the claims, stating that the initial analysis of data provided by the State Department did not support the GEC's statements. Thus, even the US GEC and Twitter disagree on the measurement and identification of inauthentic engagement.

Part of the problem rests in the very concept of inauthenticity. Twitter does not require users to identify themselves and so any attempt to form a judgement about an inauthentic account is nearly impossible. This is because an examination of individual accounts does not allow one to distinguish between maliciously inauthentic accounts and those which merely use the platform in anonymity. This norm has previously been exploited by multiple PRC-linked information operations, that have repeatedly relied on large sets of anonymous accounts for amplification purposes.[7] For example, an investigation of 23,750 accounts suspended for inauthentic engagement by researchers at Stanford University found that the operation relied on curated, recently created accounts. These were often created in batches of hundreds per day, many of which followed each other in networks.[8]

A methodological challenge in any study of information operations is gauging whether the behavior is executed by humans, by automated accounts, or a mixture of both. The referenced study on PRC-backed information operations on Twitter did not, in fact, present conclusive evidence of highly automated behavior; instead, it appeared to be a mostly human operation. Bolsover states that thanks to its vast network of state employees at its disposal for online propaganda, the PRC "does not […] normally have need to use the cheap and dirty strategies of automation and bot accounts on social media".[9] Despite this, analysis by the *New York Times* found potential signs of automation among retweeters including accounts repeatedly retweeting diplomats "at set lengths of time after the original post".[10]

Although there are a variety of approaches for detecting information operations, it is often impossible for researchers to attribute any given information operation to a specific actor. Researchers have limited access to social media data, and these operations are frequently deliberately concealed using digital anonymity tools such as Virtual Private Networks (VPNs). The focus of this study is therefore not to attribute behavior to a specific actor, but to detect patterns of inauthentic coordinated networks using traces included in the data.[4]

One of the simplest ways of establishing connections between groups of accounts is to analyze overlapping features, such as account creation dates. Information operations usually need to acquire a large number of accounts in a short period of time, and typically rely on purchased or hacked accounts. Alternatively, information operations may create these accounts themselves in bulk. In the latter case, this strategy would be visible in the account creation dates. Studies of PRC-attributed information operations have used this detection strategy. One study shows that a group of accounts used in a PRC operation uncovered in 2020 were created within a short period of time.[8]

A second detection approach considers the long-term activity of user accounts. If a group of accounts starts or stops its activity around the same time, this can be interpreted as a suspicious pattern. These kinds of patterns have recently been found in multiple analyses of pro-PRC information operations.[8], [11]

In some cases, short term account behavior patterns are also informative. Cases of overlapping long term patterns may be influenced by other variables, such as geopolitical events or the formation of a genuine grassroots movement. However, short term patterns of simultaneous retweeting within a short period of time can be indicative of inauthentic coordinated activity. For example, Keller et al. choose a maximum threshold of one minute between two concurrently amplifying users to indicate co-amplifying accounts.[5] Vargas et al. and Duh et al. find that simultaneous co-retweeting within a specific time threshold is a strong feature for detecting information operations.[12] [13] These short term behavioral patterns have previously helped to uncover PRC-linked information operations. Researchers at the University of Cardiff discovered a PRC-linked Twitter operation exhibiting various signs of coordination, such as corresponding liking behavior and accounts repeatedly sharing the same URLs in the exact same order to such an extent that it was too improbable to have occurred by chance.[14], [15]

A final approach is to examine the patterns in words and phrases used by accounts in a suspected information operations network. Thus far, attempts to identify PRC information operations using language similarity have mostly focused on topic modelling. This approach involves grouping different accounts in a network based on the topic that they have tweeted about.[8] However, beyond topic modelling, research in the field of computational linguistics commonly examines the linguistic profiles of accounts to profile authors' language use and writing style.[16]

In this report we adopt a number of these methodological approaches to investigate a network of accounts that engage with PRC diplomats in the UK. We examine account-level features, short and long term temporal features, and linguistic patterns to detect potential coordinated information operations.

2. A COORDINATED INAUTHENTIC AMPLIFICATION CAMPAIGN FOR THE PRC'S DIPLOMATS IN THE UK

In this report we examine all tweets, retweets, and replies to tweets by the PRC ambassador to the UK at the time of this analysis, Liu Xiaoming, as well as the official account of the embassy in London. This analysis takes place during an observation period from the 9th of June 2020 to the 31st of January 2021. In total, these two accounts tweeted 3,070 times during that period, 2,375 tweets from the ambassador's account, and 695 from the embassy twitter. These tweets were retweeted by third party users 45,332 times and replied to 52,733 times. For our analyses, we count every tweet that was posted at some point during the observation window, regardless of whether it was later deleted by the author or suspended by Twitter. Due to short electricity outages and other Twitter API-related factors, true figures might be slightly higher.

In total, we identify a set of sixty-two accounts that exhibit multiple signs of coordination. Of these accounts, thirty-one were suspended by the 1st of March 2021, two had been deleted by their operators, and another twenty-nine remained active. These twenty-nine remaining accounts were suspended after we reported them to Twitter on the 28th of April 2021. None of the users have a genuine profile picture or real name.

All the identified accounts focus on the United Kingdom. Nearly all accounts exclusively amplify the UK-based PRC diplomats, and rarely amplify other diplomats stationed elsewhere. As our global study demonstrates, this behavior is highly unusual. Most highly active accounts amplifying PRC diplomats usually simultaneously engage with dozens of other diplomats, in particular the account of the Ministry of Foreign Affairs officials, rather than amplifying any single diplomat exclusively.

Numerous accounts make references to the UK through their user handles, such as @JenniferatUK, @UKJenniferin, @litaoinlondon, @GraceUK5, or @londoneye826. Several accounts also claim they support football clubs in London or Manchester, and frequently use language suggesting they are UK-based, such as the phrase "Here in the UK, …".

In the following subsections, we identify multiple behavioral patterns which uniquely characterize the operation, including account-level characteristics, temporal activity, language patterns, and other digital traces.

Alongside this analysis, it is worth noting that many of the sixty-two accounts in the network replied to almost all of the PRC ambassador's tweets, and retweeted nearly all his tweets. These patterns are displayed in Figure 13 to Figure 16 of the Appendix.

We are conservative in our assessment of accounts in this network of suspected coordinated inauthentic accounts. We only include an account if it exhibits distinctive signs of coordination, or if they are highly likely to be controlled by the same operator as at least one other account in the network.

To benchmark our analysis, we compare each metric and behavior pattern against a reference group of "natural users". This reference group consists of every other user which has engaged with a UK-based PRC diplomat at least once during our window of observation and has also retweeted a UK-based or any other PRC diplomat at least once. In total, we compare our detected network of sixty-two accounts with a reference group of 6,414 users that have retweeted any PRC diplomat during the window of observation and engage with a UK-based diplomat at least once. The reference group consists of 99% of all the users amplifying the UK-based PRC diplomats, while our coordinated network represents the remaining 1%. A complete table of all included accounts and account-level metrics is included in Table 2 of the Appendix.

2.1 Coordinated Account Creation

As noted in Section 1, there are ways an information operation to acquire a large number of accounts. One option is to buy, steal, or repurpose existing accounts. We have seen multiple recent PRC-linked information operations use these strategies.[8] If this option is unavailable or inconvenient, however, an operation may also choose to create accounts en masse. Often, these "sockpuppet" amplifier accounts are created within a short period of time, making them easily identifiable by their shared creation dates. To understand whether any accounts in our network were created in this way, we check the creation dates for the accounts in our suspected network.

Of the sixty-two accounts we suspect of coordinated inauthentic behavior, eighteen accounts (29%) were created in batches within minutes of each other on two days in April and three days in August 2020. Another set of accounts was created in October and November 2020. The earliest these remaining accounts was created in 2015.

Table 1 shows the user handles and account creation dates for eighteen of the accounts, created in five distinct batches. This table illustrates that in many cases, the accounts were created within just minutes or hours of each other. As Section 2.2 will show, these accounts also operate in a coordinated manner, with many frequently posing in within short time intervals of each other.

Table 1: Coordinated Account Creation

Handle	Account Created	Status 1st of March 2021	RTs*	Replies*
@Crouchi27494110	21-04-2020 09:34	suspended	13	7
@HiddenD99075856	21-04-2020 09:42	suspended	12	5
@Diomedeidae10	21-04-2020 09:51	suspended	54	25
@Caterpi27848664	23-04-2020 11:35	suspended	667	1
@ladybug23758032	23-04-2020 11:37	suspended	566	0
@Bumbleb75459847	23-04-2020 11:40	suspended	354	0
@Hushpup16240621	23-04-2020 11:42	suspended	308	0
@MoverShaker5	04-08-2020 16:20	suspended	19	19
@Voiceof95626989	04-08-2020 16:22	suspended	340	50
@Foodfor35226217	04-08-2020 16:34	suspended	15	14
@luckycloud16	11-08-2020 12:59	active	831	7
@JenniferatUK	11-08-2020 13:42	suspended	192	86
@JoeParker135	11-08-2020 15:11	active	528	783
@SunnyWade6	11-08-2020 15:15	active	13	923
@pianotaotao	11-08-2020 17:34	active	557	606
@litaoinlondon	26-08-2020 08:02	active	1,085	301
@whwmaritime	26-08-2020 08:21	active	1,153	31
@coast59965488	26-08-2020 16:53	active	613	1

Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: On the 28th of April we alerted Twitter to the coordinated activity of these accounts. All the accounts that were "active" until then were suspended by Twitter by the 29th of April. RTs* and Replies* are the sum of amplifications of UK-based PRC diplomats (@AmbLiuXiaoMing, @ChineseEmbinUK). Due to short electricity outages and other Twitter API-related factors, true figures might be slightly higher.

Another account coordination feature involves follower networks. Here, two things stand out. First, the majority of the accounts in the coordinated network follow other prominent political figures in the UK. The majority of the accounts in the coordinated network also follow the ambassador at the time, Liu Xiaoming, as well as the account of the PRC Embassy.

Second, a qualitative analysis of following networks allows us to establish a high likelihood of coordination between three accounts, which were all created on the 26th of August 2020. All three accounts reference water or maritime issues in their profile. The first user, @litaoinlondon, has a profile picture showing a stormy ocean, while the other two accounts have no picture but are named @coast59965468 and @whwmaritime respectively. The latter account follows only two accounts, the PRC ambassador and embassy respectively, while the former two, follow a larger set of accounts. In addition to the references to water, both @coast59965468 and @whwmaritim follow similar accounts. As Figure 9 in the Appendix shows, this overlap is particularly strong for accounts that have a clear maritime focus, such as "Maritime Executive" or the International Maritime Organization.

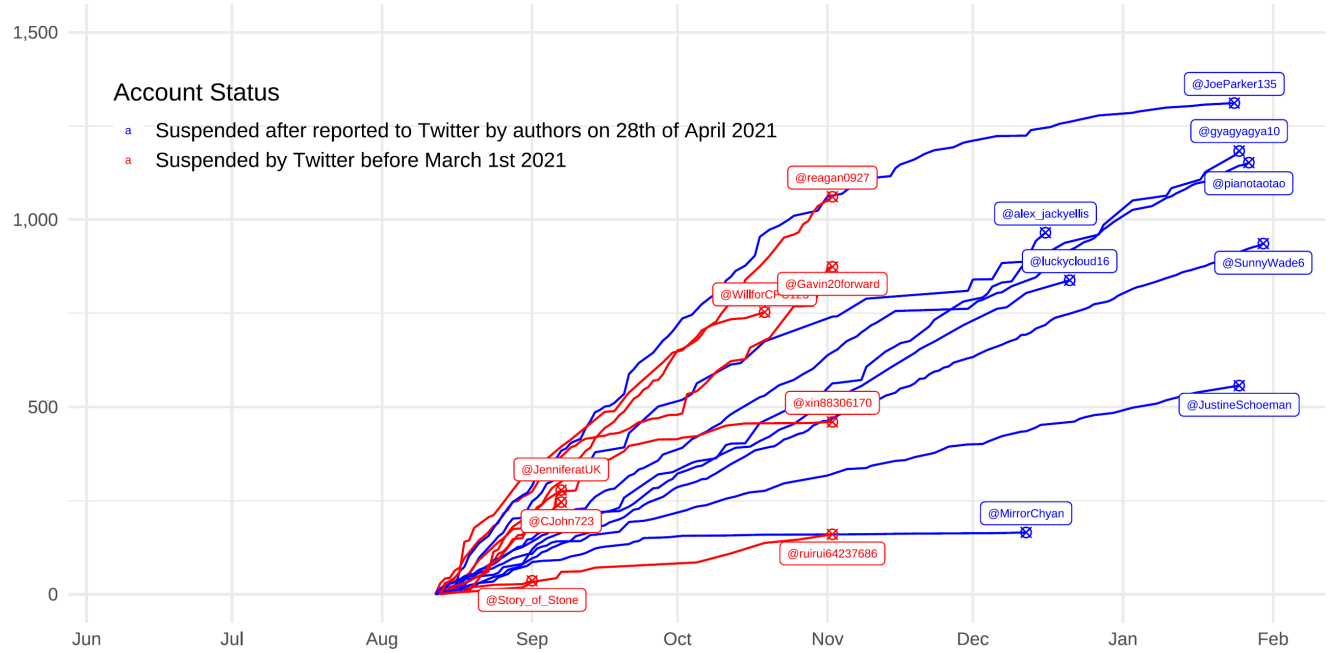
2.2 Coordinated Account Activation

The timings of account tweets and retweets also provide evidence of coordinated activity. Tweet timings can be analyzed in the short or long term. In the short term, we can conclude that a set of accounts is likely coordinated if they repeatedly tweet at similar times to a degree that cannot be explained by chance.

From a long term perspective, we observe the days and weeks when these accounts are active, and whether there are periods where they "wake up" and "fall asleep" again. Figure 1 shows that of the sixty-two accounts in our dataset, sixteen accounts did not tweet during the beginning of our observation window. Rather, these accounts woke up on the 12th and 13th of August and went on to amplify UK-based PRC diplomats several hundreds of times.

It is noteworthy that nine of these sixteen accounts were created between 2016 and 2019, and were likely kept as sleeper accounts, or later acquired and repurposed. To understand this behavior further, we also collect all tweets by the still active accounts that started amplifying UK-based PRC diplomats on the 12th or 13th of August. As Figure 17 in the Appendix shows, four accounts were inactive for multiple years before they started amplifying the UK-based diplomats, and exclusively them, in mid-August. These accounts also ceased nearly all activity after the ambassador was reposted as China's Special Representative on Korean Affairs in February 2021.

