CASM
technology

#IStandWithRussia #IStandWithPutin

Message-based Community Detection on Twitter

BACKGROUND

Across March 2nd and 3rd, two pro-invasion hashtags began to trend on Twitter across a number of geographies around the world: #IStandWithPutin and #IStandWithRussia.[1]

In the days that followed, research began to be made public that suggested that some of the activity associated with these hashtags were inauthentic. On March 3rd, Marc Owen Jones released research arguing that both automated 'bot' accounts and engagement farming had been used to drive engagement with the hashtag.[2] The next day, he released a further investigation, noting the crossover in profile pictures between those used on accounts engaged in pro-invasion messaging and dating scams.[3]

An online researcher called Conspirador Norteño raised a number of suspicious account-level patterns, including that many of the accounts sending the hashtag were created on the day of Russia's invasion, that accounts sharing the hashtag formed a dense Retweet cluster (implying overlapping sharing activity) and that the hashtag #IStandWithPutin was 'virtually nonexistent' before March 1st.[4]

A fourth piece of research on the hashtags from the Atlantic Council's Digital Forensic Lab concluded: "in this instance, a hashtag supporting the military invasion of a sovereign state shows evidence of being artificially gamed to give the impression of support for Putin and Russia."[5] On 4th of March, it was reported that Twitter removed a number of accounts that had pushed these pro-invasion hashtags for 'coordinated inauthentic behaviour'.[6] We also conducted our own preliminary analysis of the hashtag, broadly agreeing with the other research cited here.[7]

[1] The Times of London reported the hashtags to be trending in India on March 3rd, within a piece more broadly concerned with India's reticence in condemning Russia's invasion of Ukraine. https://www.thetimes.co.uk/article/why-india-has-not-spoken-out-against-invasion-pgrrx3fsr?utm_medium=Social&utm_source=Twitter

[2] https://twitter.com/marcowenjones/status/1499312091727020032?s=20&t=ywex_N7QS3zdNoAsX8WyyA

[3] https://twitter.com/marcowenjones/status/1499666250225664001

[4] https://twitter.com/conspirator0/status/1499498721964351491?s=20&t=8kg_-4PKzXhgW5xRzqYIXQ

[5] https://medium.com/dfrlab/istandwithputin-hashtag-trends-amid-dubious-amplification-efforts-2b8090ac9630

[6] https://www.nbcnews.com/tech/internet/twitter-bans-100-accounts-pushed-istandwithputin-rcna18655

[7] https://www.isdglobal.org/digital_dispatches/istandwithrussia-anatomy-of-a-pro-kremlin-influence-operation/

RESEARCH AIM

Given the emergent research finding evidence of suspicious activity associated with these hashtags, we set out to better understand the nature of the accounts that were engaging with them.

It should be noted that it was not the primary aim of our research to demonstrate that any particular account or set of accounts were automated ‘bots’ or otherwise inauthentic. Judgements such as these are difficult to definitively make with the data available to independent researchers, and whatever conclusions can be appropriately drawn about the presence, or not, of inauthenticity across the accounts featured in this research had already - we felt - been made by other researchers cited above.

Second, this research did not seek to empirically attribute the activity that is being researched. Where this is possible at all, it often requires deeper, OSINT-driven investigative work beyond the data-driven appraisal of behaviour presented here. This research sought to empirically measure online behaviour at scale; and whilst we do offer thoughts around possible intent and strategy attributing the actors ultimately responsible for this activity remain matters of interpretation and inference.

We entered into this research with two premises. First that in researching accounts heavily engaging with the hashtags, we were likely to be seeing both inauthentic and organic activity. And second, that much of this activity was patently pro-Russian. Better understanding the underlying and pre-dating activity of both authentic and inauthentic actors in this context would, we reasoned, be helpful. Analysing inauthentic actors might help us understand more about the nature of the campaign, and analysing authentic actors more about the reception of messages that it was trying to propagate.

The aim was therefore to create a way of understanding account behaviour at scale, in a way that is not solely related to the hashtags themselves. Adding to the previous work that mapped accounts on the basis of their engagements with others, we aimed to map accounts using a representation of the language they used and shared. This might nuance and enrich our understanding of the kinds of political, cultural, ideational geopolitical or any other themes that the account engaged with online beyond direct engagement with the hashtags themselves.

As we explain below, to achieve this technically, we leveraged Transformer-based models that are capable of creating general-purpose representations of text. We decided to focus only on the accounts that had most intensively engaged with the hashtags, and to build representations for these accounts based on the most-recent 200 messages each had sent. These linguistic representations take the form of vectors with which we can spatially express the linguistic similarities between accounts. This allows us to identify groups of linguistically similar accounts as linguistically-defined communities, which can then be characterised through a combination of quantitative and qualitative means.

METHODOLOGY

DATA COLLECTION

On 4th of March, Twitter was backsearched using its Search API for all Tweets and Retweets containing either “#IStandwithPutin” or “#IStandwithRussia”.[8] An ongoing collection was also established using Twitter’s Streaming API[9] to collect new Tweets containing these hashtags as they were posted. By 7th March, a total of 349,779 Tweets (and Retweets) had been collected. 304,433 Tweets were collected from the Search API, dating back to 24th of February and 45,346 from the Stream API. Together, these messages were tweeted or retweeted by 128,597 distinct users.

In order to map accounts on the basis of similarities in the textual content of their messages, between 7-9th March we collected the most recent 200 Tweets[10] from the timeline of the 128,597 distinct users previously identified, for a total of 21,927,321 Tweets. 912 users were no longer publicly available at the point of collection, either due to user- or platform-side activity. This may range from the account being deleted by the user, being changed from ‘public’ to ‘private’, or be subject to enforcement action by Twitter.

We then filtered this new dataset to create a subset of accounts that had engaged with the hashtags in question more intensively. Accounts whose past 200 Tweets contained less than five separate messages containing either #IStandwithPutin or #IStandwithRussia (or both of them) were removed. This resulted in a dataset of 1,668,919 Tweets sent by 9,907 accounts, at the basis of the network construction presented in the next section.

ACCOUNT REPRESENTATION

In this work, we have undertaken a network analysis of the accounts in our dataset. Our goal was to group accounts in this network on the basis of similarity in the textual content of the messages they posted or amplified. This message-based approach contrasts with more traditional network mappings that group accounts based on friend-follower relationships, Retweet interactions, or @mentions. Instead, our intent was to identify communities of accounts defined by common linguistic attributes based not just on their activity explicitly related to the hashtags, but on their broader history of language-use on Twitter.

We make use of an established technique that provides an approximate measurement of the semantic similarity of messages. This involves mapping messages into a common (vector) space in which the similarity of texts can be compared numerically. To do this, we used what is known as a pre-trained sentence encoder.[11]

[8] https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets

[9] https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/api-reference/post-statuses-filter

[10] It should be noted that unlike most research on Twitter, this creates a dataset that does not fall within a particular bounded time-window. An account might have sent their 200th most recent Tweet last week, last year, or never.

[11] https://arxiv.org/abs/1908.10084

A pre-trained sentence encoder is an example of a pre-trained language model that is specifically optimised for measuring similarity between sentences. In the field of Natural Language Processing, pre-trained language models[12] are currently viewed as the most effective way of capturing certain aspects of language meaning, and act as building blocks that can be adapted to suit a broad range of language processing tasks.[13]

We encoded each of the 1.67M tweets using two different pre-trained sentence encoders;[14] *all-distilroberta-v1*[15] as a monolingual (English) model, and, given that we are working with a multilingual dataset, *paraphrase-multilingual-mpnet-base-v2*[16] as a multilingual model. This encoding process places each message into a 768-dimensional space, where a pair of messages with encodings (vectors) that are closely located in this space are taken to have similar meaning.

To compute the account-level representation for a given account, we aggregated (averaged) across the message-level representations (vectors) associated with that account in the dataset. This was performed for each of the 9,907 accounts.

ACCOUNT NETWORK CONSTRUCTION

We next constructed a network in which the nodes represent the 9,907 accounts, and (undirected) weighted edges represent the similarity between two accounts, with their weights representing the degree of similarity of message content. Specifically, the weight of an edge between two accounts is the cosine similarity of the account-level representations (averaged message vectors) of the two accounts.

We set a threshold on edge weights that preserved approximately 250,000 of the most highly weighted edges. For the monolingual encoder, applying this threshold resulted in 2,151 of the accounts being excluded from the network on the grounds that they were not sufficiently similar to any other account. For the multilingual encoder, 2,018 accounts were excluded on this basis.

We graphed the network using Gephi.[17] For positioning nodes, we applied the ForceAtlas2 layout algorithm,[18] a widely used approach to spatialize a weighted undirected network in two dimensions. For community detection, we applied the Louvain method, a modularity-based algorithm,[19] which assigns a single community to each node. Both encoders resulted in networks with over 100 detected communities with a long tail of small communities. Approximately 90% of accounts were assigned to one of 8 communities in the case of the monolingual-encoder network and 7 communities in the case of the multilingual-encoder network. The discussion below is limited to these prominent communities.

[12] https://arxiv.org/abs/1810.04805

[13] https://web.stanford.edu/~jurafsky/slp3/11.pdf

[14] Tweets were pre-processed, replacing @mentions of accounts with "@user" and links were replaced with “http”.

[15] https://huggingface.co/sentence-transformers/all-distilroberta-v1

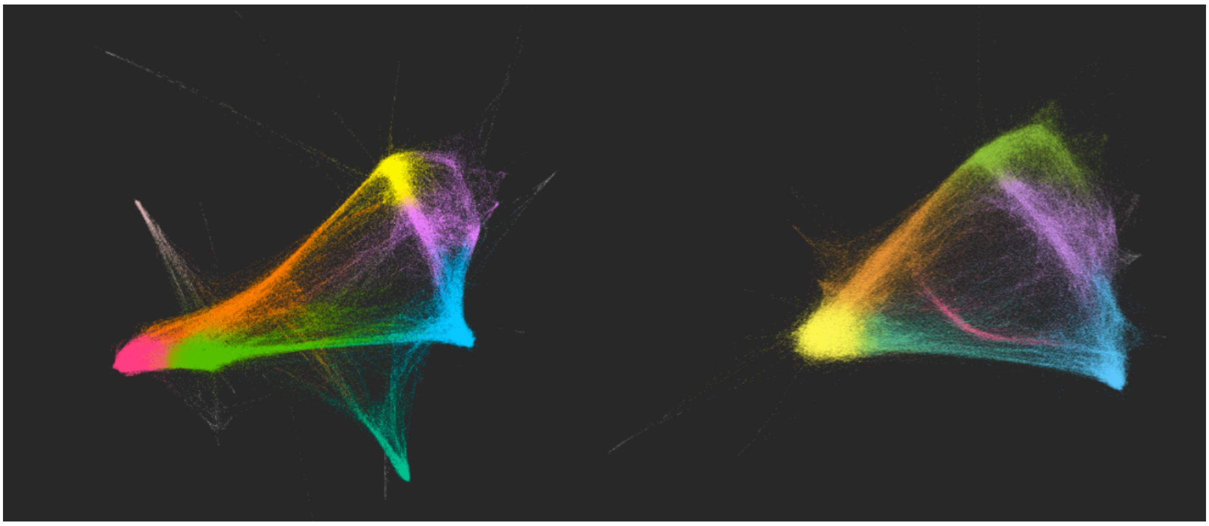
[16] https://huggingface.co/sentence-transformers/paraphrase-multilingual-mpnet-base-v2

[17] https://gephi.org/

[18] https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098679

[19] https://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008

The figure below shows the network structures for the monolingual-encoder network (left) and multilingual-encoder network (right). Note that the colours in both of these networks are only used to visually distinguish boundaries between each community, but are not aligned between networks. For a more detailed comparison of the specific community distribution between these networks, please see below.

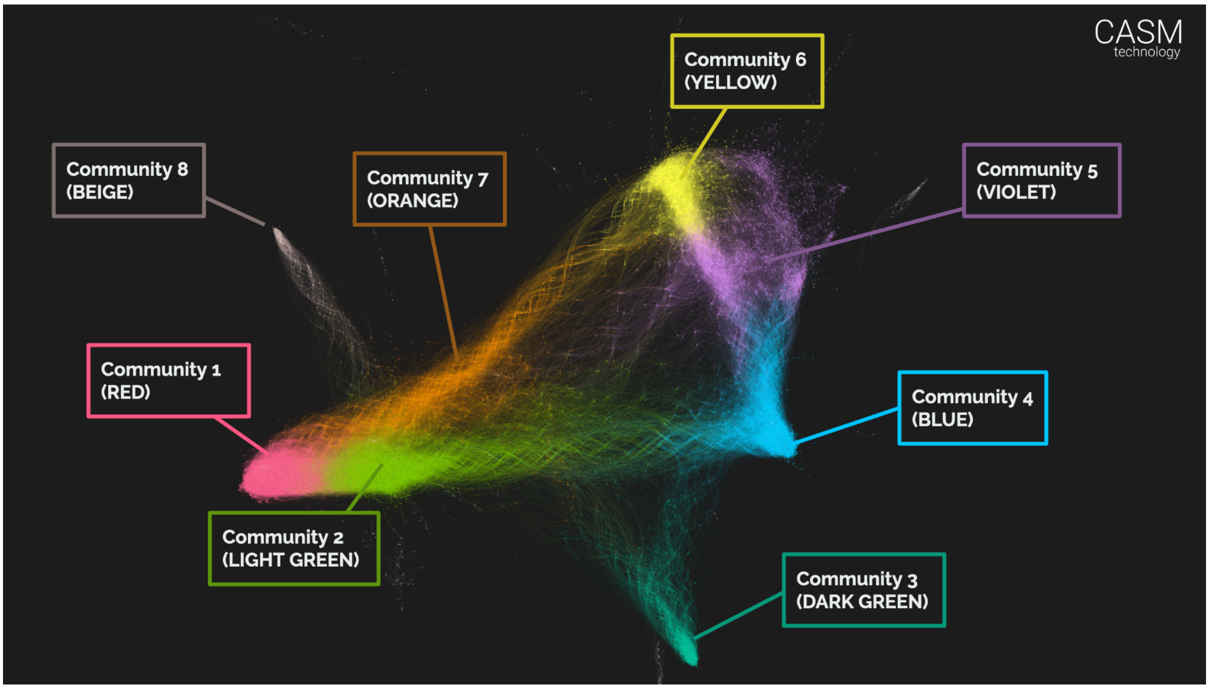


Mono- and multi-lingual model network representations

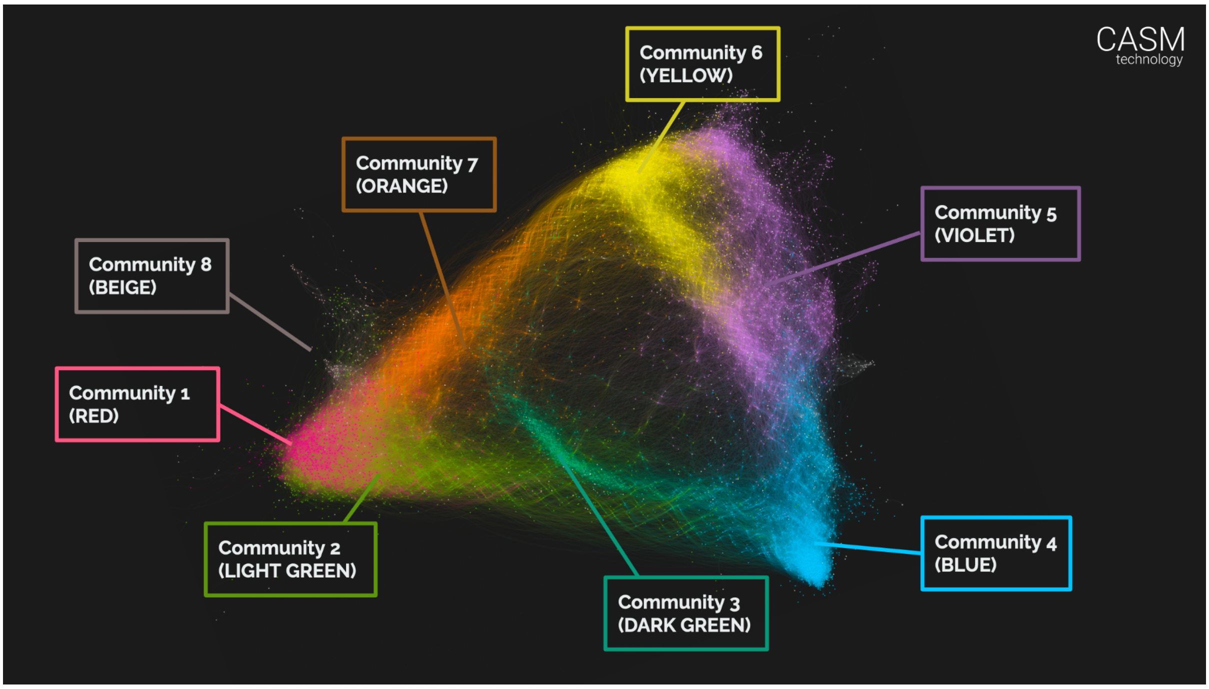
COMPARISON OF MONO- AND MULTI-LINGUAL ENCODERS

Initially, manual analysis was applied to the monolingual-encoder network. As is discussed in greater length in the next section, community identity was largely based on the mix of natural languages involved, but also suggests some geographically-specific political affiliations.

After initial characterisation of the clusters, but before the full analysis was conducted, the community segmentations of the two encoders were compared to assess whether they segmented communities on a similar basis. The figure below, i.e. the network produced with *paraphrase-multilingual-mpnet-base-v2*, shows the multilingual-encoder network coloured by the account-level community identities detected in the monolingual-encoder network. As can be seen, there is a close alignment across the two networks in terms of the communities discovered. This suggests that the most salient attributes in both networks relate to language or language-driven identities.



Network produced with all-distilroberta-v1



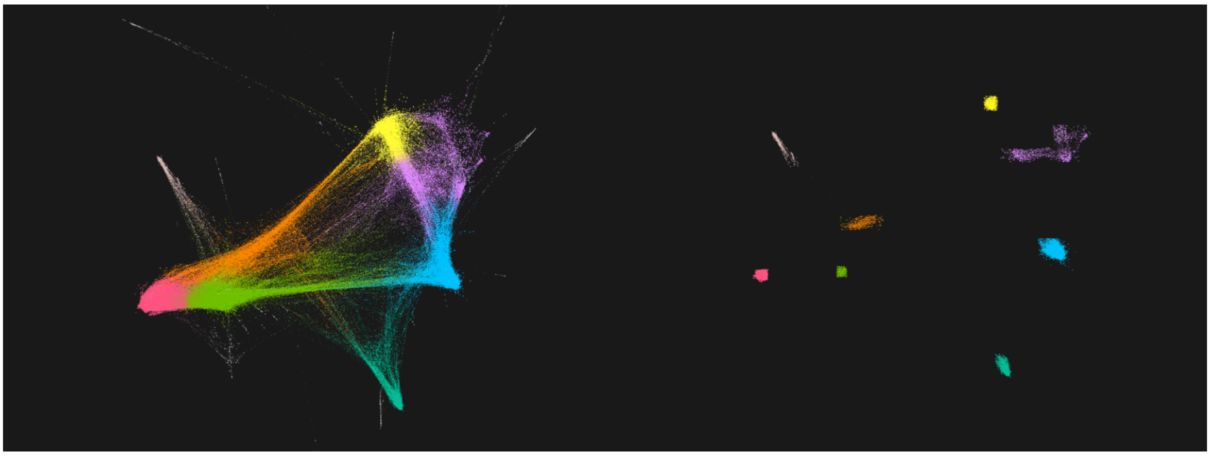
Network produced with paraphrase-multilingual-mpnet-base-v2

Given that the two encoder models have been trained on such different data, it is perhaps surprising how close the two models were in terms of their high-level segmentation of accounts. One notable difference is that the RED and LIGHT GREEN communities become more entwined in the multilingual-encoder network. This is intuitive, given that (as detailed later in this report) both communities are composed of accounts that primarily identify as Indian, and who engage in a number of common themes related to Indian politics, but are distinguished by the level of Hindi-useage.

As we began to examine samples of accounts within each cluster, it became apparent that the mix of natural languages associated with a cluster was of particular interest when characterising behaviour and putative identity. As a result, the analysis below is based on *all-distilroberta-v1*.

COMMUNITY CHARACTERISATION

The next step in the process was to manually inspect accounts from each community to attempt to identify the attributes they had in common with each other. To do this, we randomly sampled 100 accounts from the core of each community, where we consider the core of a community to be a dense region around the community's 'centre', as suggested by the network's spatial disposition. An illustrative example below shows the cores of the monolingual-encoder network that were then randomly sampled from.



The manual characterisation of accounts was conducted by three independent analysts, who, between them, appraised these 100 randomly sampled 'core' accounts from each cluster. The principal aim of this undertaking was to allow an open-ended and unstructured analysis to surface the salient features of each community.

Features analysed included:

- The profile picture of the account (especially the presence of motifs, tropes, regional or national identifiers);
- The profile description of the account (interests, hobbies, political or ideological attachments);
- The number of followers of the account;
- The number of accounts that the account follows;
- The number of Tweets sent by the account;
- The Retweet:Tweet ratio of the account

In addition, the analysts read between 100-250 messages of each account's timeline to create a narrative summary of the content that the account had either originally sent or amplified. This included the language/s used by the account, and, where possible to discern, the thematic, regional, temporal and ideological features of the messages themselves.

Once this exercise was completed for the sampled accounts for each cluster, analysts sought to identify the key attributes that were held most in common by accounts grouped within each cluster. A number of attributes were consensually identified:

- A large number of accounts had profile descriptions and pictures which located them to certain geographic vicinities;
- A large number of accounts sent messages about events and issues that were geographically specific;
- A large number of different languages were identified in the messages shared by the sampled accounts.

It was also noted that the community detection process, as described above, has identified clusters that were broadly delineated along regional, national and/or linguistic lines. Cluster characterisation proceeded on this basis.

RESULTS

CLUSTER CHARACTERISATION

Below we present the characterisation of each cluster. This combines per-cluster aggregated metrics to describe measurable behaviour, with qualitative narratives composed through the manual analysis process outlined above.

These characterisations are necessarily interpretative in nature, of course, and therefore simply an expression of the data but also the judgements, biases and predispositions of the authors. The top-level cluster descriptions do not imply that they are comprised exclusively of accounts from that country or region, nor that the languages perfectly map onto that region. Each one will contain 'noise' in the form of accounts that are from different countries, that use different languages and do not behave in the way that the overall descriptions of each cluster would suggest. Many of the clusters have important attributes (whether linguistic or thematic) which do not map onto their primary regional identities and all will contain accounts that do not conform to the salient characteristics of the cluster. Indeed, variety within a cluster is also a feature of it, such as Nigerian Edo State messages within a cluster which we broadly characterise as South African. This is a balance we try to strike between reflecting and representing the variety of accounts contained within clusters and applying an overall interpretative pattern to make the network more cognisable.

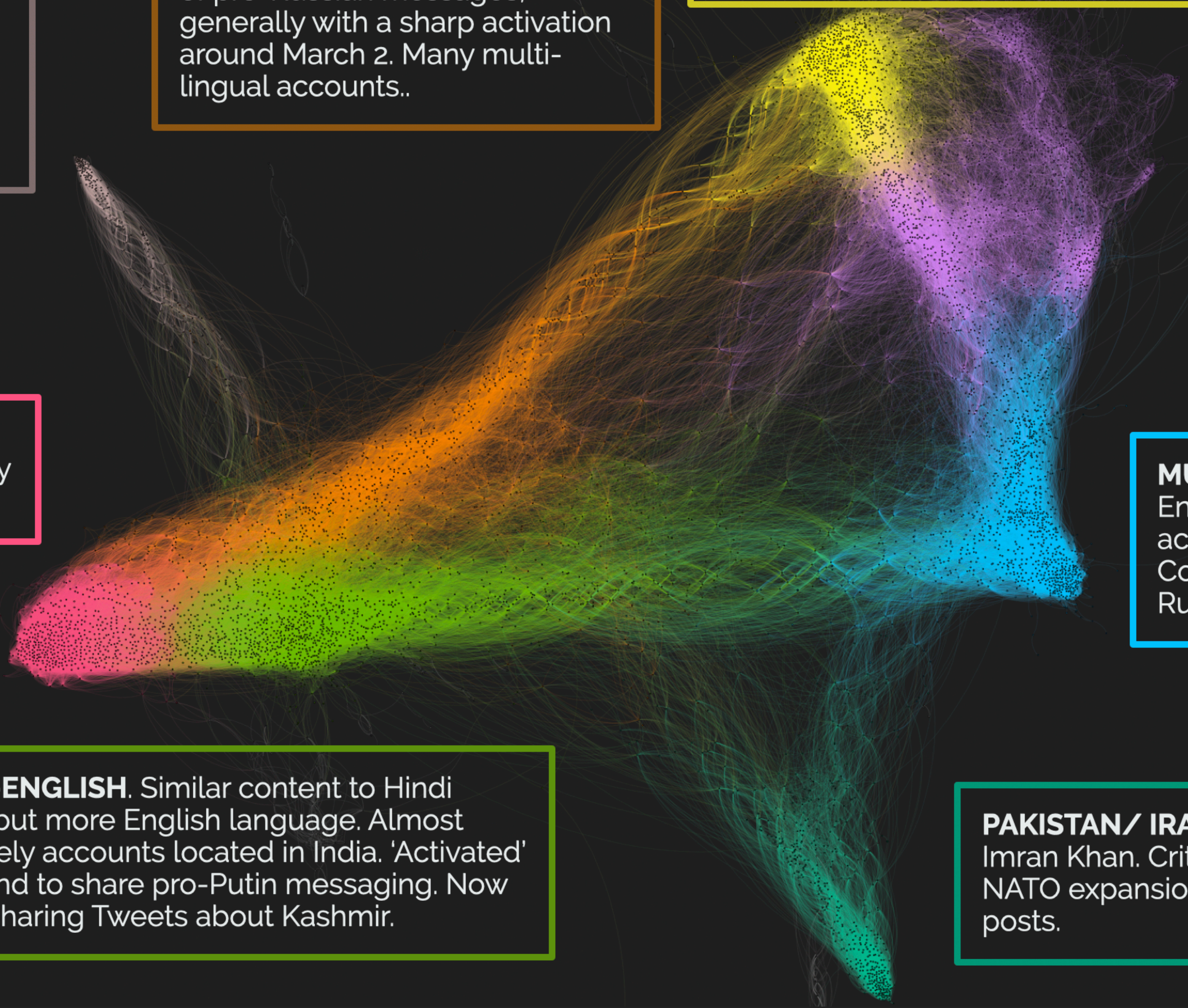
When making the volumetric measurements over time mentioned below, it should be noted that we drew on a different dataset. Rather than using the 200 most recent tweets from each timeline (which would, due to the cap, have introduced measurement artefacts), we used historic data collected using Twitter's Counts endpoint[20] to ensure Tweets for all accounts spanned the same time period. The counts of all Tweets from accounts in each cluster, and counts for all keyword related tweets from accounts in the cluster were collected between Feb 13th and March 14th (inclusive).

It was noted that a very large amount of the Tweets featuring #IstandwithPutin and #istandwithRussia contained English, regardless of the language/s that the accounts otherwise used (for a presentation of these messages, please see the next section). We have taken the usage of a small number of selected invasion-related English-language keywords (titled 'keywords' below)[21] as a simple between-cluster comparison of this English-language Russia-related activity. This is entirely indicative, and should not be taken as a measure of the entirety of invasion-related activity, and certainly not outside of English.

[20] https://developer.twitter.com/en/docs/twitter-api/tweets/counts/api-reference/get-tweets-counts-all

[21] The keywords were: 'Ukraine', 'Russia', 'Putin', '#IstandwithRussia', '#IstandwithPutin', 'Nato', 'Donbass', 'Kiev', 'Kyiv', 'Mariupol', 'Zelensky', 'Zelenskyy'. Case insensitive matching.

CASM technology



TAMIL. Majority of posts are retweets. Low density of Russia-related posts. Anti-BJP/Modi. Seems to 'activate' on March 2nd to suddenly share hashtag.

SOUTH ASIAN. Javanese, Urdu, Malaysian, Nepali Lower density of pro-Russian messages, generally with a sharp activation around March 2. Many multi-lingual accounts..

SOUTH AFRICA. Greater variety of messages (selfies, local mines, 'making edo great again)'. Pro-Zuma/ BRICS solidarity

SOUTH AFRICAN/ NIGERIAN. Zulu/English language. Amplified pro-Russian content in [March], now mainly concentrated on fuel shortages in Nigeria.

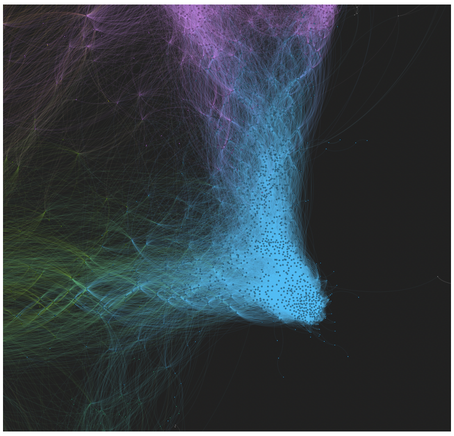
HINDI. pro-BJP/MODI. Anti-colonial, pro-BRICS solidarity messaging

MULTI-LINGUAL SPAM Hindi/ English/Chinese. Newest accounts, very few followers. Concentrated retweeting of pro-Russian memes.

INDIAN-ENGLISH. Similar content to Hindi cluster, but more English language. Almost exclusively accounts located in India. 'Activated' March 2nd to share pro-Putin messaging. Now largely sharing Tweets about Kashmir.

PAKISTAN/ IRAN. Urdu, Sindhi, Farsi. Pro Imran Khan. Criticise Western hypocrisy, NATO expansionism, some anti-Indian posts.

BLUE: ‘MULTI-LINGUAL SPAM NET’



1,128 accounts (16% of total)
Median followers:[22] 7
Retweet:Tweet ratio:[23] 3.89x

The BLUE cluster is a distinct and relatively dense nodal grouping which connects both to the predominantly African side of the network and the cluster of English-using Indian accounts.

This is the only cluster in the network that does not have a dominant natural language as its defining or most important characteristic. Accounts within it were observed to predominantly use English, but also Hindi, English, and Chinese, perhaps explaining its central location. These accounts are possibly grouped together because they see the highest concentration of pro-invasion messaging of any cluster, taking the form of a smaller number of very heavily shared pro-invasion, pro-Russian memes. Unlike other clusters, BLUE accounts do not tend to amplify related to any other major theme or region.

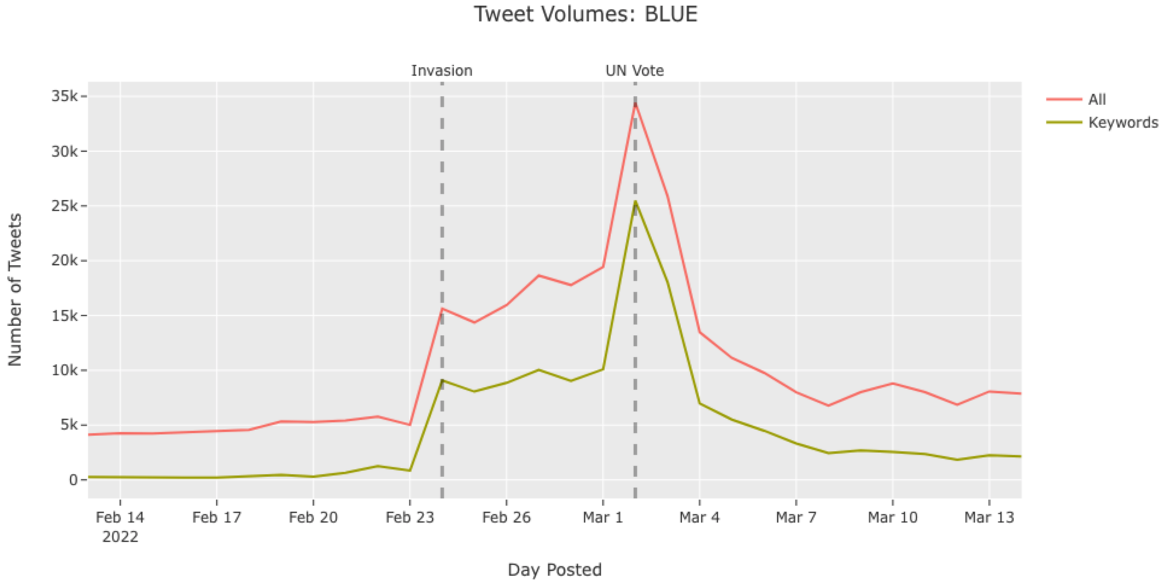
Accounts in this cluster are, on average, both the newest and have the fewest followers of any cluster. Behaviourally, most were observed to engage in high-levels of retweeting activity over the period in question, usually pro-invasion memes. Accounts engaging in counter-speech were also observed in this cluster.[24]

More than for any other cluster, we see a sharp peak in activity on 2nd March, the day of the UN vote demanding an end to Russia’s military operation. This cluster showed the highest proportion (~70%) of Tweets and Retweets on this day which used Russia-related keywords, compared to similar peaks in activity exhibited by other clusters. The cluster’s activity then quickly fell to much lower volumes after 4th March.