Figure 1: Cumulative Amplifications of UK-Based PRC Diplomats by Accounts Waking Up at the Same Time



Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

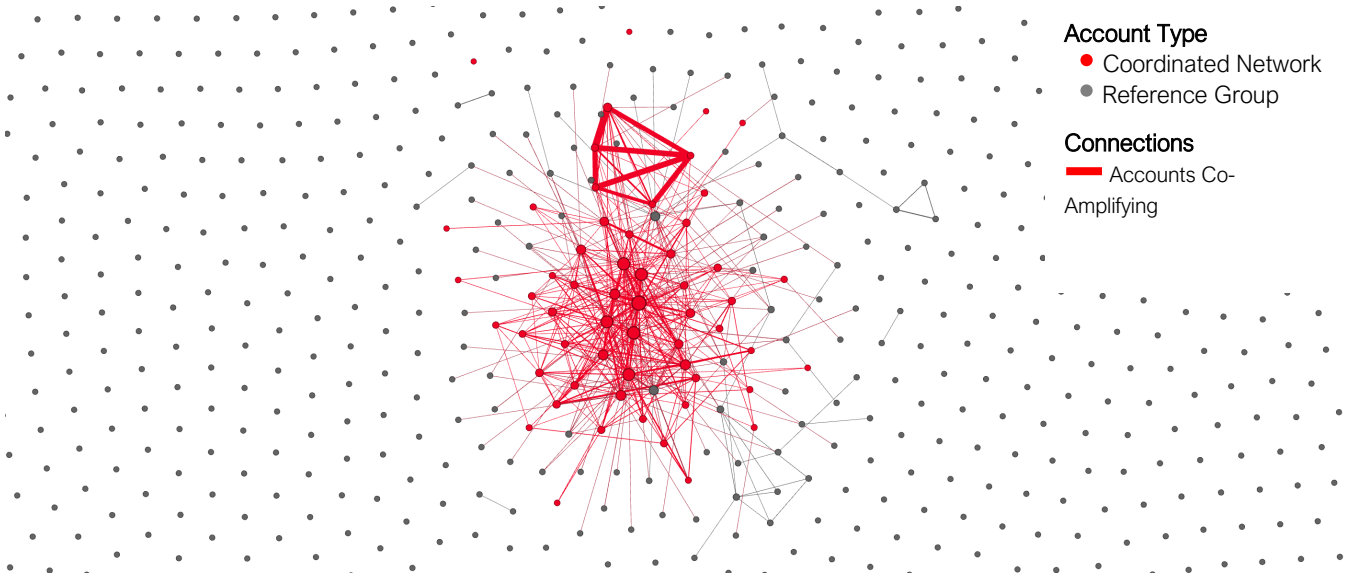
Note: Amplification is defined as the sum of retweets and replies of UK-based PRC diplomats (@AmbLiuXiaoMing, @ChineseEmbinUK). Due to short electricity outages and other Twitter API-related factors, true figures might be slightly higher.

2.3 Coordinated Account Interaction

Figure 1 illustrates that the network of accounts amplifying the PRC ambassador to the UK are coordinated insofar as many of them "woke up" at similar times. We also investigate whether these accounts act in a coordinated manner within shorter time intervals. Accounts that frequently operate within the same short-term time intervals may belong to a broader operation, or even be managed by the same operators. One method to detect suspicious micro-patterns is to examine "co-retweeting", or the frequent retweeting of the same tweet within a short period of time. One minute is a commonly used interval to assess these co-retweeting patterns.[5] We build on this framework, defining the term "co-amplifying" as an event where two accounts reply or retweet to a UK-based PRC diplomat within sixty seconds of each other.

Figure 2 shows a co-amplification network, where two dots are connected with a line corresponding to the number of times they have co-amplified a UK-based PRC diplomat in the same minute. Here, red dots represent accounts in our coordinated network, and grey dots are the aforementioned reference group of other users amplifying PRC diplomats. Following Graham, we exclude lines between accounts that co-amplified only once, thereby reducing the possibility that any two unrelated dots are connected by chance.[17] Figure 2 shows clear patterns of frequent co-amplification by numerous of the accounts in our network. Several connected accounts in our cluster co-amplified over eighty times within sixty seconds of each other. The reference group did not display these patterns, indicating that this operation is uniquely coordinated. While Figure 2 focuses on the center of the network, Figure 18 in the Appendix shows the full network of all accounts engaging with the ambassador.

Figure 2: Accounts Co-Amplifying UK-Based PRC Diplomat Within 60 Seconds of Each Other



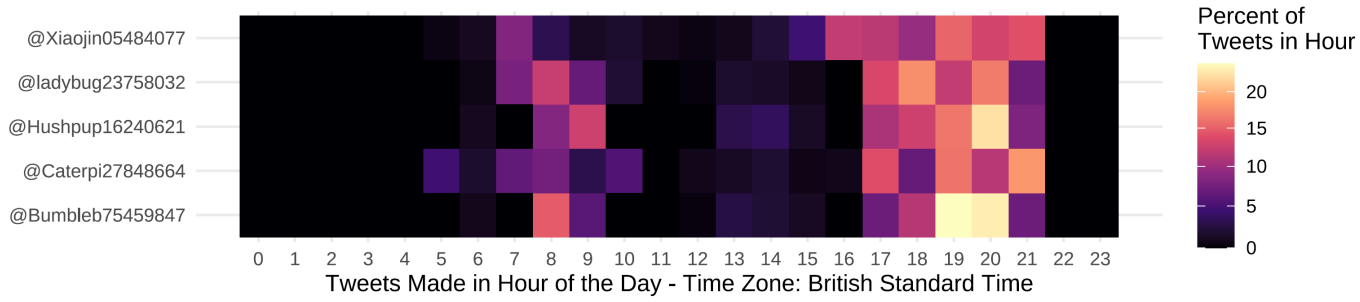
Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Red dots are accounts in our coordinated network, grey dots are a reference group of other users amplifying the PRC diplomats. The dots are connected if they co-amplified a PRC diplomat in the UK within sixty seconds of each other. Thicker ties represent frequent co-amplification.

However, the temporal co-amplification patterns described in Figure 2 should also be interpreted with some caution. While highly frequent co-amplification patterns may suggest central coordination, we also need to take into account potential confounding factors. For example, if the frequent co-amplification occurred repeatedly just seconds after the ambassador posted the original tweet, the co-occurrence may not be due to coordination but due to that third-party influence triggering both accounts to tweet. However, our data does not suggest that a large share of the co-occurrences can be explained by the timing of the original tweets. The co-amplification often happens within a specific minute, often hours or even days after the original tweet was posted by the ambassador. Furthermore, we are confident that given the narrow one-minute time window and the frequency of co-amplification between accounts, coordination is likely, especially given that the reference group exhibits no such behavior.

Figure 3 shows that many of the accounts operated in very similar time patterns corresponding to the rhythm of social media activity of UK-based users, indicating purposeful timing. This assessment is further supported by evidence of coordination in tweet timings. For multiple groups of accounts, we find sequential bulk-retweeting conducted using up to five accounts, presumably by one human coordinator. Each day, the human operator logs on and retweets the ambassador several times dozens of times within a few seconds, before switching to the next account, again within seconds. As Figure 20 in the Appendix shows, this behavior happens on numerous days, and the accounts are switched between in the same sequential order.

Figure 3: Time Patterns of Daily Tweet Activity for Selected Accounts



Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Higher color intensity suggests higher activity in that hour of the day. Dark areas indicate absence of activity. These accounts are selected because they appear to be operated by one human operator (see Figure 20 in the Appendix).

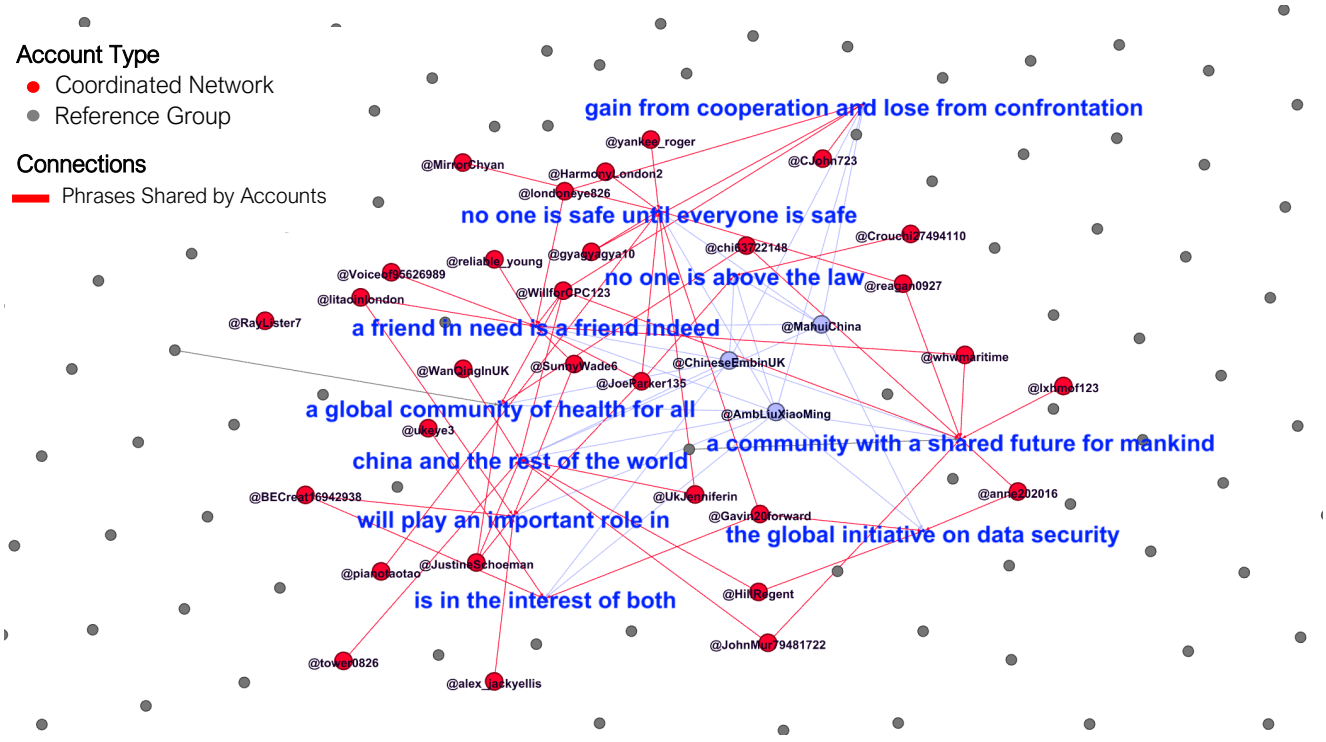
2.4 Coordinated Phrase Use

Thus far we have investigated both long- and short-term Twitter account activity patterns to highlight the signs of coordination between different accounts in our suspected network. As outlined in Section 1, another approach used to detect covert coordination is to analyze patterns in the language used by Twitter accounts in their posts. Research in computational linguistics shows that it is very difficult to obfuscate one's own writing style effectively.[16] To determine whether there are overlapping patterns in language use, we parse all 53,000 replies to a UK-based PRC diplomat into phrases of four to six words. We then analyze the language patterns which were used by at least five different users in replies to UK-based PRC diplomats during that time period.

Based on this data, we find significant overlap in the language used by accounts in our network and that of the three accounts of UK-based PRC diplomats. During our eight month study, in their replies to diplomats, accounts in our inauthentic network use an average of forty-nine phrases which were also used by the ambassador or embassy account. Reference group users, however, use only 0.5 such phrases.

Figure 4 illustrates this strong overlap for a selection of the ten most distinctive phrases shared by the PRC's three UK-based diplomatic accounts, and the larger coordinated network. In this graphic, the red dots represent accounts in the coordinated network, blue points represent the UK-based PRC diplomat accounts, grey dots represent the reference group, and the blue words represent the ten most distinctive phrases in our inauthentic network. The proximity and connecting lines of an account to a blue phrase represents the frequency with which that account used the phase. Notably, the red accounts are highly clustered around the ten distinctive blue phrases, unlike the grey reference group accounts. We can therefore infer that the use of these shared phrases is distinctive to the accounts in our network and the diplomats. The full set of accounts and phrases is included in Figure 21 and in Figure 22 the Appendix.

Figure 4: Top 10 Distinctive Phrases Used by Coordinated Network Accounts and Diplomats



Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Figure 4 shows phrases used nearly exclusively by accounts in the coordinated network and the three UK-based diplomatic accounts. This behavior was distinctive to the coordinated network and did not occur in the reference group of all other users engaging with the diplomats. The reference group consists of 1,451 grey points containing users who replied to PRC diplomats in the UK, scattered far beyond the borders visible in this graph, and visible more clearly in Figure 22 of the Appendix. Red dots are accounts that are part of the coordinated network. The blue phrases represent the ten most distinctive phrases in our inauthentic network. Overlapping sequences (e.g. "a friend in need is" and "in need is a friend") were grouped together in this figure.

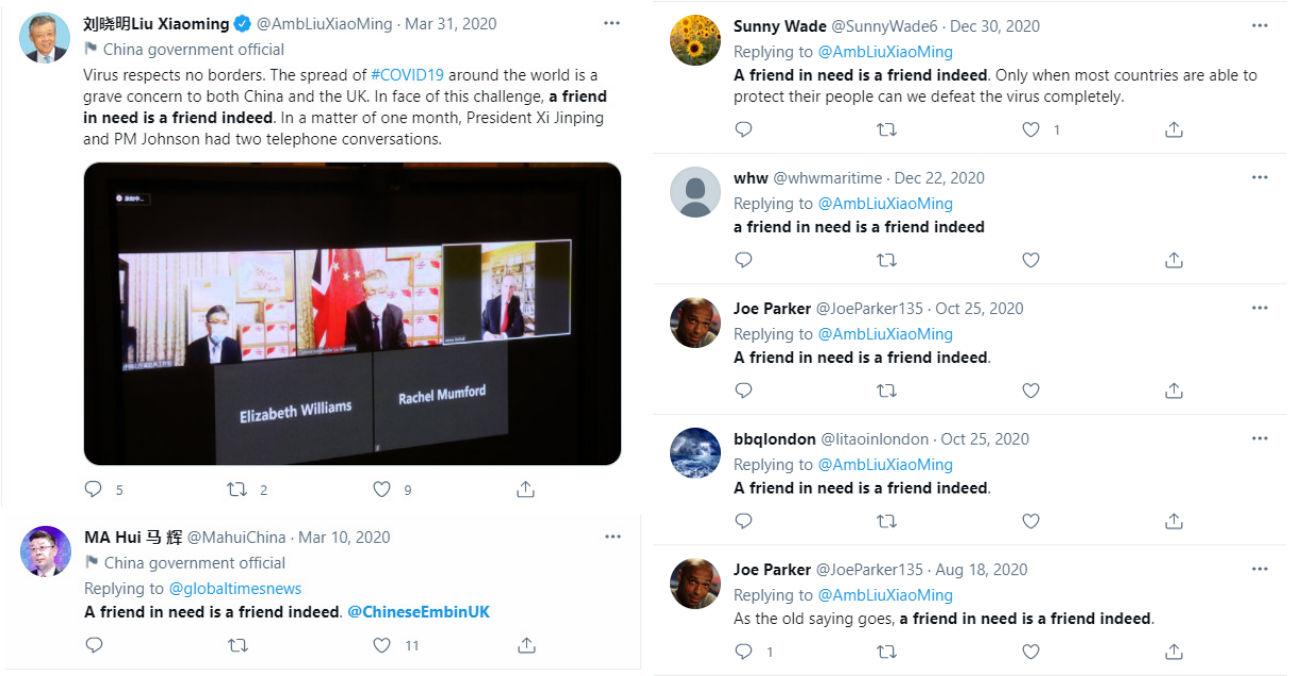
Not only are these phrases distinctive to the coordinated network, but in many cases, they had also been used by PRC diplomats months prior. As Figure 5 further illustrates, the use of this distinctive language was often spread out over months. These phrases were also rarely used by any of the accounts in our reference group.