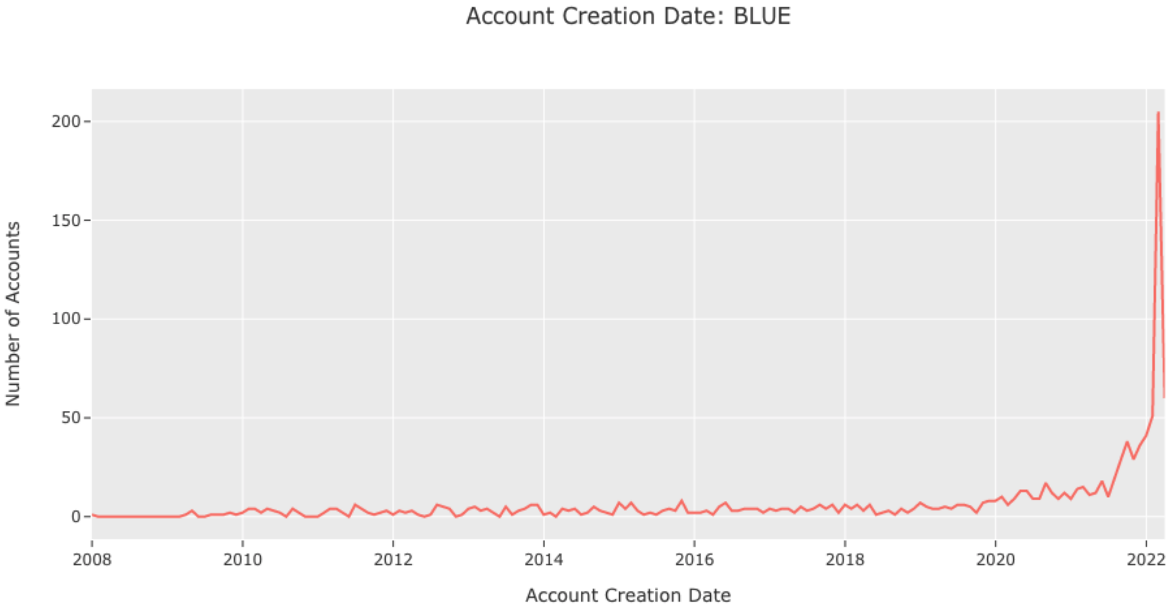
[22] We use median here as account followership can show extreme variation and we wish to avoid skew in this metric by a few extremely high follower accounts as would occur with mean averaging.

[23] The proportion of the accounts’ posts in our collected dataset - i.e. the 200 most recent posts from each account.

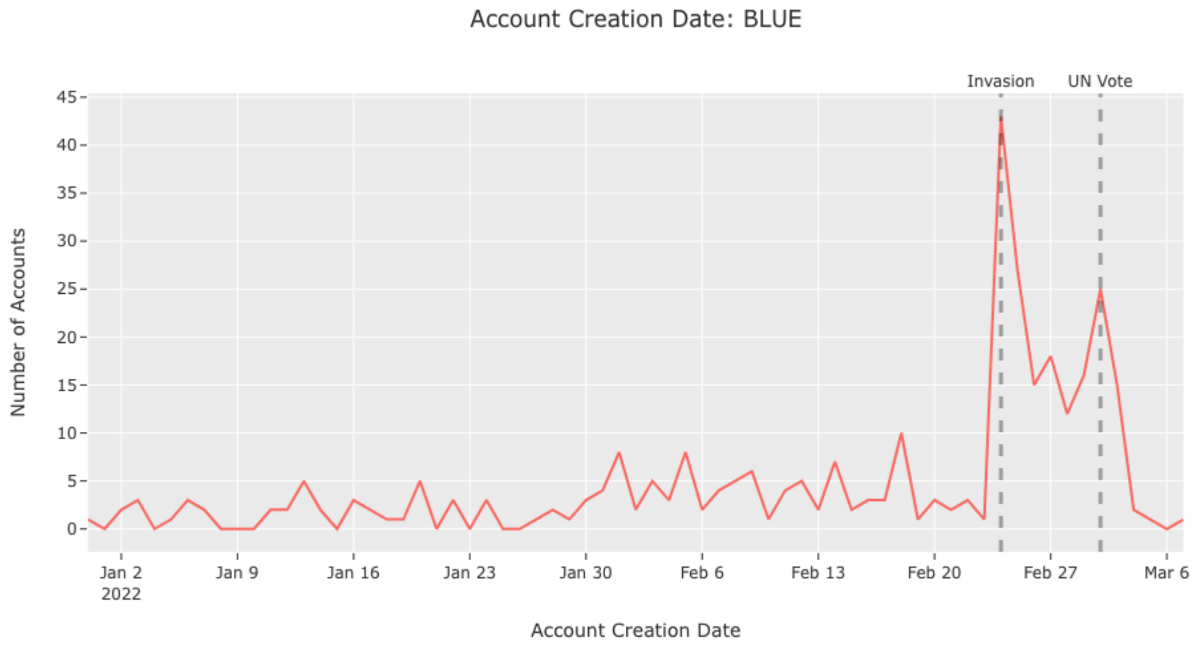
[24] Perhaps given the thresholding we apply of only analysing accounts sending five or more Tweets containing either #istandwithputin and/or #istandwithrussia, we observed less counter-speech or ‘hashtag tainting’ than other research also investigating these hashtags.



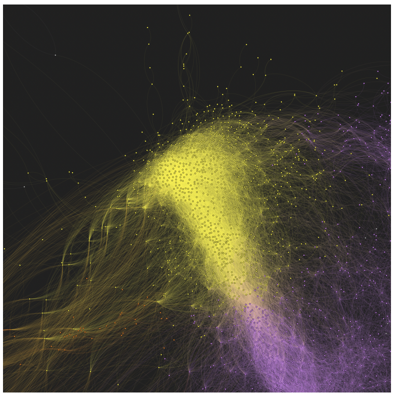
A substantial proportion (28.0%) of these accounts were created in 2022.



When the granularity of the graph is increased, we can see two sharp peaks in account creation, falling on the day of the Russian invasion of Ukraine (February 24th) and the day of the United Nations General Assembly vote to condemn the invasion (March 2nd).



YELLOW: ‘PREDOMINANTLY SOUTH AFRICAN’



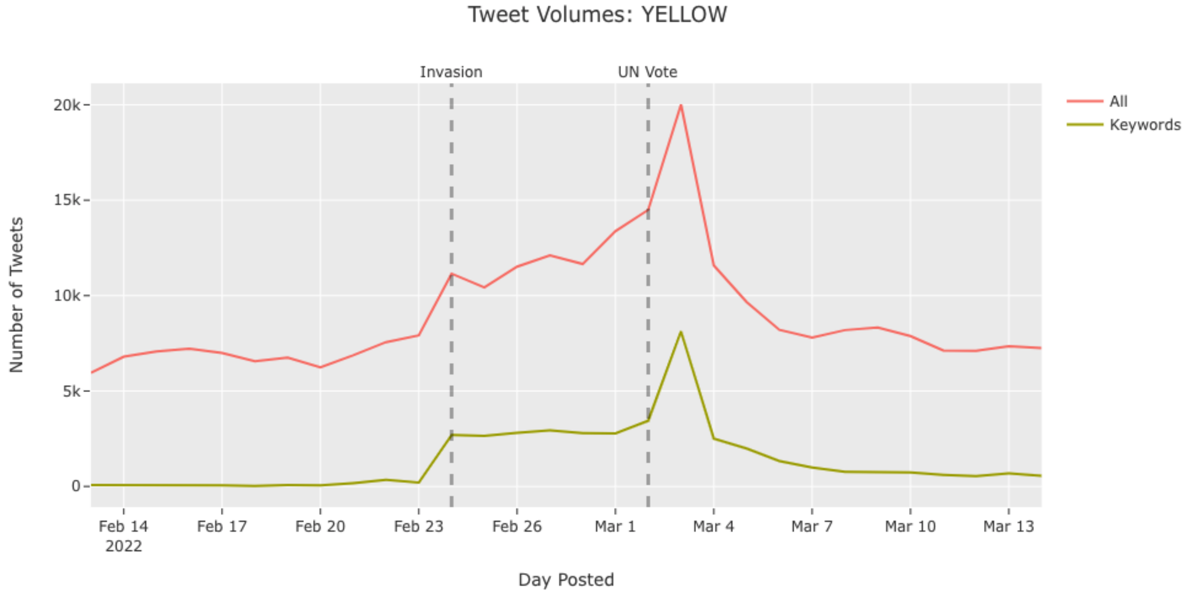
1,010 accounts (14.6%)
Median followers: 241
Retweet:Tweet ratio: 0.50x

The YELLOW cluster forms (along with the RED and BLUE communities) one of the three key vertices of the network map. It is a dense cluster of linguistically similar accounts that predominantly identify as South African, but was also observed to include accounts with Ghanian, Nigerian, and Kenyan identities.

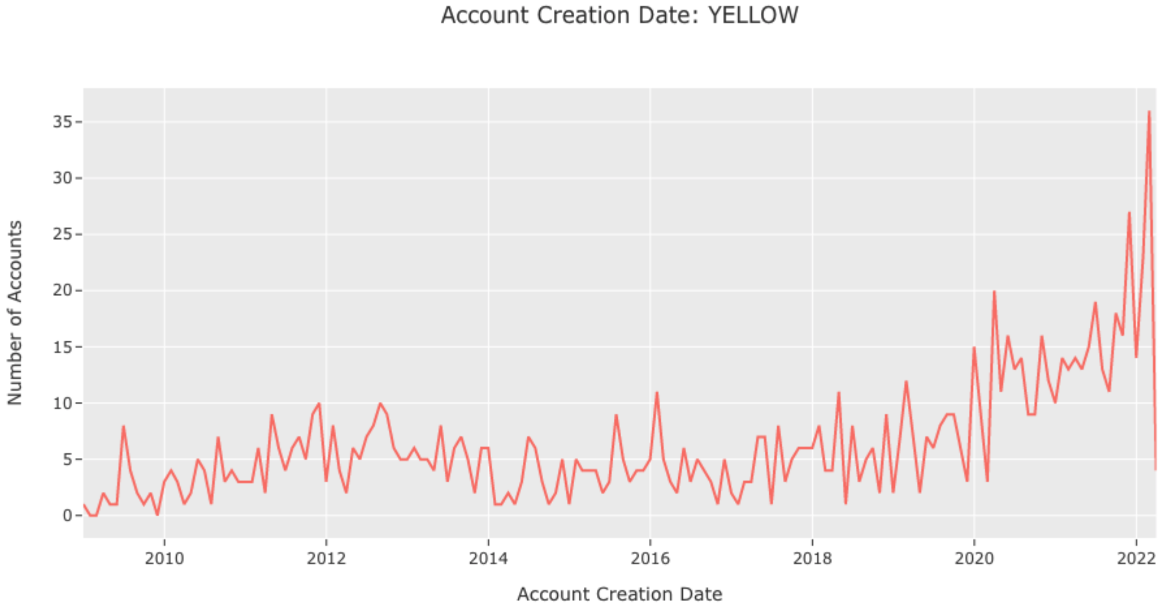
The accounts appraised by analysts overwhelmingly Retweet English-language content on a range of topics, including celebrations of Jacob Zuma, promotion of local political rallies in Africa, and expressions of BRICS and anti-colonial solidarity with Russia.

YELLOW accounts behave in ways very distinct from the Hindi- and English-language accounts identifying as Indian on the other side of the network. They have the highest number of original messages (as opposed to Retweets) of any cluster; the highest (median) number of followers; and were observed to send substantially more content. Taken together, these attributes are suggestive of a greater proportion of genuine human activity rather than spam, and accounts were observed to share consistent content such as personal pieces to camera, selfies, family pictures and so on that would be uneconomic (although still possible) for a spam operation to produce at scale.

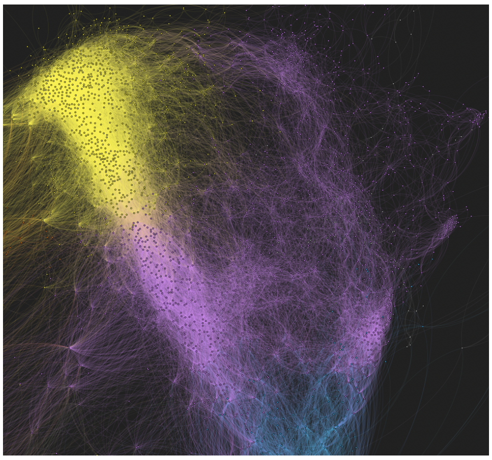
Their peak of activity came March 3rd, the day after the UN vote (and associated peak in Blue cluster activity). Of all clusters their peak activity showed the lowest proportion of Russia-related keywords, which featured in less than 40% of their Tweets on March 3rd.



A peak in account creation occurred in 2022, though this was not as distinct as for the BLUE and (as described later) VIOLET clusters.



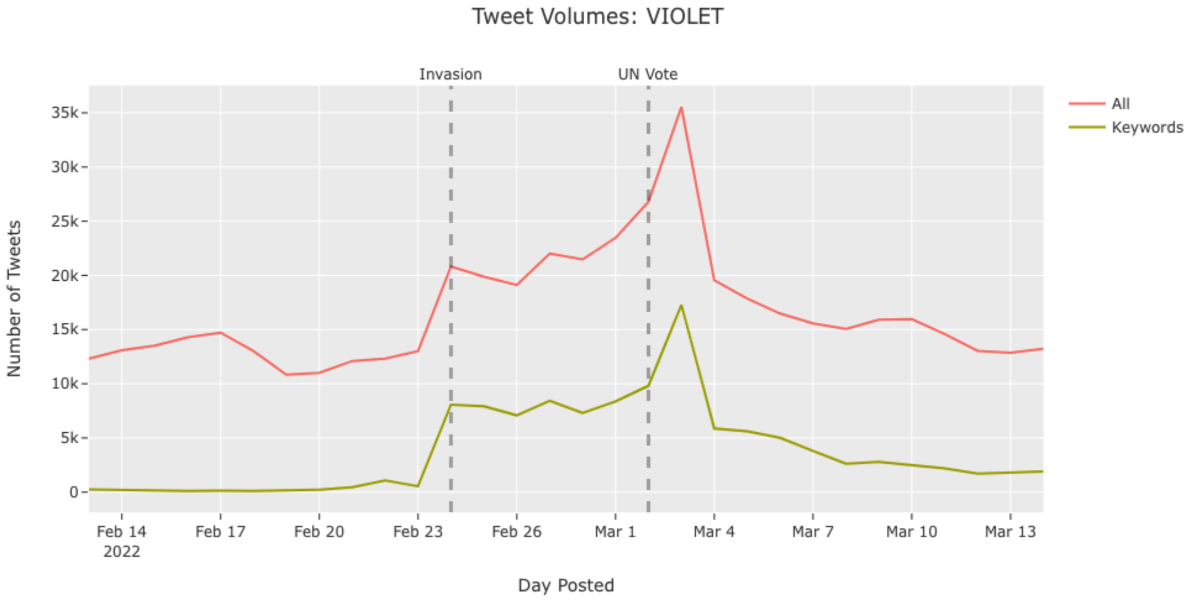
VIOLET: ‘SOUTH AFRICAN/NIGERIAN’



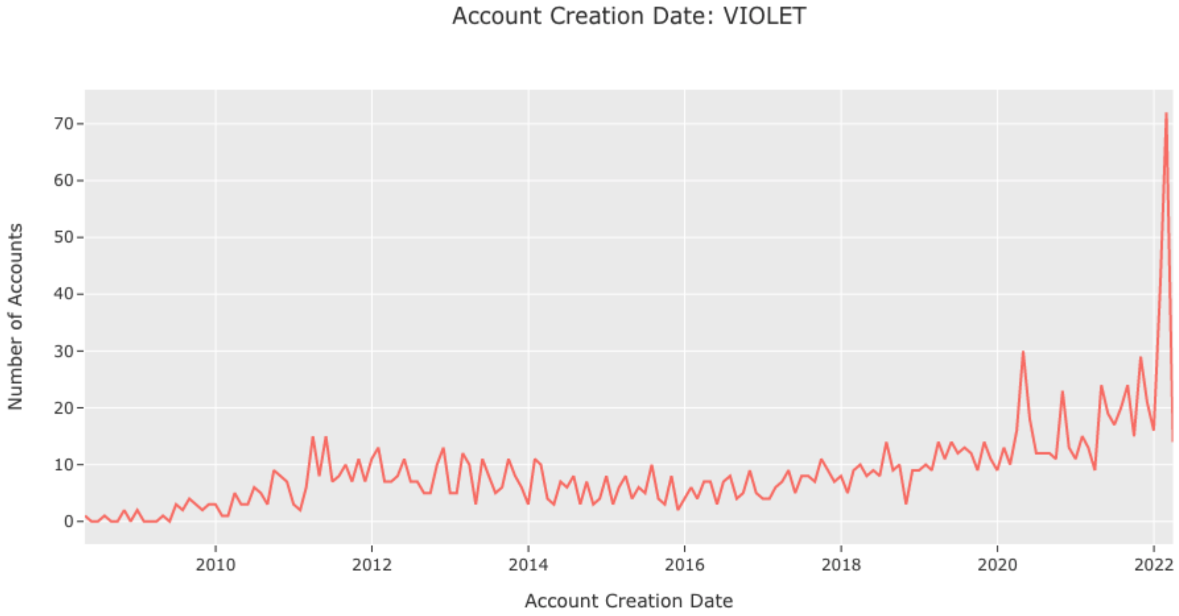
1,441 accounts (20% of total)
Median followers: 220
Retweet:Tweet ratio: 2.11x

The VIOLET cluster is the largest cluster by number of accounts, and one of the most difficult to characterise. It mixes together accounts with a number of different African identities - especially South Africa and Nigeria - and who communicate predominantly in English with Zulu-language messaging also observed. Besides Russia-related content, in our period of analysis they concentrated on a range of issues including football, fuel shortages in Nigeria and public health news.

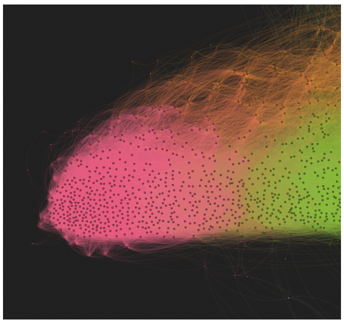
Similarly to YELLOW, there was a very sharp peak in activity and a concurrent spike in Tweets Russian-related vocabulary, which occurred *after* the UN vote. Around ~50% of the tweets in the 3rd March peak used Russia-related keywords (a higher proportion than in YELLOW).



Similarly to BLUE, a substantial proportion of these accounts were created in 2022. This, plus the adjacency of this cluster to the BLUE cluster, could suggest that spam accounts specially created to amplify particular pro-Russia content during the invasion appear across both (though, again, it is difficult to conclusively establish this).



RED: DENSE NODE OF MIXED HINDI- AND ENGLISH-LANGUAGE INDIAN

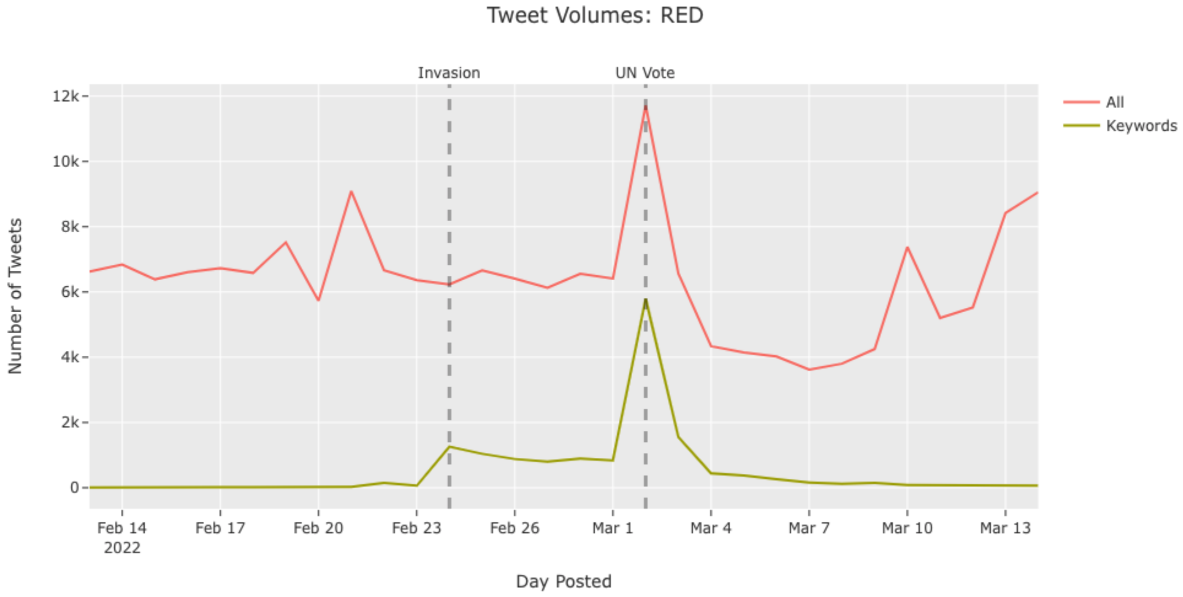


566 accounts (8.2% of total)
Median followers: 79
Retweet:Tweet ratio: 4.54x

With BLUE and YELLOW as the other two dense vertices of the network, the RED cluster forms the third. A dense and linguistically distinct cluster, it is overwhelmingly comprised of accounts identifying as Indian who share Tweets in both English and Hindi, sometimes in the same message. Gujarati was also identified.

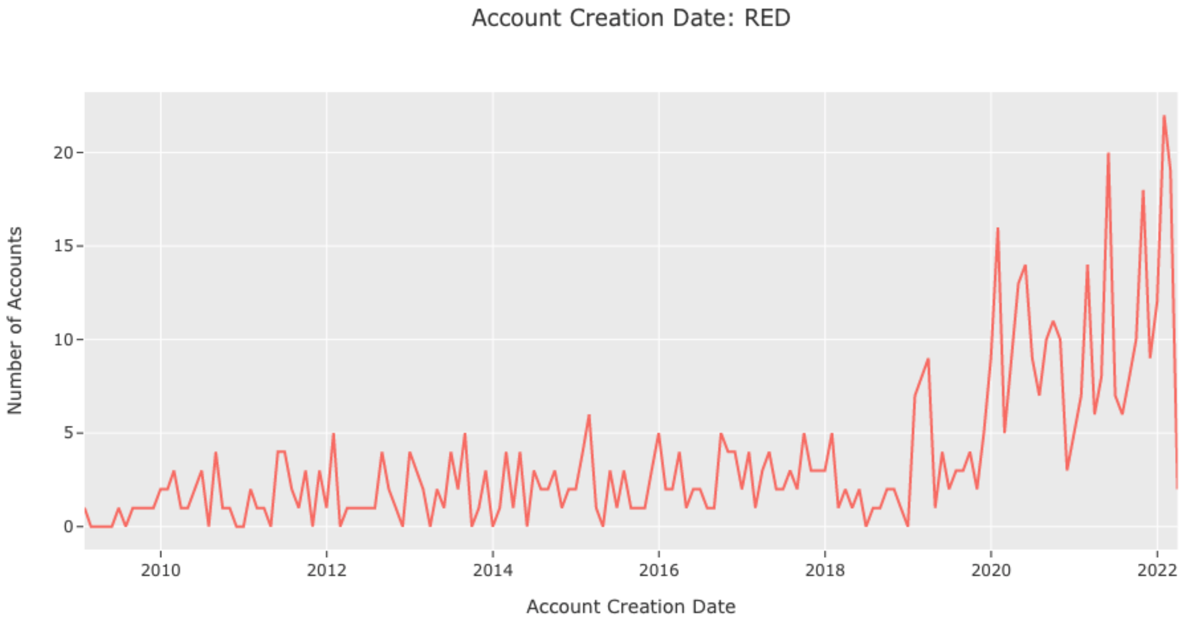
These accounts have very high Retweet-to-Tweet ratios and, and a very large number of messages for each account. The content being shared by these accounts was judged to be extremely political, and overwhelmingly related to the amplification of pro-BJP, pro-Modi messages. Related to Russia, there was a large amount of pro-BRICS solidarity in relation to the United Nations vote (as particularly demonstrated by the most-shared posts, discussed further down).

Consistent with the BLUE clusters, and also the other two Indian clusters, the day of the UN vote saw clear increases in both the overall number of Tweets sent and also the use of English-language Russian-related vocabulary. Whilst overall volumes remained high in the ensuing days, this vocabulary rapidly diminished however.

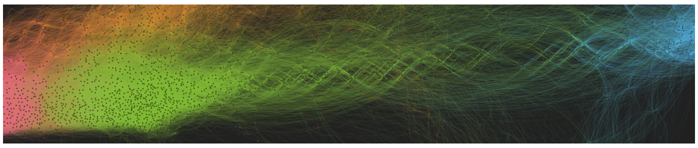


An ongoing monitoring of this cluster beyond March 13th has indicated that the 2022 Hindi-language drama *Kashmir Files* has dominated the messages amplified by this cluster: over March 11th-18th 94% of the accounts in this cluster posted about the film at least once.

Unlike other clusters previously discussed, we do not see a sharp increase in accounts created in 2022. However, the low average follower numbers and high Retweet:Tweet levels do imply spam-related activity. We also observe that Russia-related message decreases sharply after the UN vote, but overall volumes of messaging does not. Our speculation is that many of the RED accounts are members of a 'paid to engage' spam network that can be rented to supply amplification to a number of different clients, and has over our time of study been used to amplify BJP politics, a commercial cinema release and also the invasion of Ukraine.



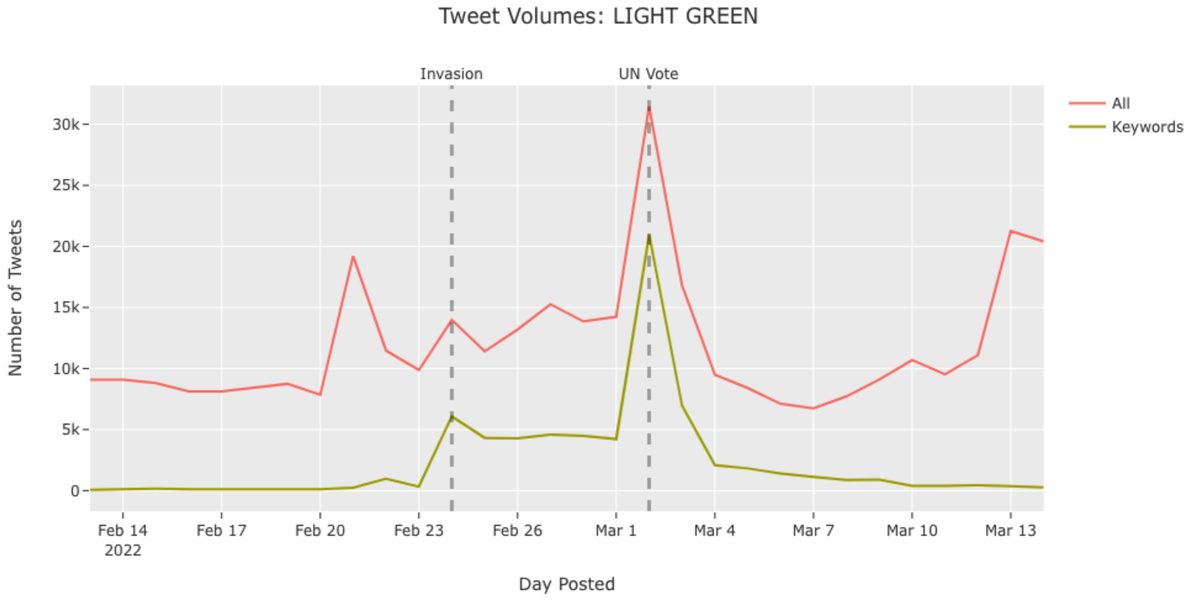
LIGHT GREEN: PREDOMINANTLY ENGLISH LANGUAGE INDIAN



1,314 accounts (19.0% of total)
Median followers: 18
Retweet:Tweet ratio: 4.45x

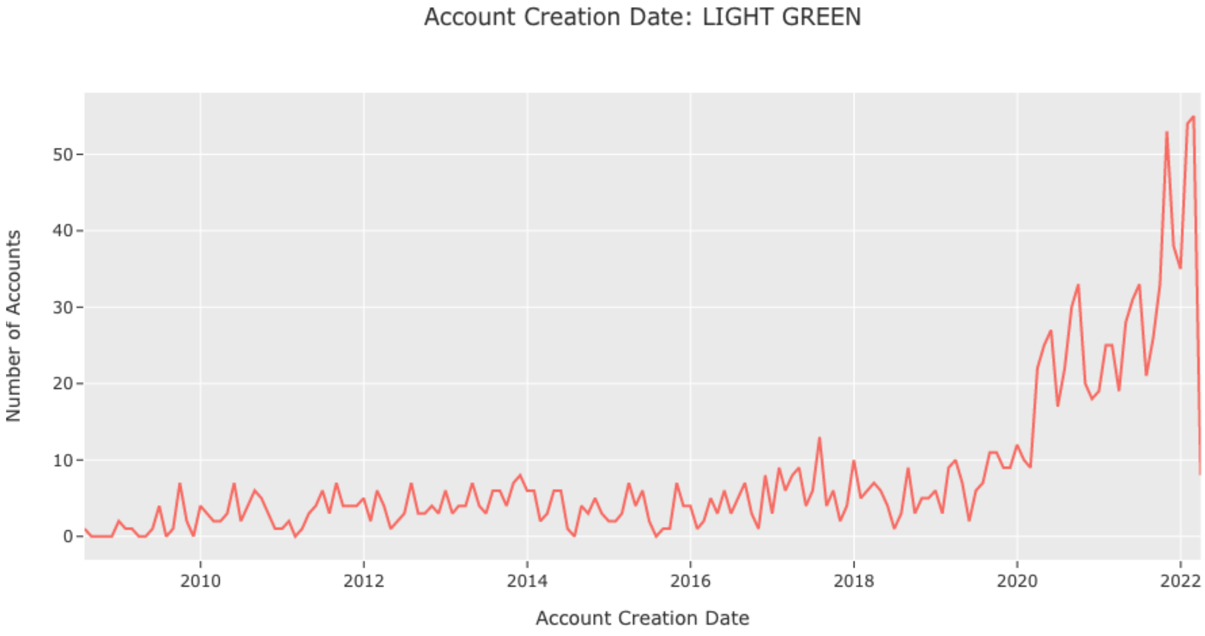
The LIGHT GREEN cluster connects the Hindi-language RED cluster with the English/multilingual BLUE cluster. Closely bunched toward the RED side of the network, this is a cluster also of predominantly Indian-identifying accounts, also amplifying pro-BJP messaging in high volumes through a high number of Retweets. The key distinction between the clusters were the LIGHT GREEN accounts were more likely to share only English-language messages, or in greater concentration than Hindi.

Volumetrically, the behaviour of LIGHT GREEN accounts is very similar to RED. There is a small increase of activity on the day of the invasion, and a much larger one on the day of the UN vote. Messages containing English-language Russia-related messages then quickly decreased whilst the overall volume subsequently saw increases.



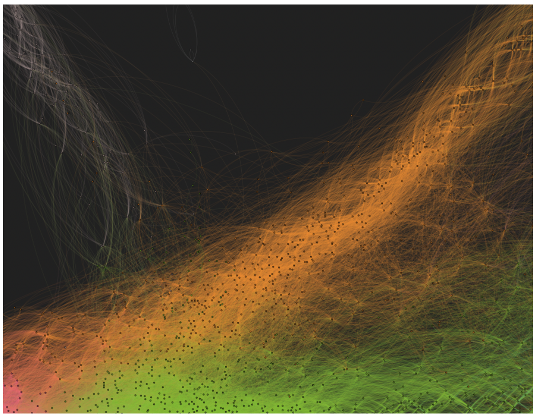
An ongoing monitoring of this cluster beyond March 13th has indicated that, again in line with the RED cluster, the 2022 Hindi-language drama Kashmir Files has been mentioned by 100% of accounts in this cluster between March 11-18, and 45% of all messages posted by accounts in this cluster in the same day contain the keyword “Kashmir”.

Many of these accounts were created relatively recently, though not overwhelmingly in 2022 (as in BLUE).