A further linguistic feature in our data is the use of simultaneous verbatim reply- and quote-tweets. We define this term here as the act of reacting to a diplomat tweet by replying or quote-tweeting the same verbatim text. Often these replies or quote-tweets occur within seconds of each other.

Figure 6 demonstrates one of many examples of this pattern. In total more than half of all users in our inauthentic network engaged in verbatim reply- or quote-tweeting at least once during our window of observation. While over half the users in our network engaged in this behavior, none of the over 6,000 accounts in our reference group did so.

Figure 25 in the Appendix shows one instance where nearly one quarter of the accounts in our network engaged in this behavior in response to a single tweet by the PRC ambassador to the UK. These reply- and quote-tweets occurred within just minutes of each other.

Figure 5: Consistent Use of Distinctive Phrase Reoccurring Over Months



Source: Authors' screen captures.

Note: Multiple uses of the same phrase spread out over multiple months. Often, the accounts in the coordinated network used the same language months after the ambassador had used the exact matching verbatim quotes months or years earlier.

Figure 6: Verbatim Quote- and Reply-Tweeting

Source: Authors' screen captures.

3. COORDINATED CAMPAIGN IMPACT

Figure 7 reveals the impact that the cluster of sixty-two suspected coordinated accounts was able to wield on a weekly basis. Both graphics in Figure 7 take a closer look at the accounts that retweet ambassador Liu Xiaoming and the PRC UK embassy.

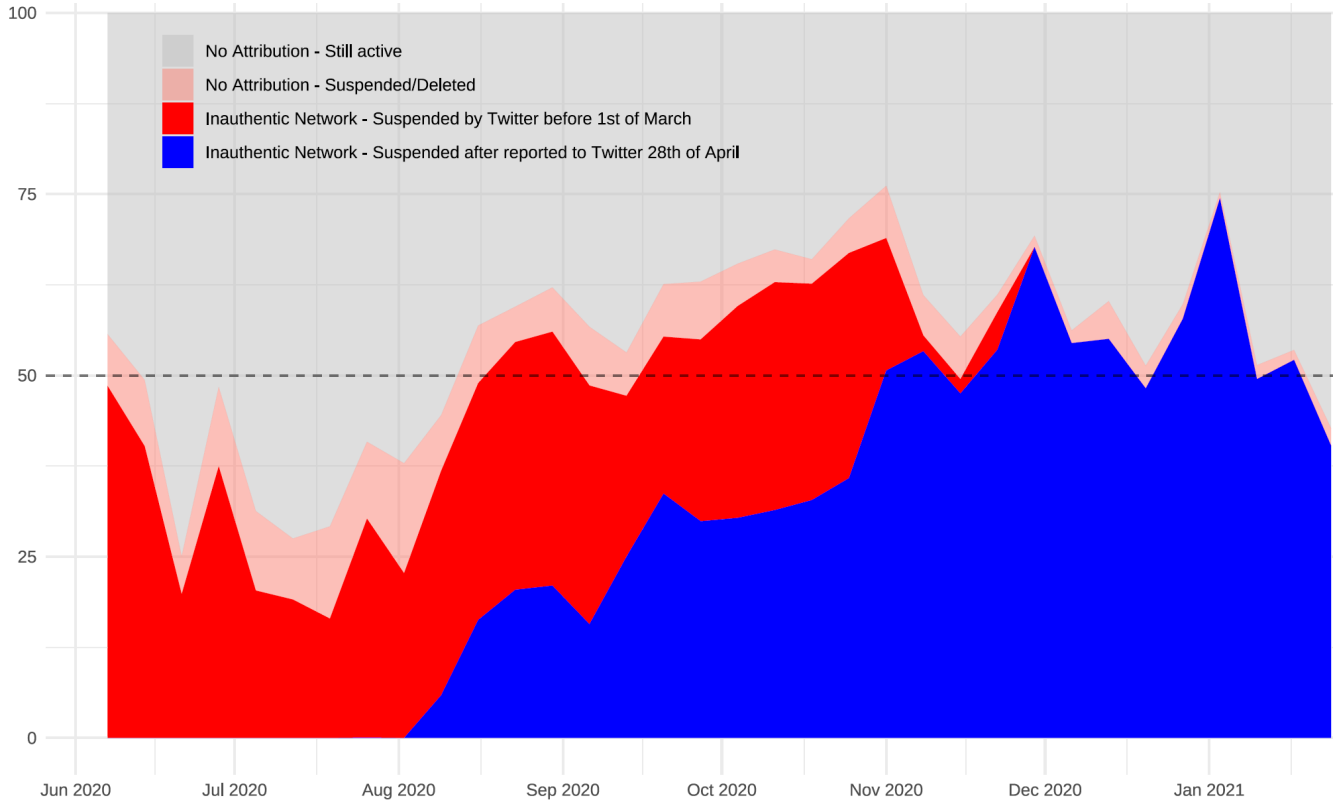
Within both graphics, the blue area represents the cumulative weekly share of the ambassador's retweets stemming from accounts that were only suspended after we reported them to Twitter on the 28th of April 2021. The solid red area depicts the weekly share of retweets by accounts in our coordinated cluster that were suspended by Twitter before March. The shaded red area represents retweets by accounts that we did not identify as part of the coordinated campaign but that were suspended by Twitter for other reasons. The remaining grey area represents retweets by other accounts that were neither found to be part of this coordinated operation, nor were suspended by Twitter.

In total, during our eight month observation window, the accounts in our coordinated network generated 18,784 (44%) of all retweets of the ambassador, and 931 (30%) for the embassy account.

As Figure 7 illustrates, the suspended part of the coordinated network was able to generate between 20% and 50% of the ambassador's weekly engagement in the summer of 2020. This figure reached 75% in some weeks in November 2020 and January 2021. Similar figures for the embassy account are displayed in Figure 8.

Furthermore, Figure 26 in the Appendix shows that accounts from our coordinated network are responsible for over half of the weekly replies to the ambassador's tweets in some weeks. In total, the sixty-two coordinated accounts accounted for 8,750 replies, representing 20% of all replies to the ambassador.

Figure 7: Share of Weekly Retweets to Ambassador (@AmbLiuXiaoMing) Attributed to Coordinated Network (Percent)



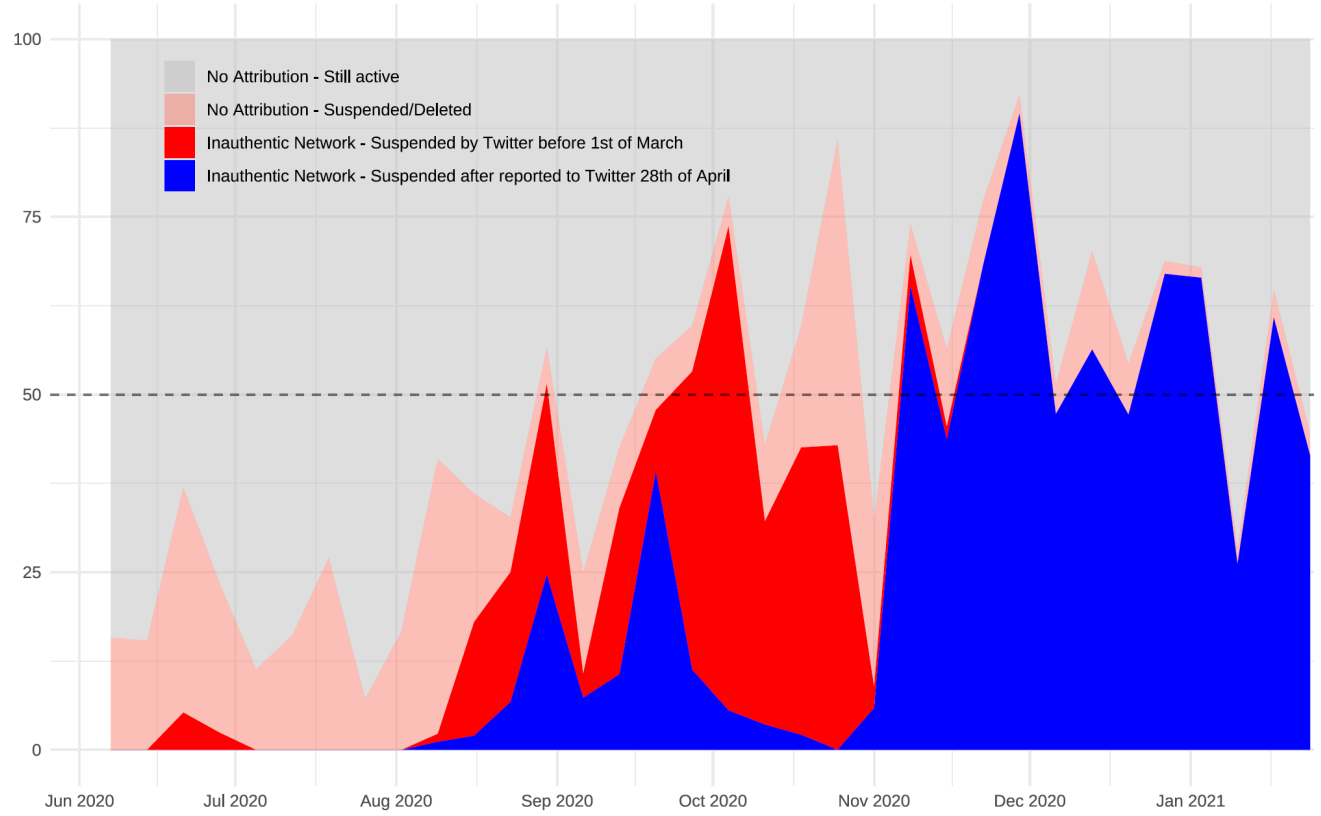
Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Blue solid area represents retweets by accounts still active before flagged to Twitter; solid red areas represent accounts in our detected operation which Twitter had already suspended, and light red represents accounts suspended by Twitter which were not part of our coordinated cluster.

These findings demonstrate that for over an extended period of time, a majority of the engagement with content posted by the PRC ambassador and embassy in the UK was generated through coordinated inauthentic behavior. However, this does not necessarily mean that the campaign had a high impact, since many of the amplifications did not see authentic engagement. Few of the retweeting or replying accounts had any genuine followers themselves, and their replies did not generate any significant further engagement by genuine users.

Despite the low levels of additional engagement among genuine Twitter users, these high levels of inauthentic engagement are remarkable for two reasons. First, artificial engagement can amplify content by manipulating Twitter's recommendation algorithm, in turn leading more genuine UK-based Twitter users to see the content. And second, artificially increasing the retweet and engagement counts of tweets may also benefit the account holder. In our case, the ambassador or the embassy's, status and reputation may have benefited, as higher retweet counts suggest broader support among their target audience.

Figure 8: Share of Weekly Retweets to Embassy (@ChineseEmbinUK) Attributed to Coordinated Network (Percent)



Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Blue solid area represents retweets by accounts still active before flagged to Twitter; solid red areas represent accounts in our detected operation which Twitter had already suspended, and light red represents accounts suspended by Twitter which were not part of our coordinated cluster.

4. CONCLUSION

This report provides evidence of a coordinated amplification network which appears to underpin the PRC's overall UK public diplomacy on social media. In our case study, which examines eight months of activity by clusters of accounts around PRC diplomats in the UK, we find that nearly half of engagement with these accounts can be identified as an inauthentic public diplomacy network.

This coordinated network of accounts consists of sixty-two accounts in total. Thirty-three of these had already been suspended or deleted before the 1st of March, and the remaining twenty-nine were suspended after we alerted Twitter to them on the 28th of April 2021. One feature which distinguishes these accounts is that many tend to impersonate UK-based individuals.

Our detailed examination of the characteristics of this inauthentic public diplomacy network reveals four ways in which the activity appears highly coordinated. First, the creation of the accounts themselves appears to be coordinated. For instance, we find that nearly a third of the accounts were created within minutes of each other.

Second, the usage of these accounts also appears to be coordinated. Evidence for this coordination is that many accounts lie dormant for extended periods and are activated together at chosen moments for particular issues. Further evidence for this coordinated usage is that the entire network tends to demonstrate a common pattern of peaks and troughs throughout the day.

Third, accounts also focus exclusively on amplifying UK-based PRC diplomats, and they often do so within a minute of a message from another account in the network. We also find evidence to suggest that some human operators manage multiple accounts that are used in a rapid and consistent sequence.

Fourth, we find coordinated content in the consistent usage of keywords and phrases, which sometimes extends to the replication of segments of speeches or commentary.

Nonetheless, these findings should be viewed with some caution. First, the evidence of inauthentic social media engagement is by its nature limited, as our data are only able to measure inauthenticity indirectly. That is, we offer analyses from different perspectives which, taken together, strongly suggest coordinated activity which in all likelihood could not have happened by chance.

Second, we were able to establish strong evidence of between account coordination for numerous clusters within the broader network of sixty-two accounts amplifying the UK-based PRC diplomats. However, the strength of coordination varies between different clusters of accounts. Some show obvious signs of coordination, indicative of a single human operator. For other accounts, the picture is more ambiguous. Thus, we cannot definitively conclude whether the sixty-two accounts in question were operated by one, a handful, or more operators. Given the complexity of the data collection and analysis process as well as scope restrictions for our disclosure report, we have so far presented only the most convincing and strongest analysis and data. Future work could build on our analysis.

And third, as our report uses open-source data, we are not able to conclusively attribute this coordinated operation to any state or non-state actor. Although we show that numerous clusters are probably controlled by the same human operator, the exact nature and full scope of coordination requires further analysis. Future work could address the challenge of attribution, for instance, by examining phrases repurposed from the tweets which the PRC ambassador posted months or years ago. One could also examine whether the language of both the accounts and the ambassador were inspired by a shared third source or centrally coordinated in some way.