ORANGE: BROAD SOUTH AND SOUTH EAST ASIAN

736 accounts (10.8% of total in network)

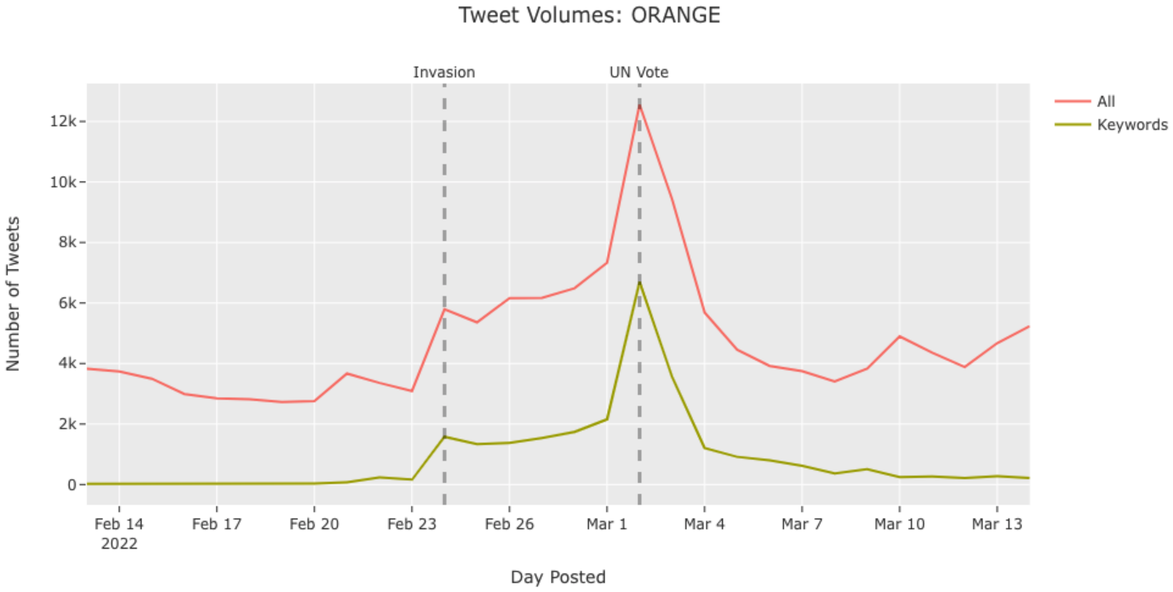


Median followers: 23
Retweet:Tweet ratio: 0.55x
Accounts created in 2022: 10.1%

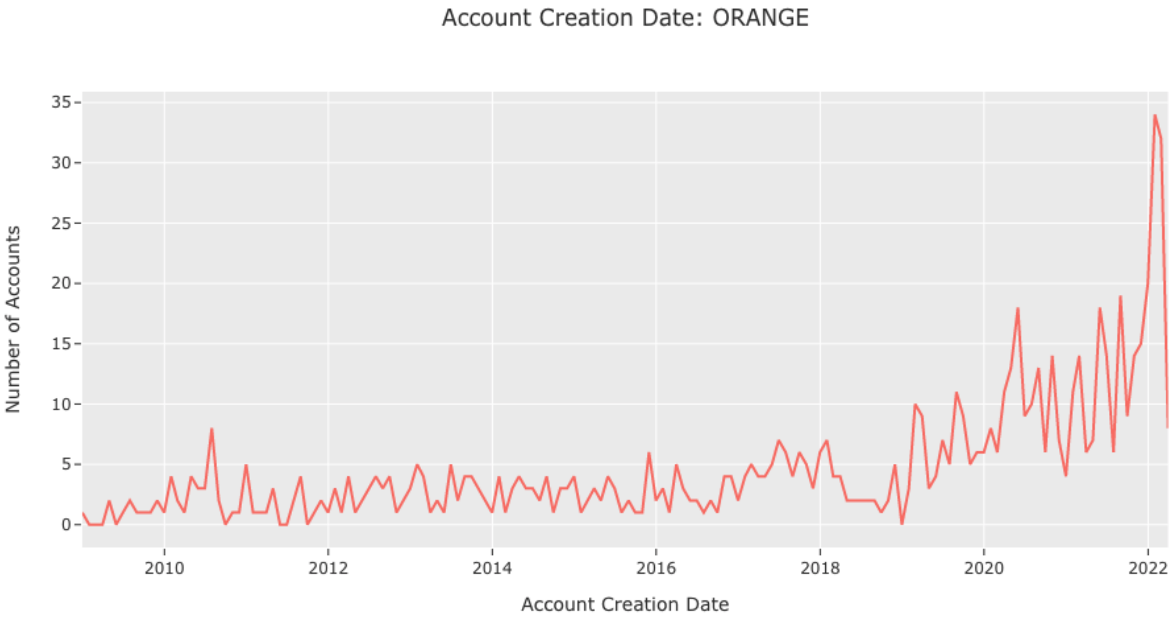
The ORANGE community is a long, tunnel-shaped distribution of accounts that link the dense RED Hindi-language node on one side with the dense YELLOW South African node on the other.

This is one of the most linguistically varied of all the clusters studied for this report. Accounts sharing messages in Hindi, English, Urdu, Malay and Javanese were all identified. What makes this cluster especially challenging to characterise is that accounts often shared messages across multiple different languages, and messages which were themselves combinations of different languages.

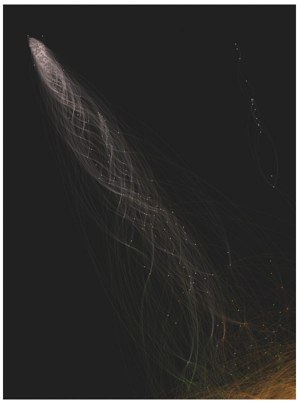
A little under ~50% of their Tweets and Retweets on 2nd March used the Russia-related English keywords.



Amongst the South Asian clusters, this cluster showed the sharpest increase in account creation in 2022 (a phenomenon also identified in the YELLOW South African cluster bordering this cluster).



BEIGE: 'TAMIL-LANGUAGE EXTRUSION’



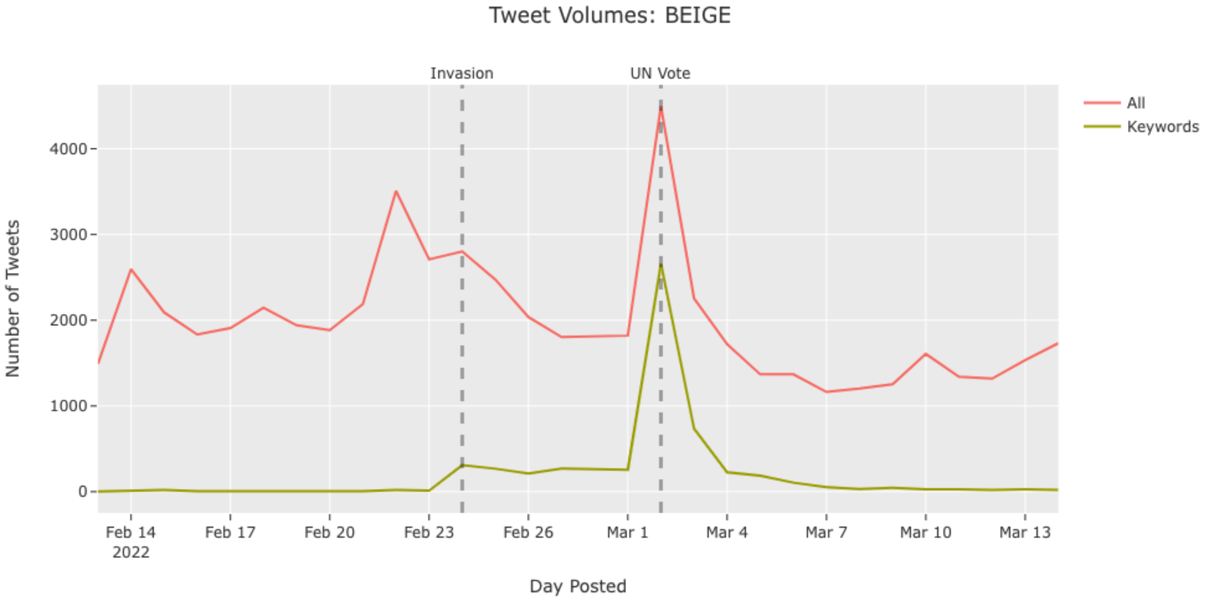
228 accounts (3.3% of total in network)
Median followers: 62
Retweet:Tweet ratio: 4.15x
Accounts created in 2022: 2.6%

BEIGE, the smallest cluster, sits far away from the other others, a distant but linguistically uniform group of (based on the sampled accounts) near-exclusively Tamil-language accounts. Almost all of the accounts appraised explicitly expressed (in Tamil) Tamil nationalism in their profile description. Many of the accounts also explicitly stated Tamil Nadu State in India as their location.

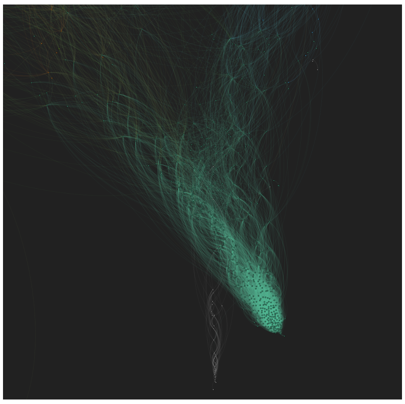
The accounts variously engaged in the amplification of motivational quotations, Indian Premier League Cricket, and popular entertainment including (and consistent with the RED and ORANGE clusters) the *Kashmir Files*. Tamil Nadu-state politics was observed as a common theme, and anti-BJP, anti-Modi messaging was also observed.

Highly shared pro-invasion memes were amongst the minority of English-language messages that this cluster was observed to produce. Unlike the other messages shared by this group, the invasion-related messages seemed to be consistent with those seen shared by the other clusters: Putin 'strong man motifs' and anti-Western/anti-NATO sentiment.

This cluster showed a very high proportion of Retweets, mostly amplification of Tamil-state politics. Around ~50% of their Tweets and Retweets on 2nd March used Russia-related keywords. They only showed a low proportion of recently-created accounts.

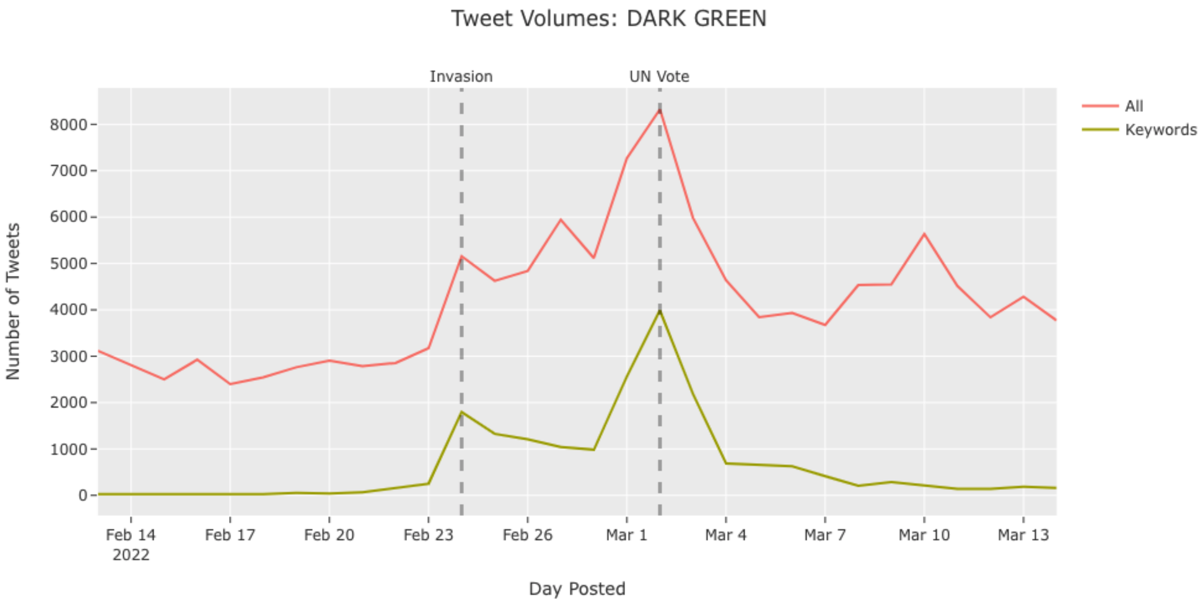


DARK GREEN: 'PRIMARILY PAKISTAN, SOME IRANIAN'



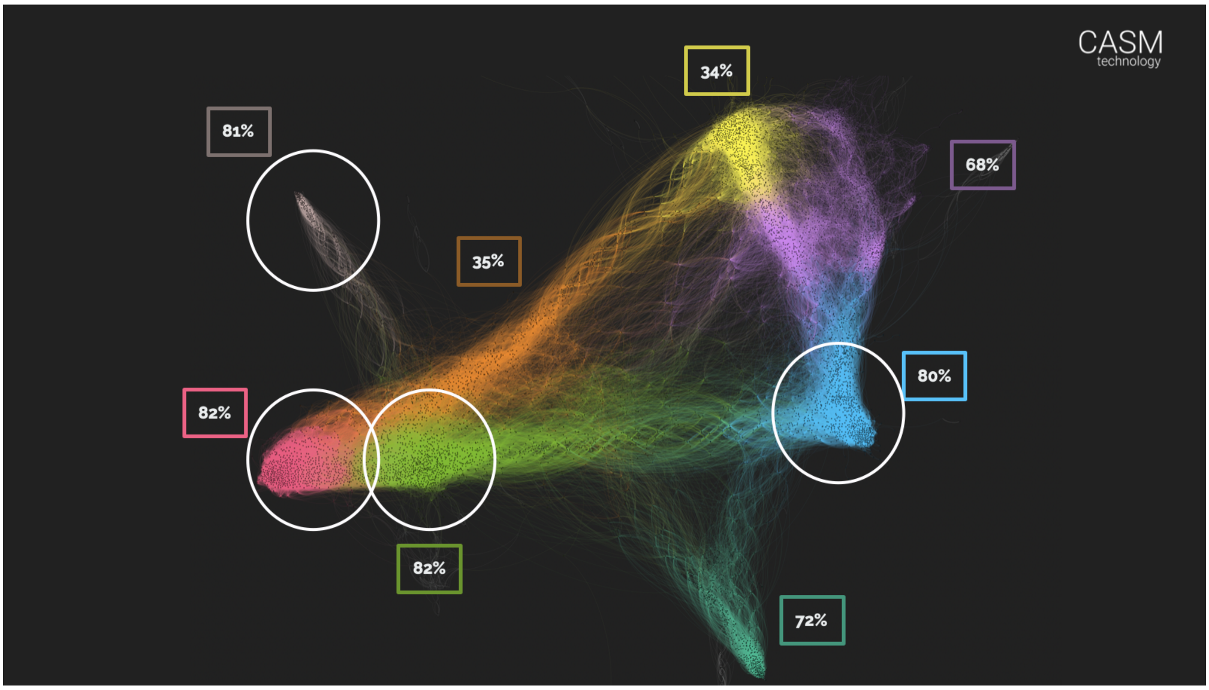
474 accounts (6.87% of total in network)
Median followers: 75.5
Retweet:Tweet ratio: 2.57x

The DARK GREEN cluster is, along with BEIGE, the other clear linguistic offshoot cluster. It is linguistically complex and alongside English many accounts were observed to share messages in Urdu, Sindhi and Farsi. These messages were often supportive of Imran Khan, the Prime Minister of Pakistan, and were frequently critical of Western hypocrisy and NATO expansionism. Anti-Indian sentiment was also observed. There was relatively low use of Russia-related keywords, which featured in ~40% of their Tweets and Retweets on March 2nd.



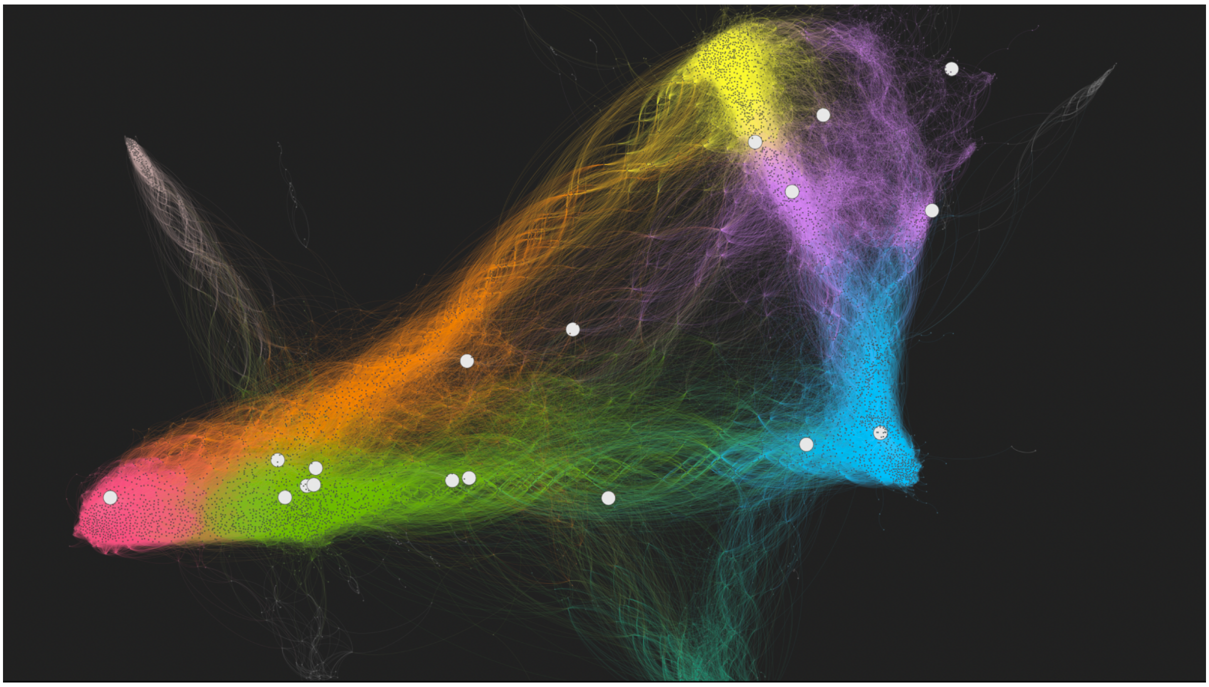
HEAVILY AMPLIFIED CONTENT

An activity in common across many of the accounts we appraised by amplification through very high quantities of Retweets and few original messages. Especially over the days of March 2nd and 3rd, this amplification activity often focussed on a smaller number of pro-invasion and anti-Western memes also using either one or both of the hashtags researched in this report. This did not happen in concert across all clusters, however. The BEIGE, RED, LIGHT GREEN and BLUE clusters had a higher proportion of their (most recent 200 tweet) timeline as Retweets than the ORANGE and YELLOW clusters, as shown below.



Percentage of Retweets in each account's 200 most recent messages (when collected), averaged by cluster

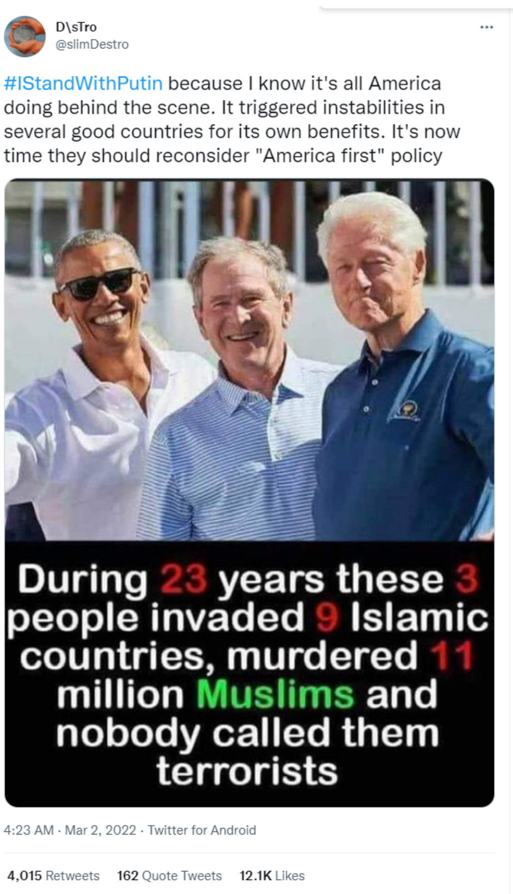
18 of 50 of the most amplified accounts sent by the network were also present within the network (i.e. they have sent Tweets containing the hashtags five or more times). Their locations (marked as larger nodes, below) were distributed across the major clusters, with a notable concentration in the GREEN English-language Indian cluster.



Highly shared memes and messages were appraised by analysts, who drew out a number of themes:[25]

Criticism of Western countries

- Particularly, but not exclusively, focused on the US.
- Often claims hypocrisy, or "whataboutism".
- Often referring to invasions of Middle Eastern countries (or "Muslim countries" in language used by many posts"). Palestine and Libya are frequently mentioned in particular.
- 6 of the 10 top-shared posts in our collected dataset are on this theme.



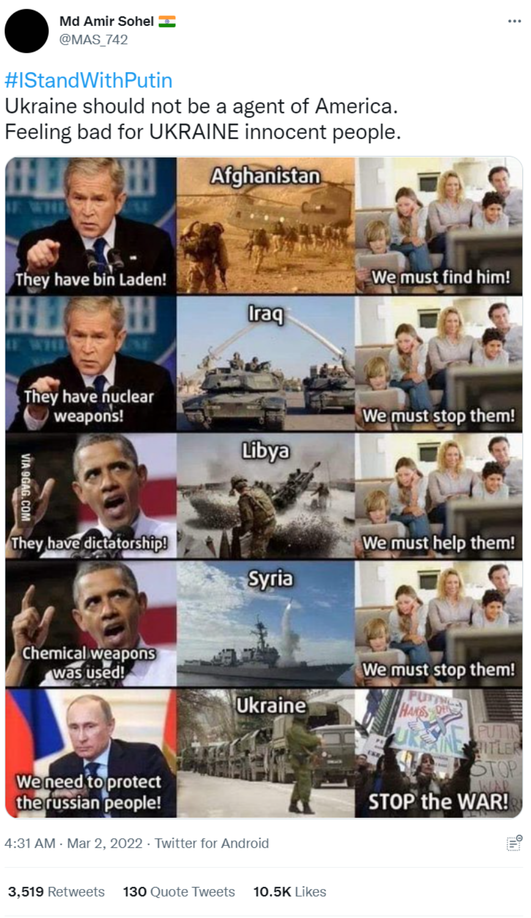
D\sTro
@slimDestro

#IStandWithPutin because I know it's all America doing behind the scene. It triggered instabilities in several good countries for its own benefits. It's now time they should reconsider "America first" policy

During 23 years these 3 people invaded 9 Islamic countries, murdered 11 million Muslims and nobody called them terrorists

4:23 AM · Mar 2, 2022 · Twitter for Android

4,015 Retweets 162 Quote Tweets 12.1K Likes



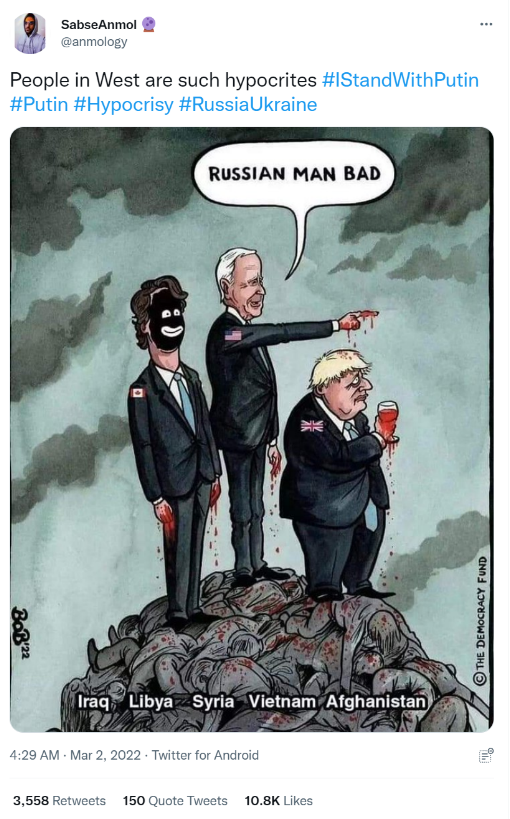
Md Amir Sohel 🇮🇳
@MAS_742

#IStandWithPutin
Ukraine should not be a agent of America.
Feeling bad for UKRAINE innocent people.

4:31 AM · Mar 2, 2022 · Twitter for Android

3,519 Retweets 130 Quote Tweets 10.5K Likes

[25] Please note a number of videos (including originally from TikTok) were also identified. They are not presented here, but broadly fall into the categories presented in this report.

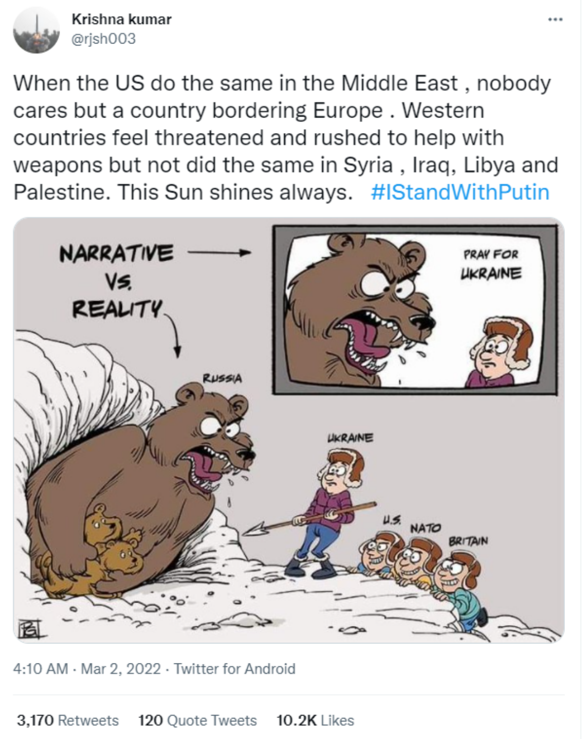


SabseAnmol
@anmology

People in West are such hypocrites #IStandWithPutin #Putin #Hypocrisy #RussiaUkraine

4:29 AM · Mar 2, 2022 · Twitter for Android

3,558 Retweets 150 Quote Tweets 10.8K Likes

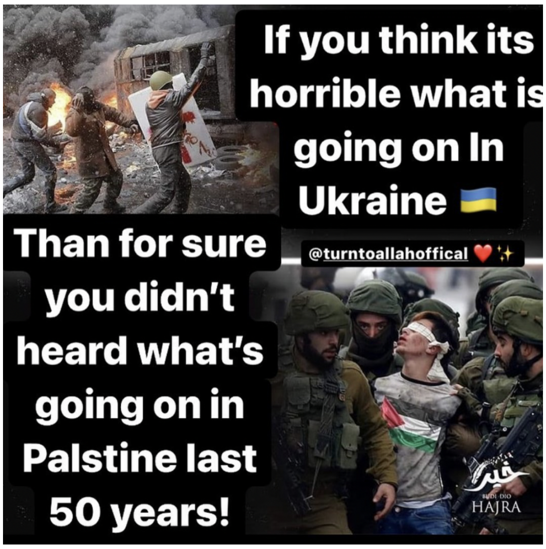


Krishna kumar
@rjsh003

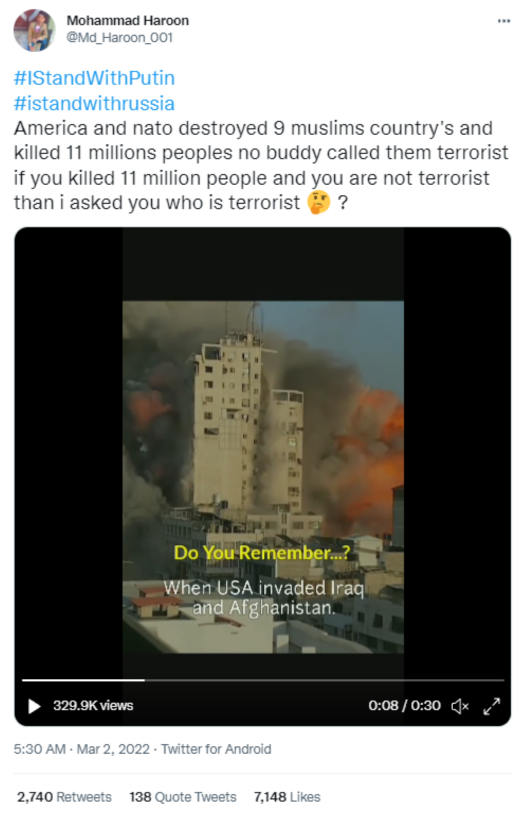
When the US do the same in the Middle East , nobody cares but a country bordering Europe . Western countries feel threatened and rushed to help with weapons but not did the same in Syria , Iraq, Libya and Palestine. This Sun shines always. #IStandWithPutin

4:10 AM · Mar 2, 2022 · Twitter for Android

3,170 Retweets 120 Quote Tweets 10.2K Likes



If you think its horrible what is going on In Ukraine 🇺🇦
Than for sure you didn't heard what's going on in Palstine last 50 years!
@turntoallahoffical ❤️✨
HAJRA

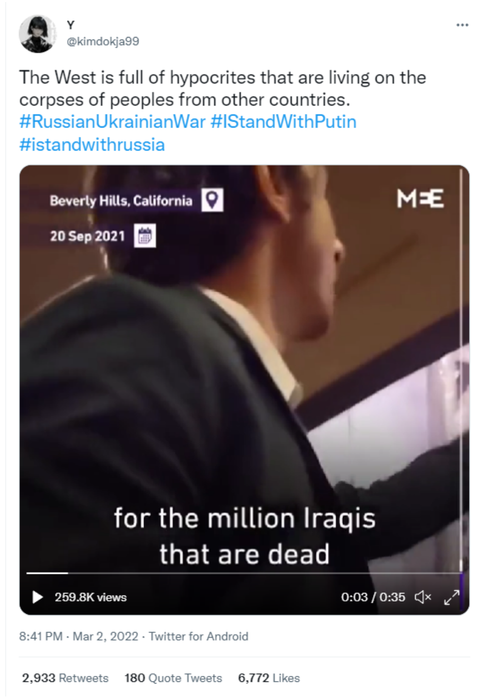


Mohammad Haroon
@Md_Haroon_001

#IStandWithPutin
#istandwithrussia
America and nato destroyed 9 muslims country's and killed 11 millions peoples no buddy called them terrorist if you killed 11 million people and you are not terrorist than i asked you who is terrorist 🤔 ?

5:30 AM · Mar 2, 2022 · Twitter for Android

2,740 Retweets 138 Quote Tweets 7,148 Likes

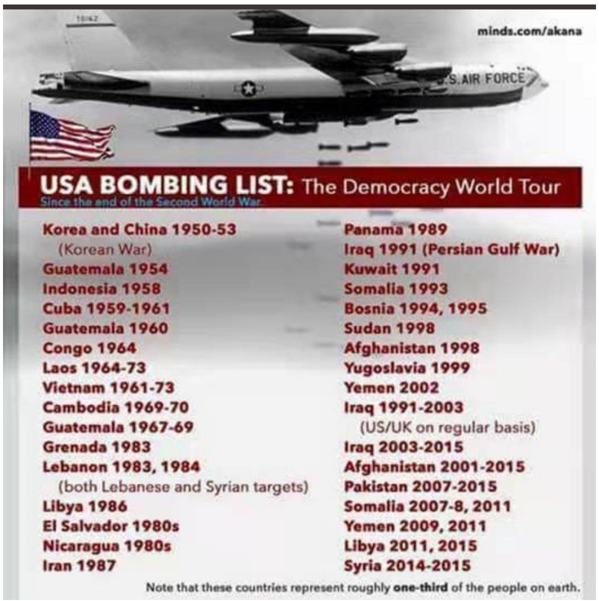


Y
@kimdokja99

The West is full of hypocrites that are living on the corpses of peoples from other countries.
#RussianUkrainianWar #IStandWithPutin
#istandwithrussia

8:41 PM · Mar 2, 2022 · Twitter for Android

2,933 Retweets 180 Quote Tweets 6,772 Likes



minds.com/akana
U.S. AIR FORCE

USA BOMBING LIST: The Democracy World Tour
Since the end of the Second World War

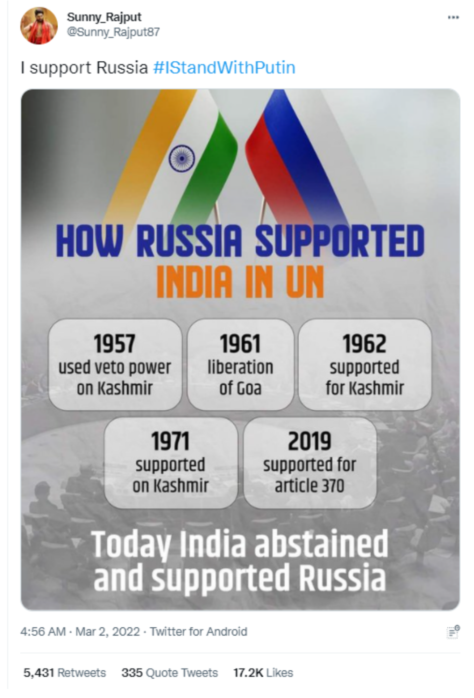
Korea and China 1950-53
(Korean War)
Guatemala 1954
Indonesia 1958
Cuba 1959-1961
Guatemala 1960
Congo 1964
Laos 1964-73
Vietnam 1961-73
Cambodia 1969-70
Guatemala 1967-69
Grenada 1983
Lebanon 1983, 1984
(both Lebanese and Syrian targets)
Libya 1986
El Salvador 1980s
Nicaragua 1980s
Iran 1987

Panama 1989
Iraq 1991 (Persian Gulf War)
Kuwait 1991
Somalia 1993
Bosnia 1994, 1995
Sudan 1998
Afghanistan 1998
Yugoslavia 1999
Yemen 2002
Iraq 1991-2003
(US/UK on regular basis)
Iraq 2003-2015
Afghanistan 2001-2015
Pakistan 2007-2015
Somalia 2007-8, 2011
Yemen 2009, 2011
Libya 2011, 2015
Syria 2014-2015

Note that these countries represent roughly **one-third** of the people on earth.