Our findings are relevant to industry, policymakers, and wider society in a few ways. First, we show that many inauthentic accounts were able to amplify PRC diplomats hundreds or thousands of times over a period of several months before being detected and suspended. However, it is certainly worth noting that when we alerted Twitter to the activity of suspicious accounts, the firm acted promptly to suspend them. Second, whoever orchestrated this campaign did so in violation of Twitter's platform manipulation rules on inauthentic behavior and coordination. And third, while we show that the inauthentic amplification campaign accounted for high relative levels of the diplomats' engagement, future research could examine how far this inauthentic content is able to penetrate genuine local audiences, and whether it is able to shape their perceptions and attitudes.

REFERENCES

[1] M. Schliebs, H. Bailey, J. Bright, and P. N. Howard, "China's Public Diplomacy Operations: Understanding Engagement and Inauthentic Amplification of PRC Diplomats on Facebook and Twitter," Oxford, UK: Programme on Democracy and Technology, Oxford University, 2021. [Online]. Available: https://demtech.oii.ox.ac.uk/china-public-diplomacy-report

[2] Twitter Inc., "Twitter's Platform Manipulation and Spam Policy | Twitter Help," 2020. https://help.twitter.com/en/rules-and-policies/platform-manipulation

[3] J. Weedon, W. Nuland, and A. Stamos, "Information Operations and Facebook," Facebook Inc., 2017. [Online]. Available: https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf

[4] P. N. Howard, *New Media Campaigns and the Managed Citizen*. New York, NY: Cambridge University Press, 2006.

[5] F. B. Keller, D. Schoch, S. Stier, and J. Yang, "Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign," *Polit. Commun.*, vol. 37, no. 2, pp. 256–280, Mar. 2020, doi: 10.1080/10584609.2019.1661888.

[6] L. Gabrielle, "Briefing With Special Envoy Lea Gabrielle, Global Engagement Center Update on PRC Efforts to Push Disinformation and Propaganda around COVID," presented at the United States Department of State - Special Briefing, May 08, 2020. [Online]. Available: https://2017-2021.state.gov/briefing-with-special-envoy-lea-gabrielle-global-engagement-center-update-on-prc-efforts-to-push-disinformation-and-propaganda-around-covid/

[7] T. Uren, E. Thomas, and J. Wallis, "Tweeting Through the Great Firewall: Preliminary Analysis of PRC-linked Information Operations on the Hong Kong Portest," *Aust. Strateg. Policy Inst. Int. Cyber Policy Cent. Barton*, 2019.

[8] C. Miller, V. Molter, I. Garcia-Camargo, and R. DiResta, "Sockpuppets Spin Covid Yarns: An Analysis of Prc-Attributed June 2020 Twitter Takedown," 2020. [Online]. Available: https://fsi.stanford.edu/publication/june-2020-prc-takedown

[9] G. Bolsover, "Social Media, Computational Propaganda, and Control in China and beyond," in *The World Information War: Western Resilience, Campaigning, and Cognitive Effects*, R. Johnson and T. Clack, Eds. Taylor & Francis, 2021. [Online]. Available: https://books.google.co.uk/books?id=JVkkEAAAQBAJ

[10] R. Zhong, A. Krolik, P. Mozur, R. Bergman, and E. Wong, "Behind China's Twitter Campaign, a Murky Supporting Chorus," *The New York Times*, Jun. 08, 2020. [Online]. Available: https://www.nytimes.com/2020/06/08/technology/china-twitter-disinformation.html

[11] J. Wallis *et al.*, "Retweeting Through the Great Firewall," 2020, [Online]. Available: https://www.aspi.org.au/report/retweeting-through-great-firewall

[12] L. Vargas, P. Emami, and P. Traynor, "On the Detection of Disinformation Campaign Activity with Network Analysis," *ArXiv200513466 Cs*, 2020, Accessed: Apr. 23, 2021. [Online]. Available: http://arxiv.org/abs/2005.13466

[13] A. Duh, M. Slak Rupnik, and D. Korošak, "Collective Behavior of Social Bots Is Encoded in Their Temporal Twitter Activity," *Big Data*, vol. 6, no. 2, pp. 113–123, 2018, doi: 10.1089/big.2017.0041.

[14] Crime and Security Research Institute, University of Cardiff, "China-linked Influence Operation Detected on Twitter Interacting with the US Presidential Election (1/2)," Crime and Security Research Institute, Cardiff, Detection Report, 2020. [Online]. Available: https://www.cardiff.ac.uk/news/view/2491763-hundreds-of-fake-twitter-accounts-linked-to-china-sowed-disinformation-prior-to-the-us-election,-with-some-continuing-to-amplify-reactions-to-the-capitol-building-riot-report

[15] Crime and Security Research Institute, University of Cardiff, "China-linked Influence Operation Detected on Twitter Interacting with the US Presidential Election (2/2)," Crime and Security Research Institute, Cardiff, Development Report, 2021. [Online]. Available: https://www.cardiff.ac.uk/news/view/2491763-hundreds-of-fake-twitter-accounts-linked-to-china-sowed-disinformation-prior-to-the-us-election,-with-some-continuing-to-amplify-reactions-to-the-capitol-building-riot-report

[16] E. Stamatatos, "A Survey of Modern Authorship Attribution Methods," *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 3, pp. 538–556, 2009.

[17] T. Graham, "Detecting and Analysing Coordinated Inauthentic Behaviour on Social Media," 2020. [Online]. Available: https://research.qut.edu.au/qutcds/wp-content/uploads/sites/257/2020/06/Coordinated_behaviour_seminar.pdf

ABOUT THE PROJECT

The Programme on Technology and Democracy investigates the use of algorithms, automation, and computational propaganda in public life. This programme of activity is backed by a team of social and information scientists eager to protect democracy and put social data science to work for civic engagement. We are conducting international fieldwork with the political consultants and computer experts who are commissioned to activate or catch information operations. We are building original databases of incidents and accounts involved in such activities, and we use our knowledge to make better tools for detecting and ending interference with democracy. We engage in "real-time" social and information science, actively disseminating our findings to journalists, industry, and foreign policy experts. Our network of experts helps civil society, industry, government, and other independent researchers develop a better understanding of the role of technology in public life.

AUTHOR BIOGRAPHIES

Marcel Schliebs is a Researcher at the University of Oxford and social data scientist at the Programme on Democracy and Technology. His research is located at the intersection of political science, statistics and computer science, and focuses on the effects of disinformation and microtargeting on political attitudes and behavior. He has developed quantitative approaches for examining state-backed information operations, and further studies the role of artificial intelligence for 21st century great power competition. Marcel holds a BA in Political Science from Zeppelin University and a MSc in Social Data Science from the University of Oxford. In the past, he has worked as a Junior U.S. Correspondent for a German Public TV/Radio Broadcaster, at the French National Election Study, and served in the German Foreign Office and NATO's Arms Control and Weapons of Mass Destruction Non-Proliferation Centre.

Hannah Bailey is a researcher at the Programme on Democracy and Technology with a focus in social data science. Her research focuses on the PRC's use of state-sponsored digital disinformation. In particular, she focusses on the effect of the PRC's digital disinformation campaigns on international audiences by assessing how they interact with this disinformation. She holds a BSc in Politics and Philosophy from the London School of Economics, as well as two MScs, in Contemporary Chinese Studies, and in the Social Science of the Internet, both from Oxford University. She has also studied Mandarin at Fudan University (Shanghai).

Jonathan Bright is an Associate Professor, Senior Research Fellow at the Oxford Internet Institute who specialises in computational approaches to the social and political sciences. He has two major research interests: exploring the ways in which new digital technologies are changing political participation; and investigating how new forms of data can enable local and national governments to make better decisions.

Philip N. Howard is a professor and writer. He teaches at the University of Oxford and directs the Programme on Democracy and Technology. He writes about information politics and international affairs, and he is the author of ten books, including The Managed Citizen, Pax Technica, and Computational Propaganda. He has won multiple best book awards, and his research and commentary writing has been featured in the New York Times, Washington Post, and many international media outlets. Foreign Policy magazine named him a "Global Thinker" for 2018 and the National Democratic Institute awarded him their "Democracy Prize" for pioneering the social science of fake news. He has testified before the US Senate, UK House of Parliament, and European Commission on the causes and consequences of fake news and misinformation. His latest book is Lie Machines: How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives. He tweets from @pnhoward.

ACKNOWLEDGMENTS

We are particularly grateful to Erika Kinetz from the Associated Press (AP) Global Investigations Team for her invaluable contributions. Furthermore, we thank Franziska B. Keller, David Schoch, and three anonymous reviewers for their review and useful feedback on this manuscript.

The authors gratefully acknowledge the support of Craig Newmark Philanthropies, the Economic and Social Research Council (UKRI Grant Number 2260175), Ford Foundation, and Luminate.

Project activities were approved by the University of Oxford's Research Ethics Committee, SSH_OII_CIA_20_041. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the University of Oxford, the Oxford Internet Institute, or our funders.

CITING THIS WORKING PAPER

Please cite this working paper as:

Marcel Schliebs, Hannah Bailey, Jonathan Bright, Philip N. Howard. " China's Inauthentic UK Twitter Diplomacy: A Coordinated Network Amplifying PRC Diplomats." Working Paper 2021.2. Oxford, UK: Programme on Democracy and Technology, Oxford University, 2021. https://demtech.oii.ox.ac.uk/china-public-diplomacy-casestudy-uk. 35 pp.

APPENDICES

A.1 Data Collection, List of Included Accounts, and Descriptive Statistics

As part of this research project, we collected all tweets by PRC diplomats based in the UK over an eight month period from the 9th of June 2020 to the 31st of January 2021, as well as all retweets and replies to one of the target accounts. The data collection was conducted using the Twitter Streaming API and Facebook's CrowdTangle API. Data collection was interrupted for several hours on 6 December, 13 January, and 11 February due to power outage in the University of Oxford's computing center. Because of these outages, we estimate that we captured 99% of the activity shared from the Twitter API. Furthermore, the API is known to sometimes exhibit slight under-coverage, meaning that a small share of tweets or retweets may not be included in data from the Streaming API. However, the impact of this on sampling is not fully understood. It is likely that our estimates are conservative and that the findings and implications are not impacted by these small uncertainties.

Also, we do not capture engagement with older tweets that were engaged with after the data collection window ended on the 31st of January 2021. Due to the design of the Twitter Streaming API, quote-tweets of the PRC are not included in the systematic data collection. The examples of over half of the coordinated accounts verbatim quote- and reply-tweeting to the ambassador were thus collected by accident because they quoted tweets that included a state-baked media URL.

In this detection report, we have carefully sought not to disclose any private data by genuine individuals, including neither real names or identities nor any still active account. Instead, we include account information only for accounts who we assess were part of this operation and were suspended by Twitter for platform rule violations. Approximately half of these suspensions happened gradually over the second half of 2020, while the other half occurred after we shared the user ids and handles with Twitter on the 28th of April 2021.

The tweet and user ids will be made available in accordance with Twitter's data sharing policy as well as the Oxford University Research Ethics guidelines (CUREC). The complete R and Python code used to collect data and produce all statistics, figures, and tables will be released alongside this publication. See the project website.

Table 2 on the subsequent pages includes account-level data for all metrics or categories of patterns developed and applied for this paper. It includes the following features and variables:

- **Handle**: User handle
- **Account Created**: Account creation date
- **Status 1st of March**: Account status on the 1st of March. (Note: All active accounts were later suspended after reported by us to Twitter)
- **Times Amplified PRC Diplomat**:
- **Share Dedicated to UK-based (in%):** Share of all PRC diplomat amplifications dedicated to UK-based diplomat
- **UK-based Diplomat Retweets:** Number of retweets of UK-based PRC diplomats
- **Share Main Tweets Retweeted (in %):** Share of main tweets by ambassador retweeted (excluding threads) The relative share is based on all tweets that the ambassador made during the time the account was active
- **Share Thread Tweets Retweeted (in %):** Share of thread tweets by ambassador retweeted
- **Med. Lag between Tweets (in s)**: Median lag time between two consecutive retweets in seconds
- **Sequential Coordination Patterns**: Includes 🔄 for accounts which frequently co-amplify diplomat in a sequential pattern with other accounts in the network
- **UK-based Diplomat Replies:** Number of replies to UK-based PRC diplomats
- **Share Replied to (in %):** Share of ambassador's tweets replied to by account
- **Share Overlapping Language (in %):** Share of replies to ambassador including overlapping language patterns
- **Language Patters**: Includes 📣 for accounts who frequently use distinctive shared phrases
- **Verbatim Quote-Replying**: Includes “” for accounts who engage in verbatim quote- and reply-tweeting
- **Other**:
  - 1) Created on same day, maritime references and overlapping following network
  - 2) Matching the name of another user except for one character or digit

Table 2: Full List of All 62 Accounts Included in the Coordinated Network, With Distinctive Features

Handle	Account Created	Status 1st of March 2021	Times Amplified PRC Diplomat	Share Dedicated to UK-based (in%)	UK-based Diplomat Retweets	Share Main Tweets Retweeted (in %)	Share Thread Tweets Retweeted (in %)	Median. Lag between Tweets (in s)	Sequential Coordination Patterns	UK-based Diplomat Replies	Share Replied to (in %)	Share Overlapping Language (in %)	Language Patters	Verbatim Quote-Replying	Other
@litaoinlondon	2020-08-26 08:02	active	1386	98	1085	94	70	14		301	28	9	📣		1)
@JoeParker135	2020-08-11 15:11	active	1311	100	528	47	5	220	🔄	783	71	33	📣	“”	
@whwmaritime	2020-08-26 08:21	active	1184	100	1153	96	85	4		31	3	23			1)
@gyagyagya10	2019-12-06 14:45	active	1183	100	176	8	18	26	🔄	1007	82	19	📣	“”	
@Xiaojin05484077	2019-10-20 18:03	susp.	1166	100	1160	93	83	3	🔄	6	1	17			
@pianotaotao	2020-08-11 17:34	active	1163	99	557	50	4	804	🔄	606	52	32	📣	“”	
@reagan0927	2016-11-16 21:34	susp.	1063	100	559	89	20	149		504	79	30	📣		
@chi63722148	2018-04-11 17:06	active	1042	100	894	55	54	4		148	10	39	📣	“”	
@alex_jackyellis	2019-06-06 08:57	active	965	100	773	80	39	4		192	22	33	📣	“”	
@SunnyWade6	2020-08-11 15:15	active	936	100	13	1	0	>3600	🔄	923	82	37	📣	“”	
@Gavin20forward	2020-02-20 09:27	susp.	874	100	464	75	19	183	🔄	410	68	32	📣	“”	
@luckycloud16	2020-08-11 12:59	active	838	100	831	85	35	7		7	1	14			
@springer000111	2019-10-21 13:46	susp.	828	100	715	86	20	26	🔄	113	15	23	📣		
@N**********	2017-10-19 20:13	deleted	820	100	669	74	22	4		151	19	23		“”	
@axer97964843	2019-10-26 18:32	susp.	795	100	694	80	21	21	🔄	101	14	28	📣	“”	
@gjslxh2020	2020-10-19 14:16	active	774	100	758	88	79	21		16	3	31	📣		
@WillforCPC123	2019-11-20 17:41	susp.	756	99	495	77	81	4	🔄	261	50	49	📣		
@Caterpi27848664	2020-04-23 11:35	susp.	668	100	667	57	42	3	🔄	1	0	0			
@coast59965488	2020-08-26 16:53	active	614	100	613	60	33	4		1	0	0			1)
@anne202016	2020-11-10 09:15	active	590	99	360	60	11	4		230	45	25	📣	“”	
@ladybug23758032	2020-04-23 11:37	susp.	566	100	566	53	30	3	🔄	0	-	-			
@JustineSchoeman	2017-12-24 12:25	active	557	100	39	3	1	9	🔄	518	39	40	📣	“”	
@wangqichn	2019-10-29 13:21	active	517	99	516	42	7	12	🔄	1	0	0			