Explicit References to the United Nations

- As noted in our per-cluster analysis, there were substantial peaks in posting in line with the UN vote to compel Russia to withdraw troops. This was particularly the case for the South Asian cluster, as well as the BLUE spamnet. The theme of the UN vote also appeared in top content.
- #1 most-shared post in our data flagged Russia's support for India in past UN votes.



Sunny_Rajput
@Sunny_Rajput87

I support Russia #IStandWithPutin

HOW RUSSIA SUPPORTED
INDIA IN UN

1957
used veto power
on Kashmir

1961
liberation
of Goa

1962
supported
for Kashmir

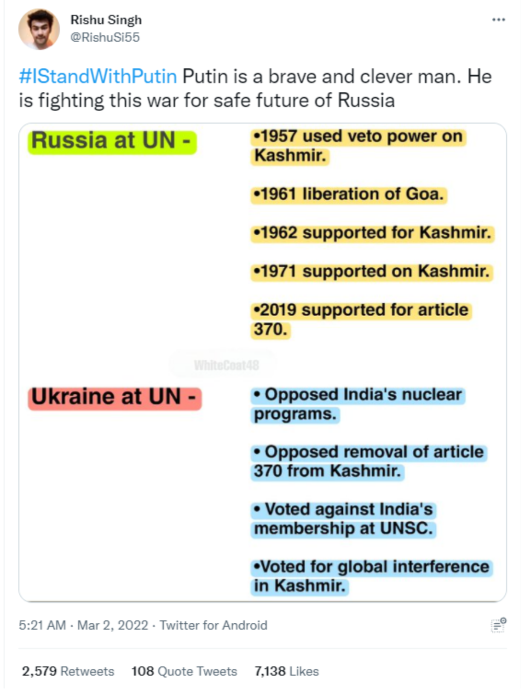
1971
supported
on Kashmir

2019
supported for
article 370

Today India abstained
and supported Russia

4:56 AM · Mar 2, 2022 · Twitter for Android

5,431 Retweets 335 Quote Tweets 17.2K Likes



Rishu Singh
@RishuSi55

#IStandWithPutin Putin is a brave and clever man. He is fighting this war for safe future of Russia

Russia at UN -

•1957 used veto power on Kashmir.

•1961 liberation of Goa.

•1962 supported for Kashmir.

•1971 supported on Kashmir.

•2019 supported for article 370.

Ukraine at UN -

• Opposed India's nuclear programs.

• Opposed removal of article 370 from Kashmir.

• Voted against India's membership at UNSC.

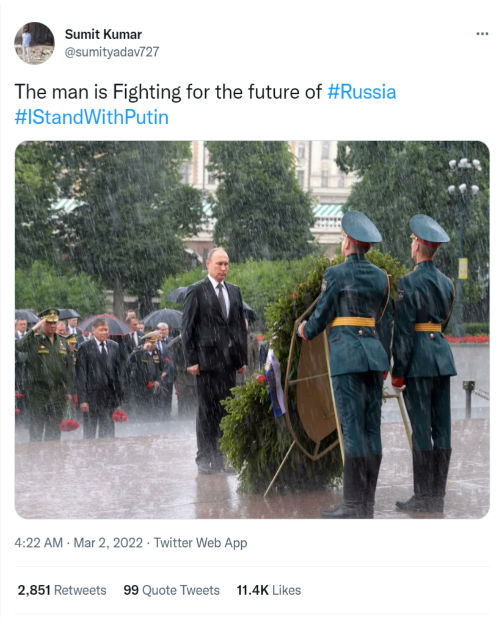
•Voted for global interference in Kashmir.

5:21 AM · Mar 2, 2022 · Twitter for Android

2,579 Retweets 108 Quote Tweets 7,138 Likes

‘Strongman’ Imagery of Putin

- This sometimes involved (favourable) comparisons with other leaders, including Zelenekyy - occasionally this included accusations that Zelenskyy was not actually in Ukraine.

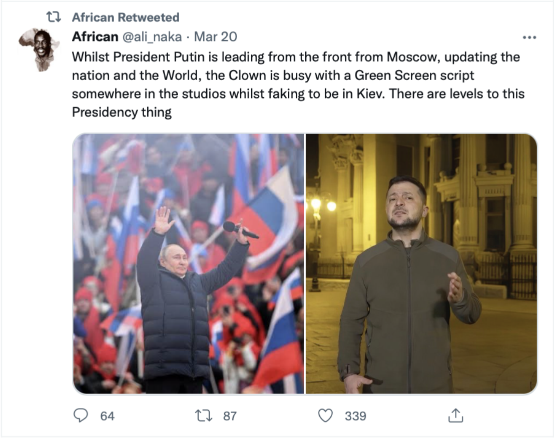


Sumit Kumar
@sumityadav727

The man is Fighting for the future of #Russia #IStandWithPutin

4:22 AM · Mar 2, 2022 · Twitter Web App

2,851 Retweets 99 Quote Tweets 11.4K Likes

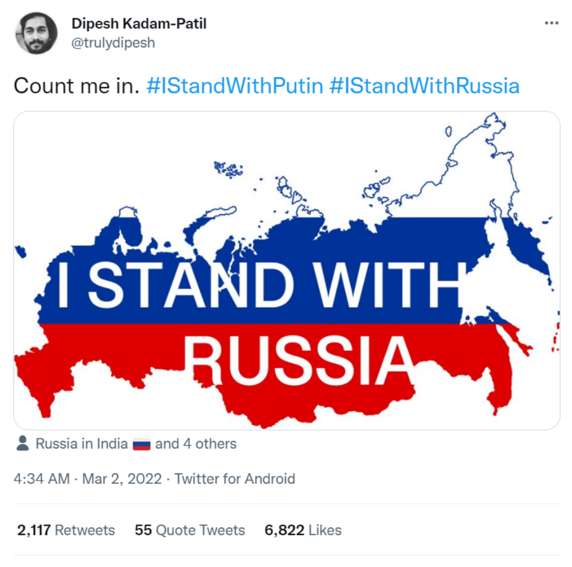


African Retweeted
African @ali_naka · Mar 20
Whilst President Putin is leading from the front from Moscow, updating the nation and the World, the Clown is busy with a Green Screen script somewhere in the studios whilst faking to be in Kiev. There are levels to this Presidency thing

64 87 339

General Memetic Portrayal of pro-Russian support

- Imagery often with echoes of Soviet Realism, or used a similar flag overlay used by a lot of pro-Ukraine imagery.



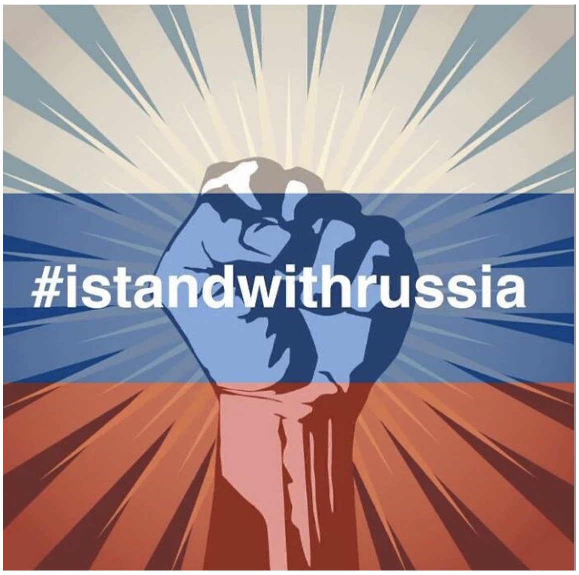
Dipesh Kadam-Patil
@trulydipesh

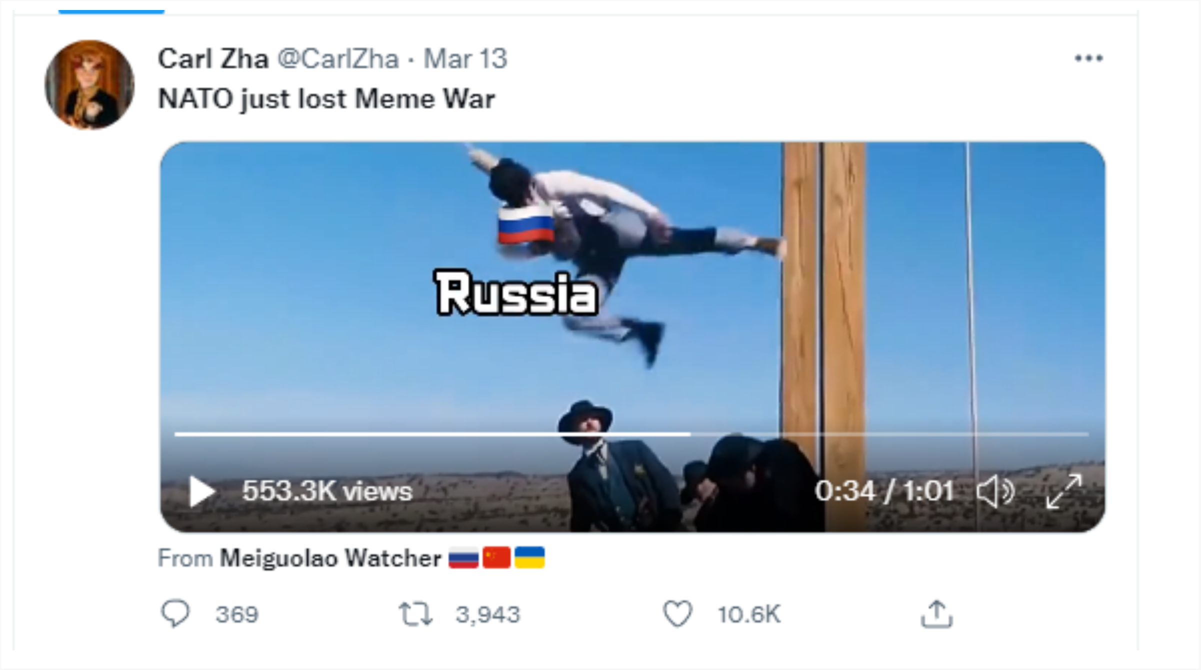
Count me in. #IStandWithPutin #IStandWithRussia

Russia in India 🇷🇺 and 4 others

4:34 AM · Mar 2, 2022 · Twitter for Android

2,117 Retweets 55 Quote Tweets 6,822 Likes





Carl Zha @CarlZha · Mar 13

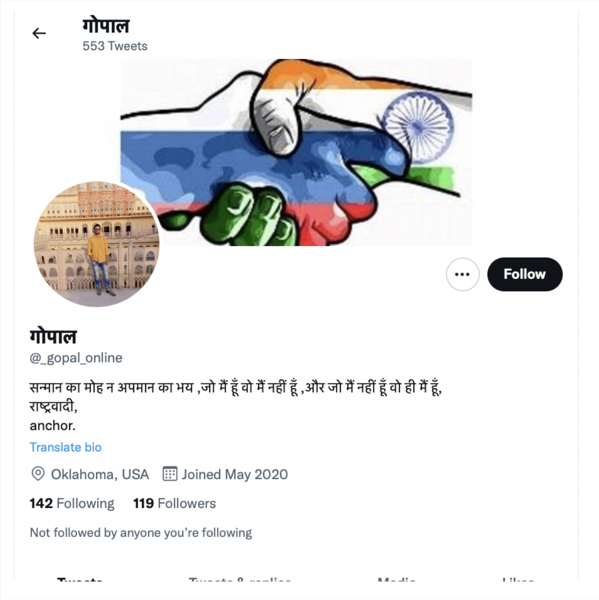
NATO just lost Meme War

From Meiguolao Watcher 🇷🇺🇨🇳🇺🇦

369 3,943 10.6K

BRICS Solidarity

- This theme appeared across both Indian and African accounts, with explicit references to pro-Indian and pro-African solidarity.
- Appeared in some profile pictures as well as content.



गोपाल

553 Tweets

गोपाल

@_gopal_online

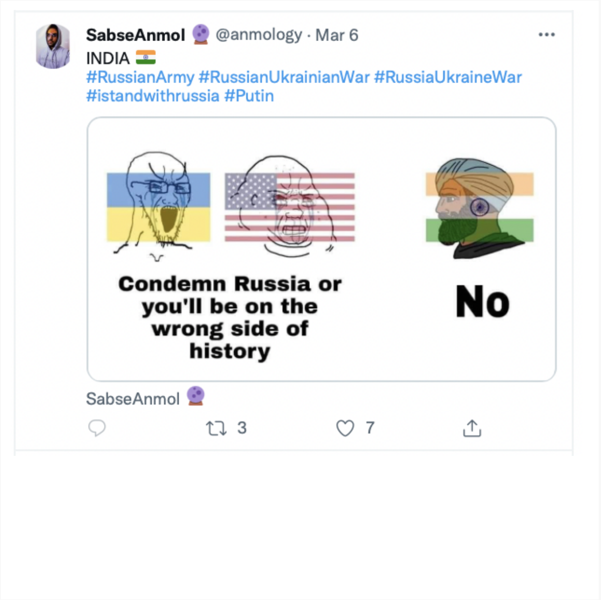
सन्मान का मोह न अपमान का भय ,जो मैं हूँ वो मैं नहीं हूँ ,और जो मैं नहीं हूँ वो ही मैं हूँ,
राष्ट्रवादी,
anchor.

Translate bio

Oklahoma, USA Joined May 2020

142 Following 119 Followers

Not followed by anyone you're following

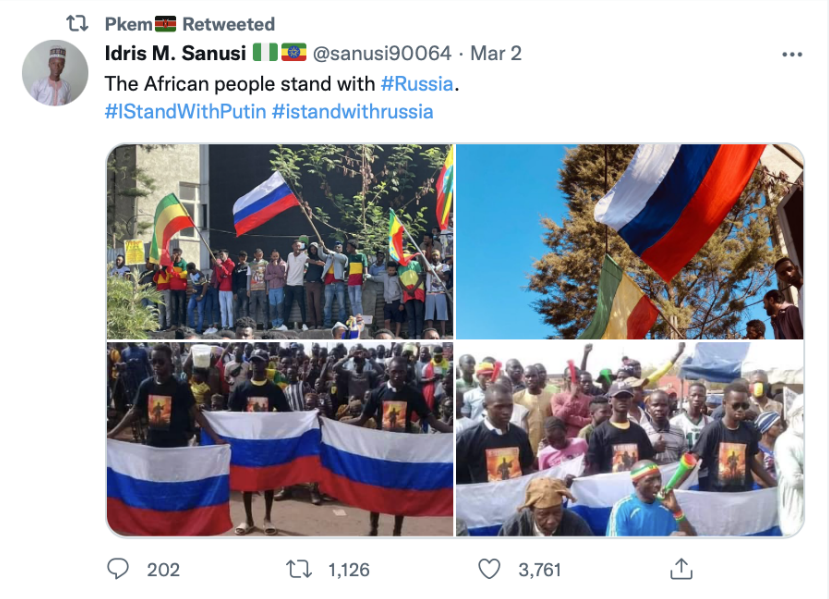


SabseAnmol @anmology · Mar 6

INDIA 🇮🇳
#RussianArmy #RussianUkrainianWar #RussiaUkraineWar #istandwithrussia #Putin

SabseAnmol

3 7



Pkem🇰🇪 Retweeted

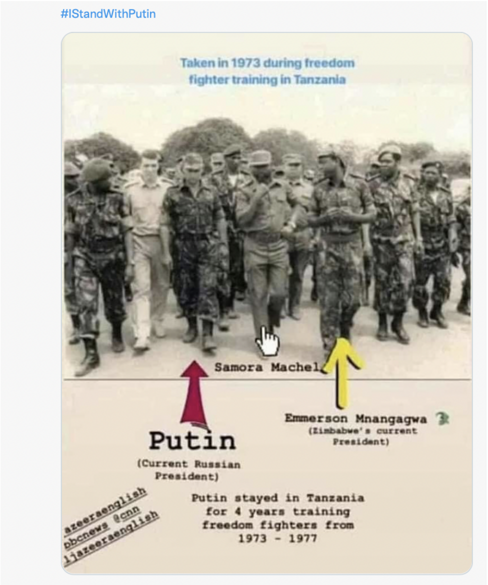
Idris M. Sanusi 🇳🇬🇪🇹 @sanusi90064 · Mar 2

The African people stand with #Russia.
#IStandWithPutin #istandwithrussia

202 1,126 3,761

Anti-colonialism

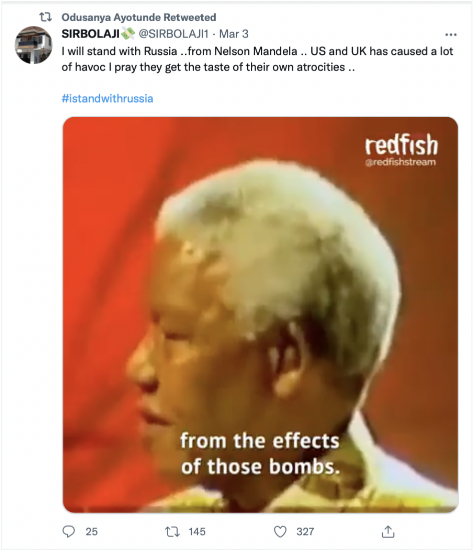
- Some anti-western / pro-Russia content was refracted through historical lenses of colonialism.