Handle	Account Created	Status 1st of March 2021	Times Amplified PRC Diplomat	Share Dedicated to UK-based (in%)	UK-based Diplomat Retweets	Share Main Tweets Retweeted (in %)	Share Thread Tweets Retweeted (in %)	Median. Lag between Tweets (in s)	Sequential Coordination Patterns	UK-based Diplomat Replies	Share Replied to (in %)	Share Overlapping Language (in %)	Language Patters	Verbatim Quote-Replying	Other
@BECreat16942938	2020-11-06 19:47	active	493	100	401	76	45	6		92	14	26	📣	“”	
@WanQingInUK	2020-10-26 20:37	active	466	73	440	74	53	11		26	6	15	📣		
@xin88306170	2020-07-24 16:01	susp.	460	100	245	42	4	77		215	37	24	📣	“”	
@reliable_young	2020-05-26 09:17	susp.	444	87	396	31	17	59	🔄	48	5	15			
@LiminRicky	2015-08-15 19:36	susp.	423	100	267	24	2	71		156	15	33	📣	“”	
@Voiceof95626989	2020-08-04 16:22	susp.	390	100	340	56	19	53		50	8	24	📣	“”	
@ukeye3	2020-09-19 11:29	susp.	389	100	247	70	0	42		142	40	49	📣		
@Bumbleb75459847	2020-04-23 11:40	susp.	354	100	354	40	9	3	🔄	0	-	-			
@GraceUK5	2020-11-22 16:14	active	340	100	253	52	25	5		87	20	18	📣	“”	
@yankee_roger	2020-11-13 09:56	active	332	98	302	52	46	55		30	6	10	📣		
@londoneye826	2019-10-19 09:08	susp.	317	100	164	90	1	120		153	85	62	📣	“”	
@Hushpup16240621	2020-04-23 11:42	susp.	308	100	308	37	8	3	🔄		0				
@JenniferatUK	2020-08-11 13:42	susp.	278	100	192	95	0	29	🔄	86	45	42	📣	“”	
@HillRegent	2020-11-16 14:17	active	277	100	145	36	4	46		132	34	30	📣	“”	
@G**********	2020-06-11 11:34	active	265	91	264	28	21	89		1	0	0			
@UkJenniferin	2020-10-07 07:21	active	264	100	5	1	0	149.5		259	38	32	📣	“”	
@HarmonyLondon2	2020-11-29 21:49	active	261	98	136	34	9	136		125	38	24	📣		
@CJohn723	2019-12-09 15:03	susp.	248	100	109	60	0	210.5		139	76	35	📣	“”	
@X**********	2020-06-04 12:25	deleted	219	99	132	26	4	6		87	14	36	📣	“”	
@lxhmof123	2020-08-14 09:43	active	184	100	158	65	4	64		26	11	42	📣		
@MirrorChyan	2016-11-21 03:12	active	165	99	3	0	0	>3600	🔄	162	19	9	📣		
@ruirui64237686	2019-08-21 12:54	susp.	160	100	60	11	0	9		100	17	63	📣	“”	
@tower0826	2020-11-03 19:50	active	144	100	62	15	0	242		82	20	48			
@tea_for_tweet	2015-01-26 11:21	susp.	125	100	4	1	0	192	🔄	121	27	24			

Handle	Account Created	Status 1st of March 2021	Times Amplified PRC Diplomat	Share Dedicated to UK-based (in%)	UK-based Diplomat Retweets	Share Main Tweets Retweeted (in %)	Share Thread Tweets Retweeted (in %)	Median. Lag between Tweets (in s)	Sequential Coordination Patterns	UK-based Diplomat Replies	Share Replied to (in %)	Share Overlapping Language (in %)	Language Patters	Verbatim Quote-Replying	Other
@SwordMotherland	2020-02-13 06:17	susp.	109	100	106	99	0	4		3	1				
@Visione75692167	2019-11-05 10:17	susp.	95	100	64	9	0	>3600	🔄	31	4	23		“”	
@JohnMur79481722	2020-11-23 09:24	active	90	100	41	9	10	4		49	10	63	📣	“”	
@Diomedeidae10	2020-04-21 09:51	susp.	79	100	54	7	0	>3600	🔄	25	4	28		“”	
@DaggerShield	2019-10-30 17:47	susp.	65	100	44	6	1	>3600	🔄	21	3	38		“”	
@eagleey62611149	2020-10-29 13:41	active	44	100		0	0	-		44	13		📣		
@RayLister6	2020-11-22 17:54	active	44	100	32	13	0	4		12	2	92	📣		2)
@MoverShaker5	2020-08-04 16:20	susp.	38	100	19	7	0	>3600		19	7	21		“”	
@Story_of_Stone	2018-01-26 21:44	susp.	36	100	13	8	0	5		23	16	65	📣	“”	2)
@Foodfor35226217	2020-08-04 16:34	susp.	29	100	15	9	2	>3600	🔄	14	9	21		“”	
@awakeninglions	2020-07-17 15:57	susp.	27	100	18	4	0	>3600		9	2	11		“”	
@Crouchi27494110	2020-04-21 09:34	susp.	20	100	13	2	0	>3600	🔄	7	1	43		“”	
@HiddenD99075856	2020-04-21 09:42	susp.	17	100	12	2	0	>3600	🔄	5	1	60		“”	
@story_of_stone_	2020-09-15 17:37	susp.	16	100	1	0	0	-		15	4	69			2)
@RayLister7	2020-11-30 09:38	active	3	100		0	0	-		3	0				2)

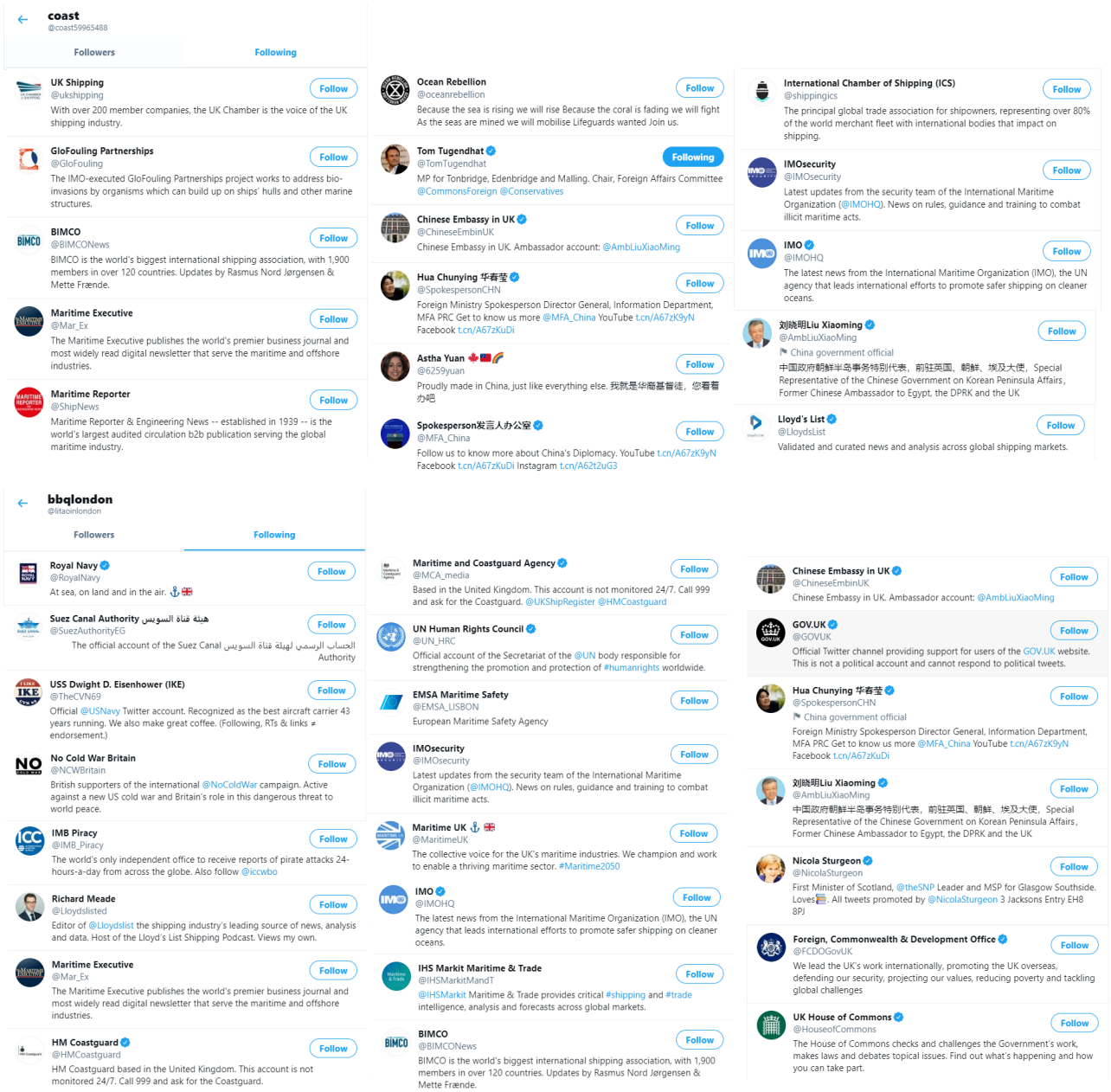
Note: The user handles of users who deleted their accounts themselves are censored for privacy reasons.

A.2 Additional Evidence for Coordinated Networks of Accounts

Figure 10: Three Accounts Created on the 26th of August 2020 With Unusual Water and Maritime References

Source: Author's screen captures.
Note: All three users were created on the 26th of August 2020, and reference water or maritime topics in their profile picture or name.

Figure 9: Accounts That Follow Unusual Maritime-Focussed Accounts



Source: Author's screen captures.

One distinctive pattern of the inauthentic coordination campaign was that some users alluded to their alleged UK-background through different means. Figure 11 shows an example of an account who described themselves as a London-based political commentator and used a picture by former Arsenal London footballer Thierry Henry as their profile picture. Other users, such as the ones displayed in Figure 12, suggested a UK-focus either by including the acronym UK in their username or by alluding to alleged events that had occurred "here in the UK".

Figure 11: Examples of Account Claiming to Be UK-Based

Source: Authors' screen captures

Figure 12: Examples of Accounts Exhibiting to a UK Focus

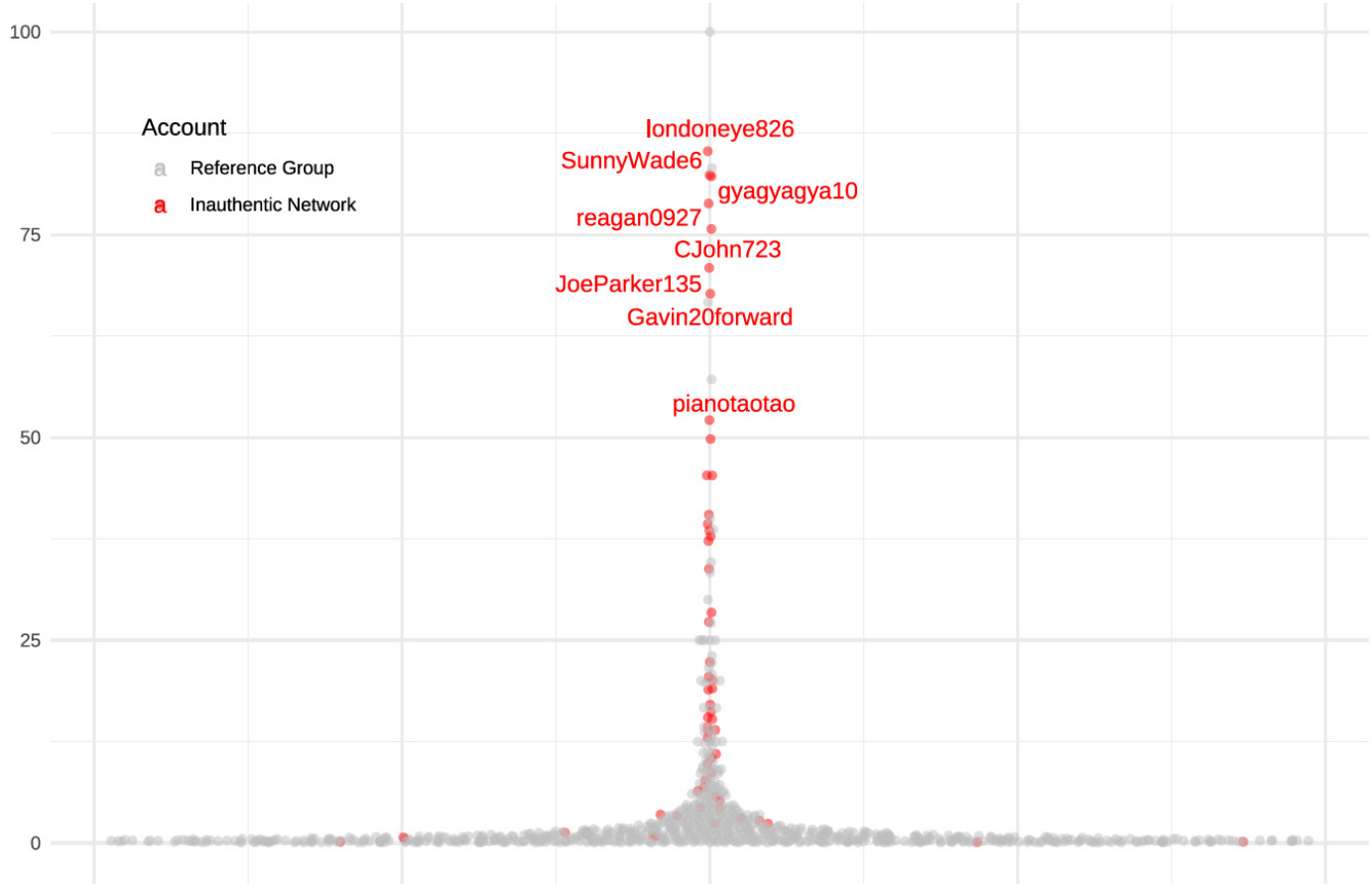
Source: Authors' screen captures.

One trace in the behavioral data that made the coordinated accounts in our network distinctive was that many of them interacted with nearly every tweet by the PRC ambassador to the UK. As visible in Figure 13, a large number of accounts in our cluster replied to a high share of the ambassador's tweets, while the majority of accounts in the reference group did so rarely if ever. Some users, many of which belong to a group created on the 11th of August, even replied to nearly or above three quarter of all tweets by the ambassador. In total, the sixty-two users in our network on average replied to 21% of all main tweets by the ambassador, while the 6,416 reference group users who replied to the ambassador on average replied to only 0.3% of his tweets, which is more than five standard deviations below the average reply quota for accounts in our network. Figure 14 further shows that a high share of these replies use the overlapping distinctive language patterns described in section 2.4.

A similar pattern can be observed not only for replies but also retweets. Figure 15 shows that inauthentic many of the sixty-two coordinated accounts were distinctive because they retweeted nearly every tweet that the ambassador authored while they were active.

The discrepancy between our coordinated network and the reference group is particularly strong for retweets not of the initial tweet from the ambassador's "thread tweets", but his subsequent thread tweets. Thread tweets are a series of connected tweets from one person. Figure 16 shows this pattern graphically.

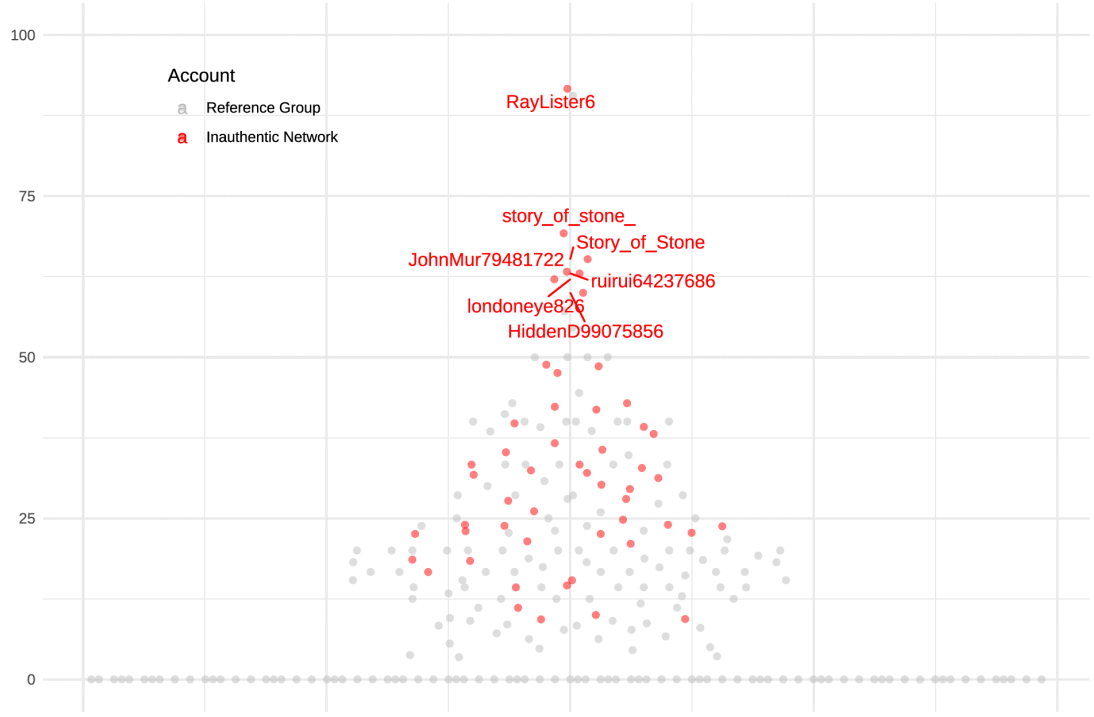
Figure 13: Share of Ambassador Liu Xiaoming's Tweets Replied to by Each Account (in percent)



Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Y-axis shows the shares of all main tweets (excluding threads) by the ambassador that a user replied to. Measures the relative share of replies compared to the total of ambassador tweets that a user could reply to, thus including all ambassador tweets while the user account was active. The accounts are scattered across the full range of the x-axis to show the illustrate the breadth of the distribution. Red accounts are from coordinated network and are labelled if the share is above 50%. The grey dots are the reference group of all other users engaging with the ambassador.

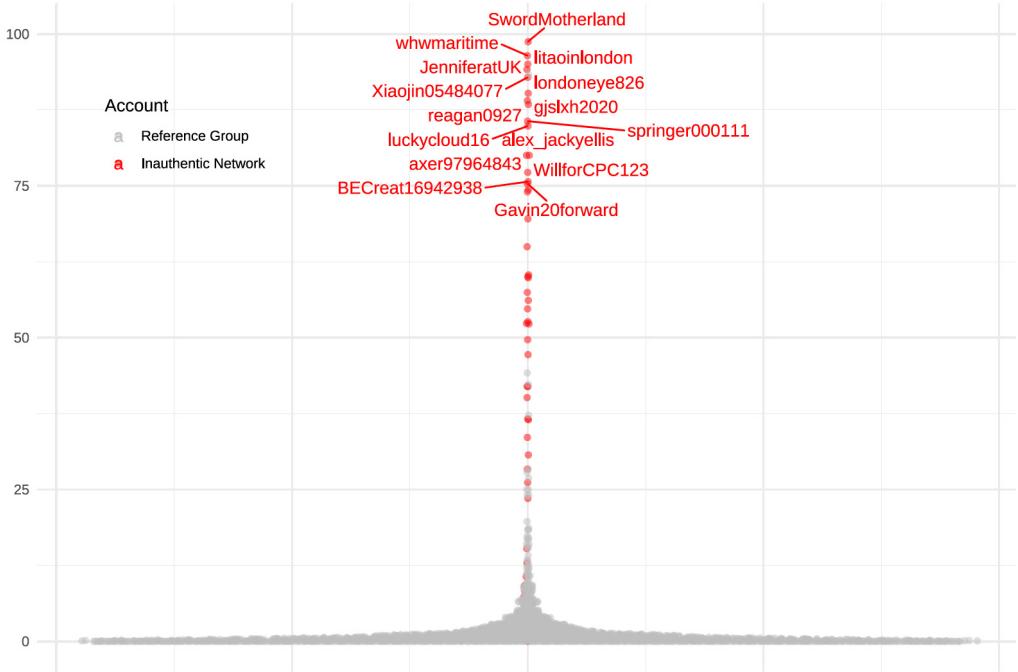
Figure 14: Share of Replies by a User Containing Overlapping Language Patterns Also Used by Diplomats (in percent)



Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Y-axis shows the shares of all replies which contained four, five, or six-word long sequences also earlier or later used by the ambassador. The accounts are scattered across the full range of the x-axis to show the illustrate the breadth of the distribution. Red accounts are from coordinated network and are labelled if the share is above 50%. The grey dots are the reference group of all other users engaging with the ambassador. For better visibility, figure includes users who replied at least 5 times.

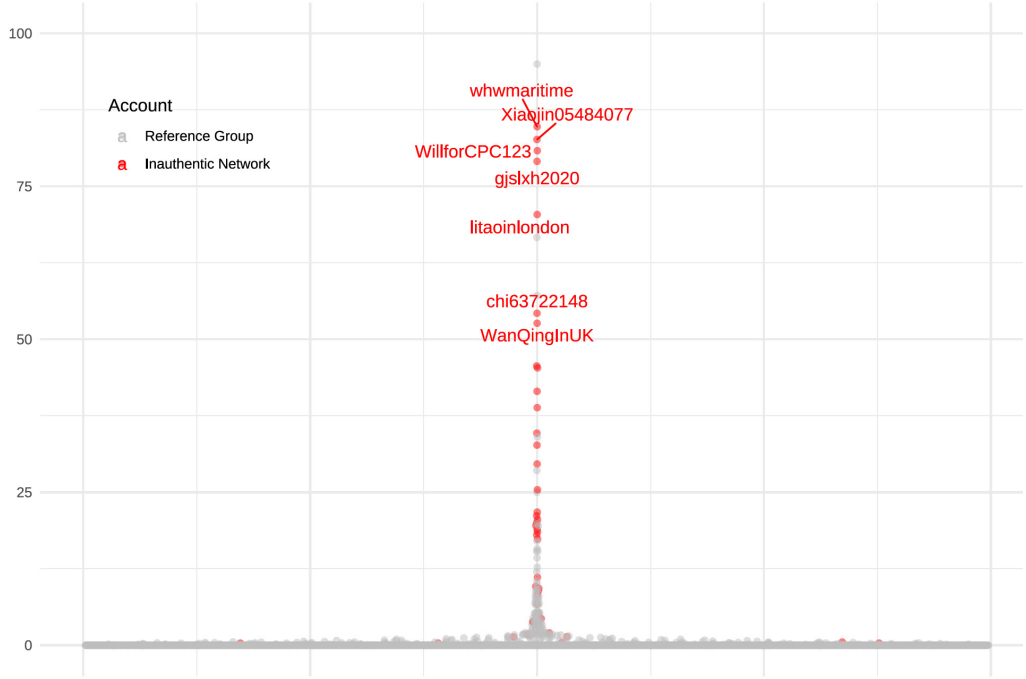
Figure 15: Share All Ambassador Main Tweets (No Thread Tweets) Replied to by User (in percent)



Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Y-axis shows the shares of ambassador main tweets (no thread tweets) replied to by a user. The accounts are scattered across the full range of the x-axis to show the illustrate the breadth of the distribution. Red accounts are from coordinated network and are labelled if the share is above 75%. The grey dots are the reference group of all other users engaging with the ambassador.

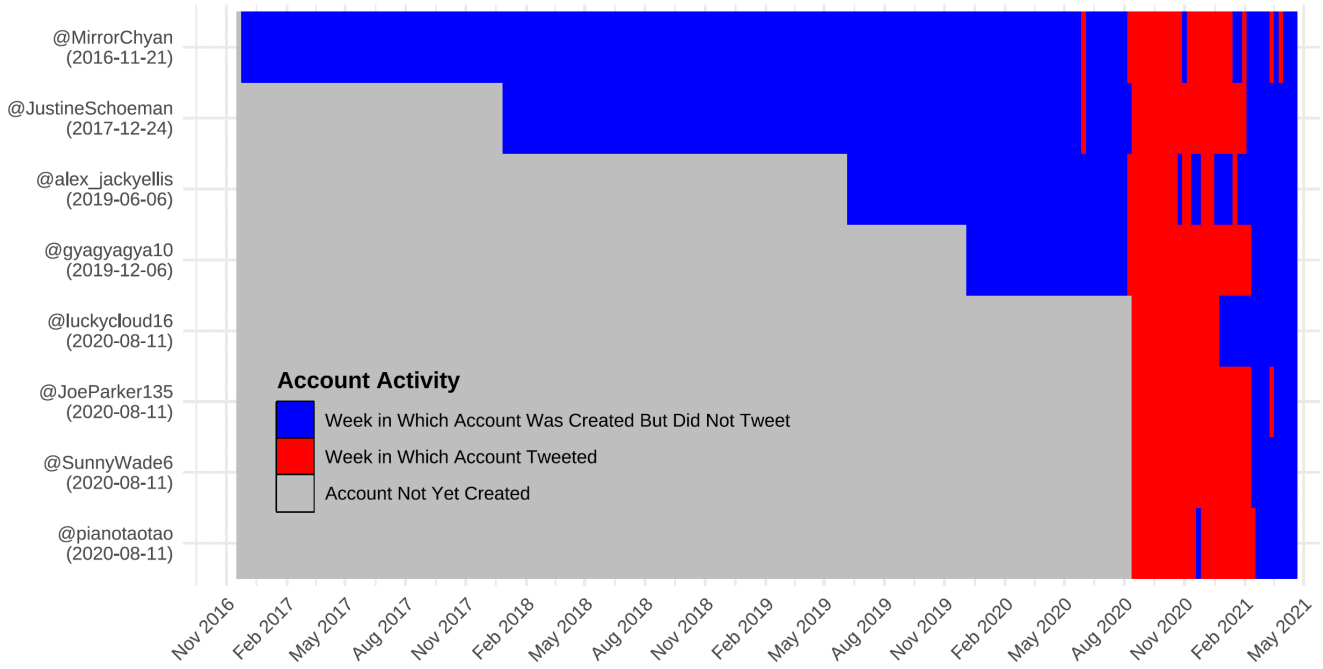
Figure 16: Share of Ambassador Liu Xiaoming's Thread Tweets that were Retweeted by All Accounts (in percent)



Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Y-axis shows the shares of ambassador thread tweets, not the initial tweet, but only subsequent tweets, retweeted by a user. The accounts are scattered across the full range of the x-axis to illustrate the breadth of the distribution. Red accounts are from the coordinated network and labeled if the share of retweeted thread tweets is above 50%. The grey dots are the reference group of all other users engaging with the ambassador.

Figure 17: Long Term Activity by Accounts "Waking Up" on the 12th or 13th of August 2020

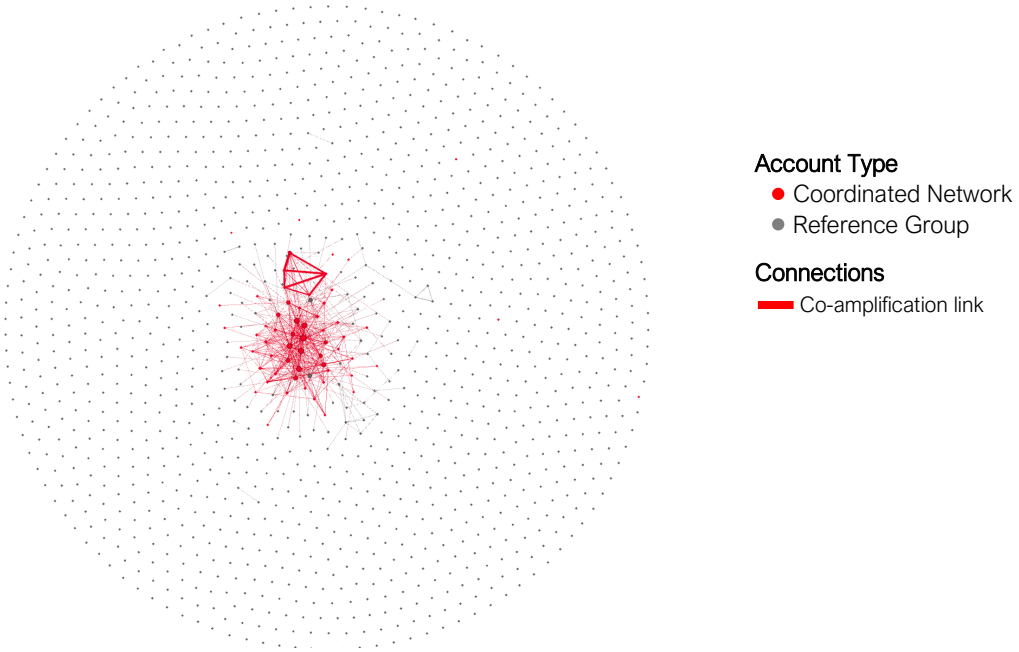


Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Red area represents weeks in which an account was active; blue areas are weeks in which an account had already been created but did not tweet or retweet anything; grey areas are times during which an account had not yet been created. Account creation date in parentheses.

Figure 2 in the main body shows a close view of a network that connects accounts if they have co-amplified PRC diplomats in the UK at least two times within sixty seconds of each other between June 2020 and January 2021. Figure 18 extends that perspective by including the full set of all accounts who amplified the ambassador, showing that an overwhelming majority of the other amplifiers in the reference group nearly never did so within sixty seconds of each other. Figure 19 focuses into the centre of the graph and provides user handles, allowing the reader to derive which users frequently co-amplified each other.

Figure 18: Full Set of Accounts and Co-Amplification Within 60 Seconds



Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Red dots are accounts in our coordinated network, grey dots are a reference group of other users amplifying the PRC diplomats. The dots are connected if they co-amplified a PRC diplomat in the UK within sixty seconds of each other at least twice. Larger ties represent more frequent occurrences of this pattern, which occurs over eighty times for some of the accounts in the sample, as opposed to zero times or once for most accounts in the reference group.

Figure 19: Network of Co-Amplifying Accounts That Retweet/Reply Within 60 Seconds of Each Other (Handles Labeled)

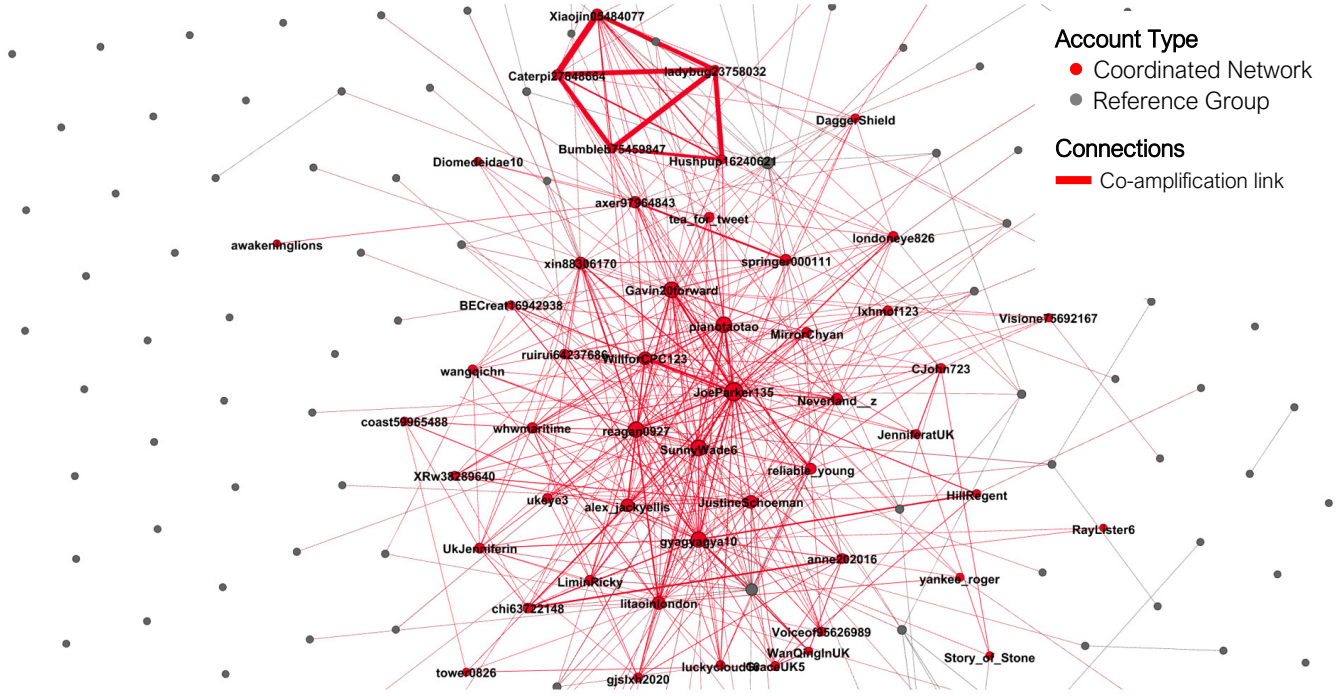
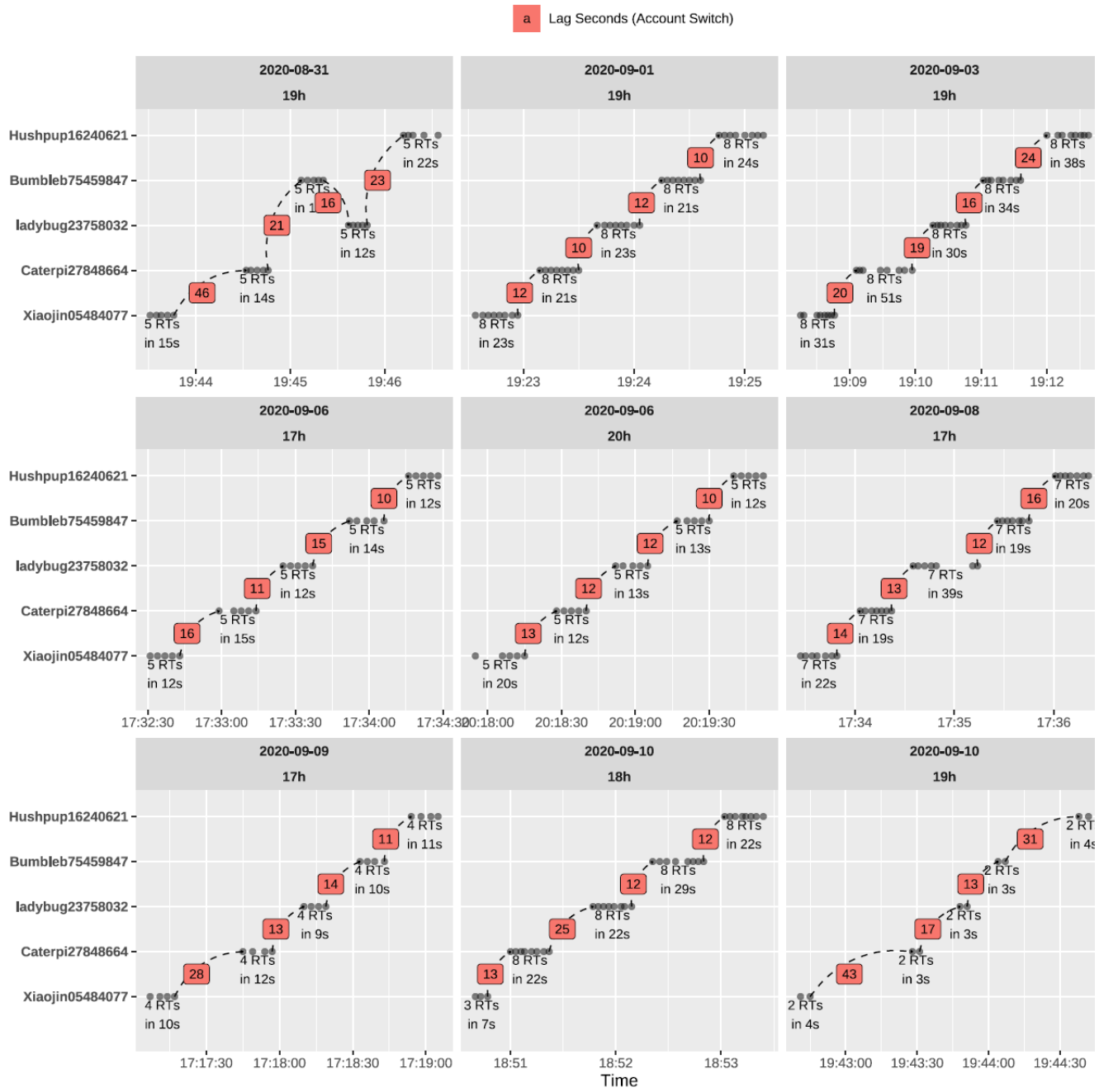


Figure 20 highlights short-term temporal coordination patterns. It shows retweeting patterns of five of the accounts in the network during a selected period of nine days in the beginning of September. On each day, presumably one single human operator logs on and retweets the ambassador several times (grey dots) in few seconds, before switching to the next account in a short amount of times (duration of account switch contained in red label). For example, the top row middle cell shows the period from 19:22 to 19:25 on the 1st of September 2020. Shortly before 19:23, @Xiaojin05474077 logs on and retweets the PRC ambassador to the UK eight times in twenty-three seconds. After twelve seconds, the next account picks up and retweets the same eight tweets in twenty-one seconds, and so on. This behavior pattern is sequential and occurs in the exact same order on multiple days between June and September 2020.

Figure 20: Sequential Retweeting on Selected Days and Hours in Early September 2020



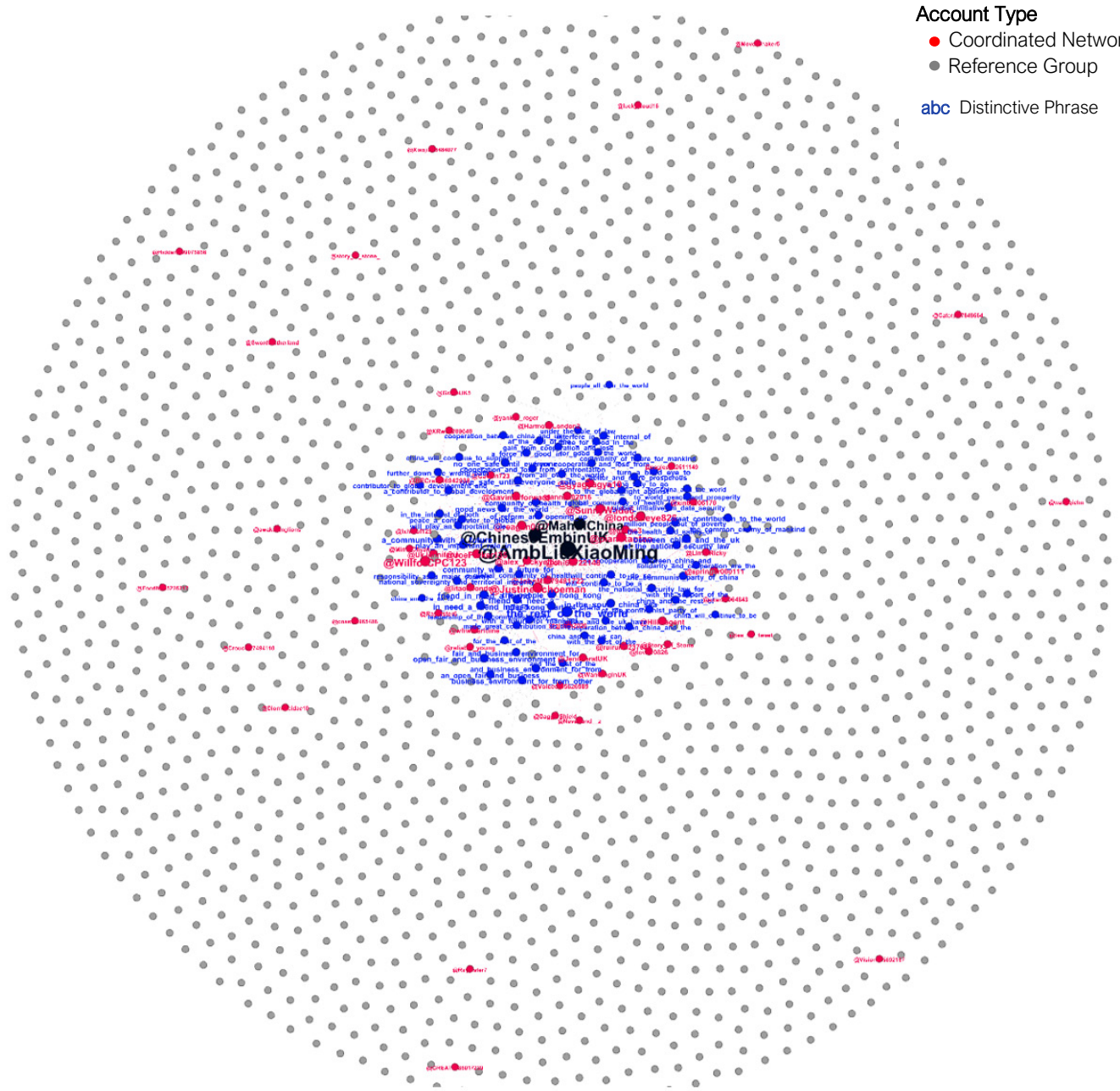
Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Figure shows retweeting patterns of five of the accounts in the network on selected days in the beginning of September. On each day, the human operator logs on and retweets the ambassador several times (grey dots) in few seconds, before switching to the next account in a short period of time (duration of account switch contained in red label).

For space and readability reasons, Figure 4 in the main body was limited to the ten most distinctive phrases. However, it is useful to showcase how distinctive this language use was for the overall coordinated network.

Figure 21 shows the full network of five word sequences which were used by at least five users when replying to the PRC diplomats. As visible in the figure, the large majority of red dots, which represent the accounts in our network, are clustered around the distinctive phrases. At the same time, nearly all the grey reference group accounts are spread in the periphery of the graph, indicating that they never used any of the distinctive phrases.

Figure 21: Full Set of 5-Word Sequences Used by All Accounts Engaging With the PRC Diplomats

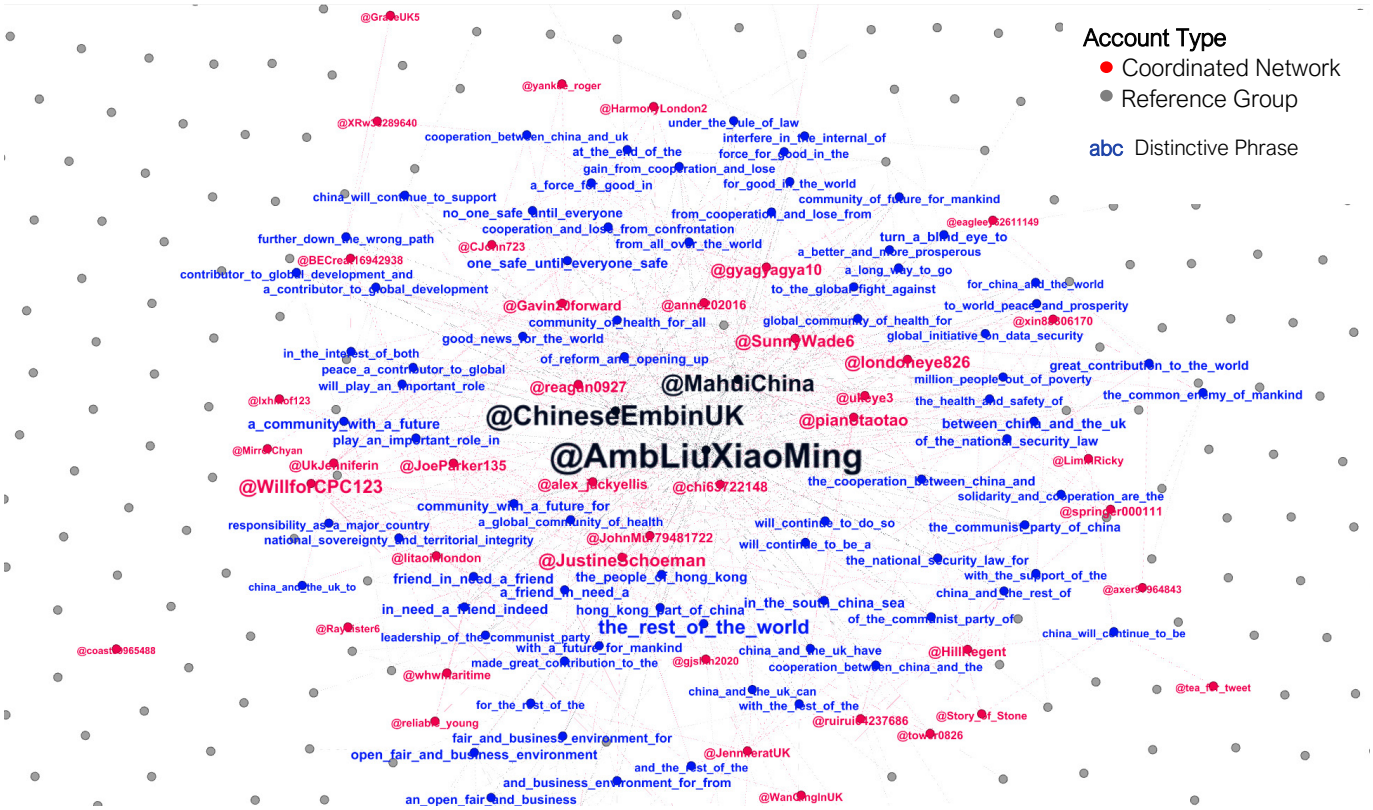


Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Figure shows phrases of at least five sequential words used nearly exclusively by accounts in the coordinated network and the UK-based diplomatic accounts. This behavior was distinctive to the coordinated network and did not occur in the reference group of all other users engaging with the diplomats.

Figure 22 zooms into the center of the previous figure, making the users and distinctive chunks of words readable. The Figure demonstrates that many of the inauthentic coordinated accounts frequently repurposed distinctive phrases and chunks of words from previous tweets by the diplomat accounts. Other users in the control group used these phrases only very rarely.

Figure 22: Network of Distinctive Five Word Phrases Being Used by Inauthentic Network and Diplomats

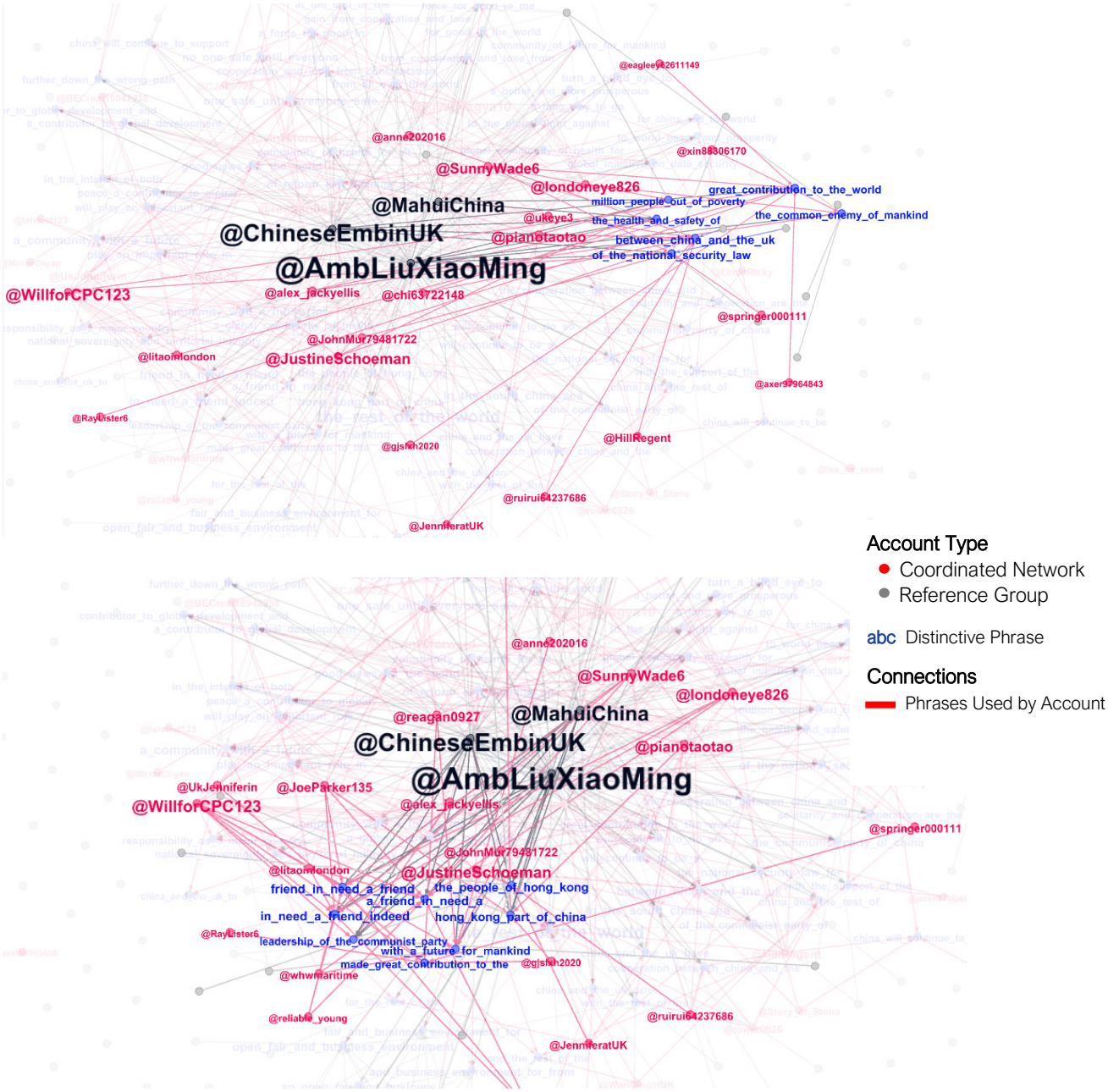


Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Figure shows phrases of at least five sequential words used nearly exclusively by accounts in the coordinated network and the three UK-based diplomatic accounts. This behavior was distinctive to the coordinated network and did not occur in the reference group of all other users engaging with the diplomats.

Figure 23 shows that the several highly distinctive phrases, such as "million people out of poverty", "(china has made) great contribution to the world", or "a friend in need is a friend indeed" were used by a large share of the accounts in our inauthentic network, and very rarely by any other account in the reference group (grey dot). Nearly all these phrases were also contained in earlier tweets by the accounts of PRC diplomats.

Figure 23: Accounts from Coordinated Network Using Subset of Overlapping Phrases



Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: This graph shows highlighted subsets of the network. All red accounts in the coordinated network have used the highlighted blue phrases multiple times, while nearly no other users from the reference group of all other users engaging with the ambassador have.

Figure 24 shows that the several subgroups of accounts exhibited very high overlap in the distinctive phrases they used. For example, @SunnyWade6 and @pianotaotao are both accounts which were created on the 11th of August. Similarly, two users called @londoneye and @ukeye each used numerous matching patterns that nearly no reference group users ever used.

Figure 24: Distinctive 5-word Phrases Used by Small Example Subgroups of Accounts

Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Lines between users with the distinctive phrases they used.

Figure 25 shows one of numerous examples in which multiple accounts, in this case fifteen, replied and quote-tweeted to the same tweet by the PRC ambassador with matching verbatim phrasing. In total, more than half of the accounts in the inauthentic network engaged in this behavior at least once. While we did not observe an instance where an account from the large reference group engaged in the same type of behavior.

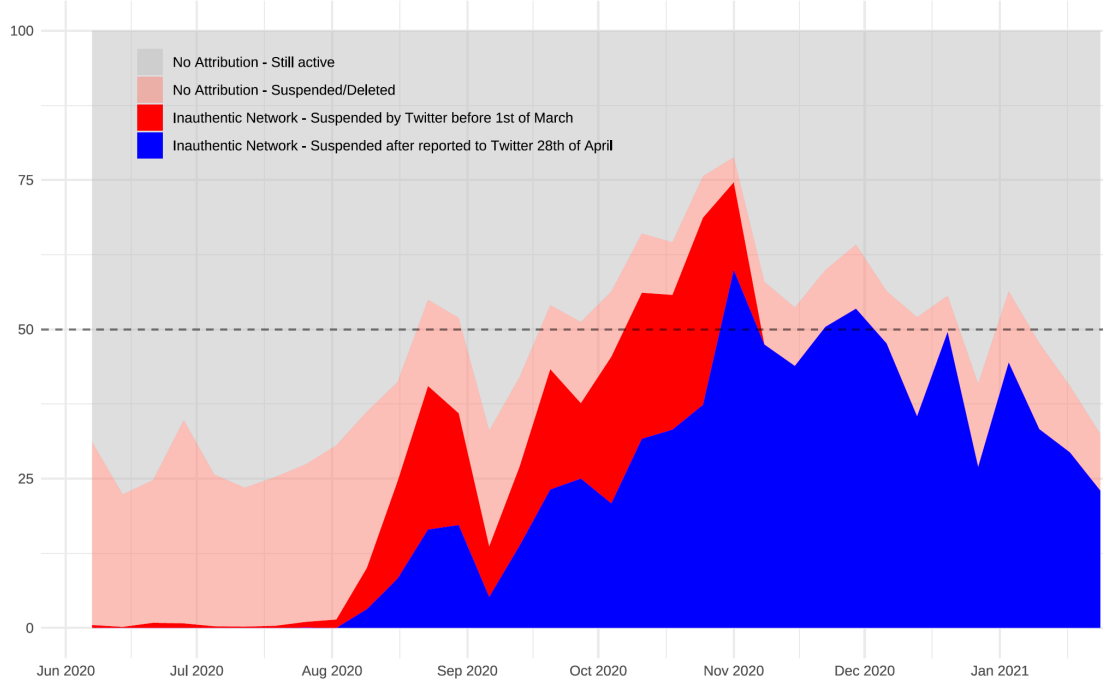
Figure 25: Verbatim Quote- and Reply-Tweeting to a Single Tweet by a PRC Ambassador

Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: The blue tweet is the original tweet, the subsequent lines represent the verbatim replies and quote-tweets authored by the accounts in the coordinated network.

Figure 26 and Figure 27 complement the results in section 3 with the share of weekly replies. As visible by the blue and solid red areas in the figures, the coordinated inauthentic network accounted for up to three quarters of weekly replies to the ambassador in November 2020. Over the whole observation window, 20% of all replies to the ambassador were attributed to the inauthentic network of sixty-two account.

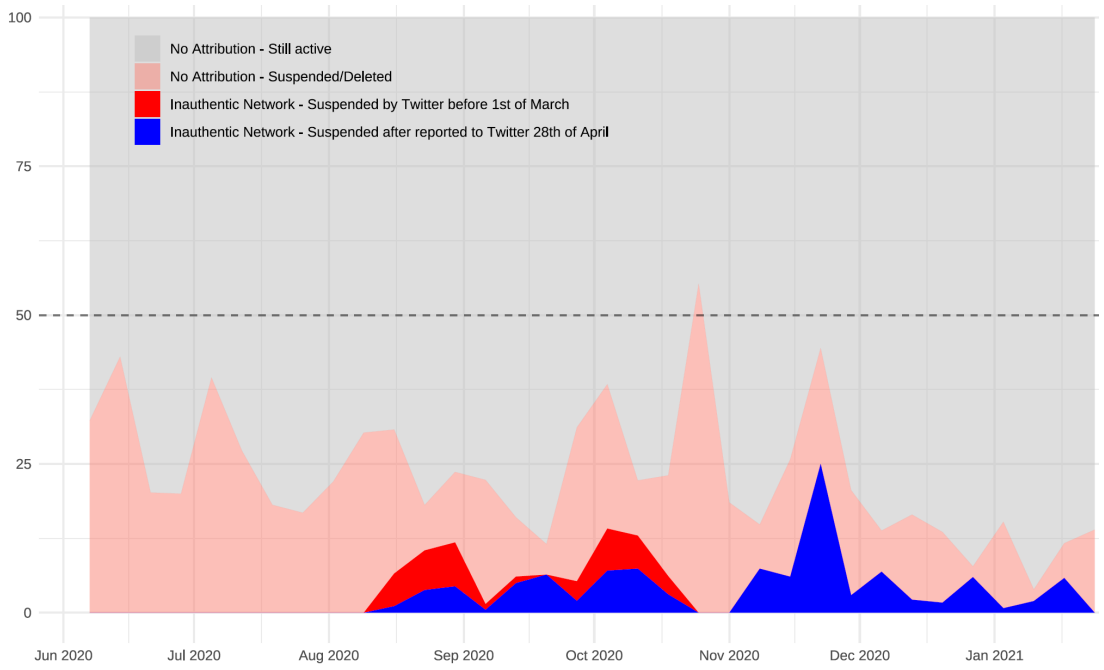
Figure 26: Share of Weekly Replies to Ambassador (@AmbLiuXiaoMing) Attributed to Coordinated Network (Percent)



Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

Note: Blue solid area represents retweets by accounts still active before we flagged them to Twitter; solid red areas represent accounts in our detected operation which Twitter had already suspended. Light red represents accounts suspended by Twitter which were not part of our coordinated network.

Figure 27: Share of Weekly Replies to Embassy (@ChineseEmbinUK) Attributed to Coordinated Network (Percent)



Source: Authors' calculations based on data collected between the 9th of June 2020 and 31st of January 2021.

The Programme on Democracy & Technology

at the Oxford Internet Institute

University of Oxford

1 St Giles • Oxford OX1 3JS

Website: www.demtech.oii.ox.ac.uk