#IStandWithPutin

Taken in 1973 during freedom fighter training in Tanzania

Putin stayed in Tanzania for 4 years training freedom fighters from 1973 - 1977



Odusanya Ayotunde Retweeted

SIRBOLAJI @SIRBOLAJI1 · Mar 3

I will stand with Russia ..from Nelson Mandela .. US and UK has caused a lot of havoc I pray they get the taste of their own atrocities ..

#istandwithrussia

INTERPRETATION

The research of suspicious activity online, including suspected influence campaigns, frequently confronts a common problem: any number of different underlying motivations can manifest as the same behaviour. Inauthentic activity, organic engagement, commercial motivations activity, deep and genuine commitment to a cause all mix together to drive online behaviour. Influence campaigns are similarly heterogeneous, and can involve false identities amplifying truths, real people sharing falsehoods, the use of harassment to deny voices from the debate, attributed and non-attributed actors and both those that are centrally controlled with those that are completely self-directing.

Definitively distinguishing between all of these different forms of activity is extremely challenging. The nature of the activity often relies on the underlying intent and motivation of the actors involved and so is beyond the ability of research such as this, which is aimed at describing behaviours on an online platform. We venture here a series of interpretations which take us beyond a strict presentation of the data itself. Other ways of reading the data are also perfectly viable of course; but we hope those included below are helpful.

These interpretations begin with the interplay between the two things that this dataset very likely represents: both inauthentic messaging and its amplification organised around two hashtags on the one hand, and organic engagement with the selfsame hashtags on the other. Insofar as the data represents the former, it can tell us something about the nature of the inauthentic campaign. Insofar as it represents the latter, it can tell us something about its reception.

Considering inauthenticity first, then, a number of observations can be made. As has been pointed out by other research too, many of the accounts studied here were created very recently, have very few followers and post a very small amount of original content themselves, preferring to amplify instead. They all follow a common volumetric pattern: a small uptick on the day of the invasion, a large spike on the day of the UN vote and a sharp decrease in the days thereafter.

The linguistic clustering we have performed may have surfaced different kinds of inauthenticity, however. The RED/GREEN/ORANGE have account-level attributes that can correlate with spam, but they also send a large amount of material entirely unrelated to the Russian invasion, but that is in-common with other accounts across these clusters including, most recently, an extremely common amplification of messages related to Kashmir files. A possibility is that these are either one or a series of 'paid-to-engage' services; consumer-facing retail offerings where Retweets, followers and replies (as well as other forms of social media engagement on other platforms) can be readily purchased online. These networks could have been rented to amplify the hashtags and are now being rented to amplify entirely different things.

The accounts found in the BLUE cluster seem to represent a different kind of spam. They also have a very high Retweet:Tweet ratio but were created later on average than accounts from any other cluster and have markedly fewer followers. They also amplify much higher concentrations of invasion-related messaging. These cannot be members of a longer-standing paid-to-engage network, but may have been part of a new network created with the specific intention of amplifying the hashtags in question.

Assuming that some of the activity across our network is indeed inauthentic and represents a deliberate attempt to amplify the hashtag, our research implies several things about this campaign.

- It is possible that the putative identity, and regional and linguistic concentration of accounts engaged in hashtag amplification behaviour simply reflects the *supply* of accounts from paid-to-engage services and that any, claiming to be from anywhere in the world, might have been used to achieve the hashtag amplification effects.

- However, some of the most heavily shared memes and messages shared by the network (and frequently associated with the hashtag) explicitly address regions where these clusters are putatively located, especially India. More have a rhetorical positioning - BRICS-solidarity and Western hypocrisy for instance - that is unlikely to be intended for Western audiences.

- Twitter's trending curation overwhelmingly occurs at the regional (and indeed local) level. To 'game' Twitter to make a hashtag trend is to make it predominantly visible to Twitter users within that region and to have it discernibly trend in that region.

Of course, discerning the intent or motivation of any actor responsible for this activity becomes inherently speculative. Our interpretation, however, is that, both the accounts that participated in the hashtag amplification and the messaging that they sent imply a focus on BRICS and more broadly Asian, South Asian and African audiences. We read in this data an attempt to couch Russia's activities as essentially a defensive reaction to NATO expansionism and the response of the international community as driven by Western hypocrisy and colonialism. Sharp increases in activity occurred on the day of and following the emergency session of the United Nations General Assembly to vote to condemn the Russian invasion of Ukraine and we judge it likely that this data, in part, represents a deliberate attempt to target a series of Asian and African countries in this context.

It is also very likely that our research describes genuine, organic engagement with the hashtag as well, of course, driven by deeper and genuinely held beliefs. This activity may imply where this hashtag (and its associated messaging) resonated. As noted above, there are some indications that the yellow cluster, especially, saw more organic interaction. The accounts send, on average, many more original messages than those in other clusters. There is also the presence of notable highly followed accounts, including @DZumaSambudla, which claims to be the daughter of Jakob Zuma and has over 196,000 followers.

We judge there to be a notable absence in attempts to either reach or address Western audiences. We did not identify the participation of large numbers of accounts with explicit US, UK or European identifiers. Much of the messaging was *about* the West of course, but we observed extremely few messages that seemed to explicitly address any constituencies in the West. English-language messaging was prevalent across the data and almost all of the pro-invasion memetics were in English, so we cannot entirely rule out an attempt to either reach or address Western audiences, but it was notably absent from the data that we looked at.

LIMITATIONS AND CAVEATS

As with any methodology, the approach used here carries with it a series of strengths and weaknesses. When interpreting the data, the following caveats should be regarded.

The cluster descriptions are impressionistic. Other researchers may have drawn different contrasts or similarities from an appraisal of accounts in this network, or may have placed emphasis on different places.

The cluster descriptions do not hold true for every account that is a member of them. We have specifically characterised the 'core' of each cluster, to draw out features that are most distinctive of each cluster. Manual analysis, while the best method for developing holistic impressions, does limit the number of accounts that can be used to characterise clusters. Each cluster will contain 'noise' of different sorts; including accounts that are from different countries, use different languages and do not behave in the way that the overall descriptions of each cluster would suggest. However, it should also be noted that analysis of behaviour after the period of investigation in this report corroborates many of the characterisations developed in this analysis.

For narratives, in addition to themes gleaned from our inspection of accounts (with the caveats identified above), we surfaced and displayed specific content which has been shared in identical or near-identical form across many accounts. This is a good approach for capturing the 'memetic' nature of much online amplification, in which specific content is shared by a wide range of accounts. However it is possible that there are narratives shared through a variety of non-identical posts, which would exist 'below the surface' of the top-shared content and may not be captured by our method.

An obvious limitation but an important one is that this research is confined to Twitter only. Influence campaigns often exploit a number of channels to reach desired audiences however, and in concentrating on just one, we are possibly only describing a fraction of the activity and missing activity on other platforms that could be systematically different.

NEXT STEPS

We have focused here on identifying distinct features of each cluster. However further characterisation of nodes on the borders between clusters could draw out cross-cutting characteristics further. Some of the borders between clusters are non-intuitive from a linguistic point of view - for instance the border between the South Africa and South Asia clusters. We have drawn out some features of similarity between clusters (e.g. account creation date), which could be expanded further by focusing on nodes between clusters.

We continue to monitor these accounts. As noted in our analysis, many of the accounts seem to have moved on from Russia-related material and now post about other topics (for some of the Indian-associated clusters, the Kashmir Files particularly). However, isolating those accounts which continue to post about Russia could add further information about the relationship between these clusters and invasion narratives. What proportion of remaining accounts show behaviour which looks like genuine pro-Russia activism, and what proportion seem to exhibit less authentic behaviour? Monitoring such accounts could give advance notice of narratives which, either through co-ordination and/or through genuine resonance, could suddenly spread back through the entire network.

We are keen to better understand the relationship between the clusters produced by the mono-lingual and multi-lingual encoders. While they broadly agree on the top-level communities, a more fine-grained analysis is needed in order to better understand how to most effectively exploit this message-based approach when dealing with a dataset involving a mixture of different natural languages.

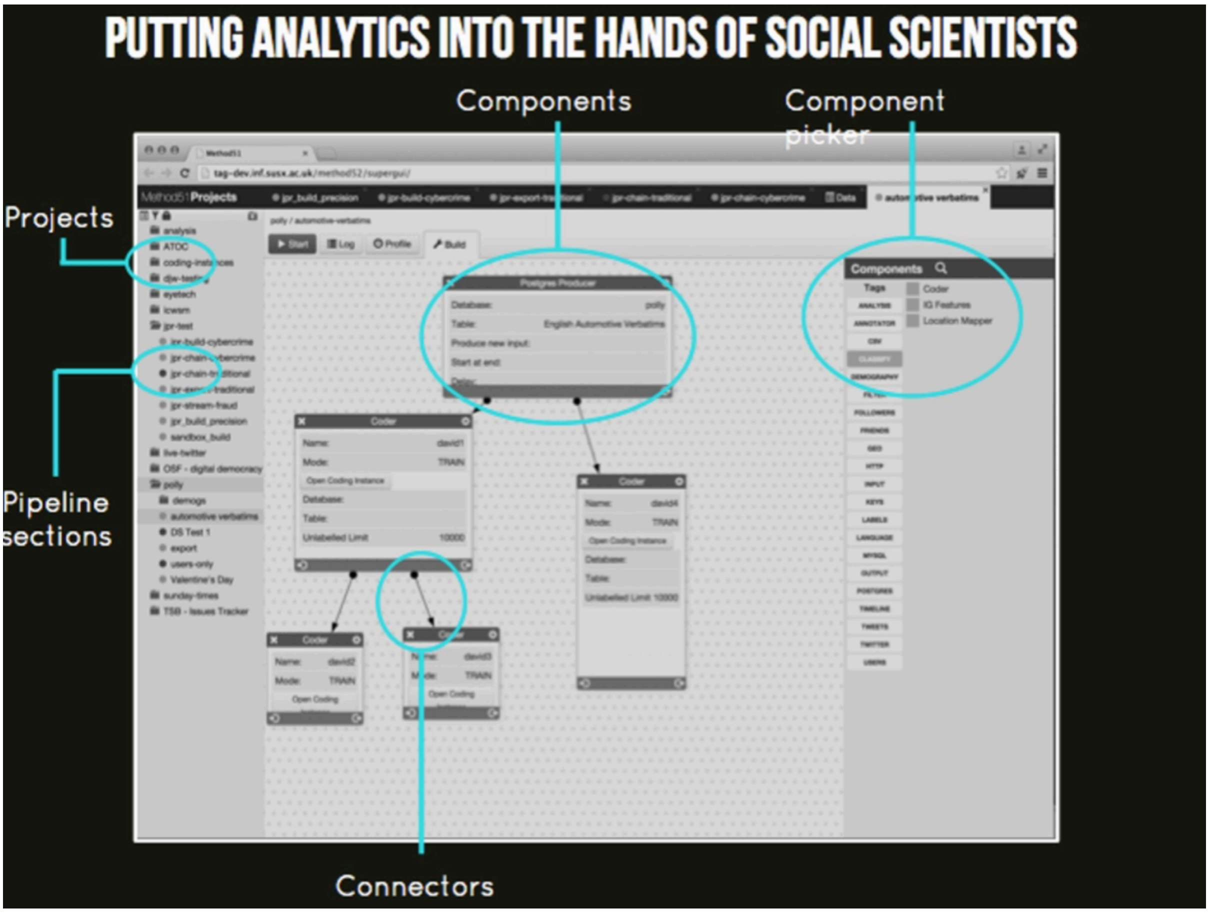
TECHNOLOGY

The core technology used by CASM Technology is a platform it developed called Method52.[26] It is an online research environment that has been built over the course of ten years to allow analysts to collect, analyse and visualise datasets that are very large and unstructured. This is especially the case for large, text-based datasets, such as those drawn from social media, but also includes websites, mainstream media data, free-text survey responses and a variety of internal and proprietary data held by large organisations.

Method52 was built to give social media researchers flexibility to perform analytical operations not possible through the social listening dashboards commonly used across social media marketing and advertising which typically offer a series of pre-set, pre-trained and predefined workflows, counts and measures. The design principle of Method52 is instead to allow analysts to connect and configure components to build their own data collection-analysis-visualisation pipelines through a graphical user interface (GUI). Each of these pipelines is designed to perform a particular task and often a number of pipelines are themselves connected together to create a particular research-driven architecture.

There are 82 components in Method52, and more are added as we conduct new projects that require new capability. Some components are dedicated to collecting data, including from Facebook, Instagram, Telegram, 4Chan, Twitter, YouTube, mainstream news aggregators and websites. Analytical components typically leverage capabilities that have emerged from the field of Natural Language Processing, including for the training of bespoke semi-supervised NLP classifiers, unsupervised semantic clustering, language recognition, Named Entity Recognition and geo-parsing. Other components are built to interact with other APIs external to CASM, including third-party data providers and external models.

[26] For more information, please see https://www.casmtechnology.com/pages/technology



PUTTING ANALYTICS INTO THE HANDS OF SOCIAL SCIENTISTS

The GUI of Method 52

ABOUT CASM TECHNOLOGY

We are a small technology company focussed on developing social media research methods and technology in order to confront online harms. With a range of partners drawn across Governments, journalism, civic society, we work on data-driven responses to climate and health disinformation, online electoral interference, targeted harassment, hate and extremism, state information operations and the illicit trade in wild-life and in support of fact-checking.

With the Institute for Strategic Dialogue since 2015, CASM has developed the ‘Beam’ capability to expose, track and confront information threats online. In 2021, it was the joint-winner of the US-Paris Tech Challenge for innovative approaches to counter disinformation. It has expanded to cover Facebook, Instagram, Telegram, 4Chan, Twitter, YouTube, mainstream news aggregators and hundreds of websites. Beam operates in French, German, Italian, Arabic, Dhivehi, Somali, Spanish, English, Chinese Mandarin, Vietnamese, Farsi, Macedonian, Albanian, Russian, Bengali and Portuguese. Beam has been used by over 350 civil society organisations to confront disinformation during the German (2017, 2021), European Parliamentary (2019), UK (2019), Swedish (2018) and US (2020) Elections. In 2021 Beam was deployed to detect climate disinformation during COP26, producing regular reporting for civic society organisations, the COP presidency, the Counter-Disinformation Unit at the DCMS, the Open-Source Unit at the FCDO and the Rapid Response Unit at the Cabinet Office. Beam has been used in 10 countries overall, and has created over 90 non-public data briefings for partners, including legal, security and government partners, 28 public investigations, 150 media exposes of disinformation and fifteen reports of credible threats to the authorities.

CASM's team mixes together analysts and subject matter experts with data scientists and software developers. The predominant research interest of the team is in foundational Natural Language Processing and social media research methodology innovation. It is broadly drawn from two institutions: the Centre for the Analysis of Social Media at Demos and the Text Analytics Laboratory within the Department of Informatics at the University of Sussex.

No funding was received to conduct this work. All errors and omissions for this research remain the authors' own. None of the views expressed in this report reflect those of any of CASM Technology's partners or funders.

THE AUTHORS

Carl Miller is a co-founder and Partner at CASM Technology. He is also the co-founder of The Centre for the Analysis of Social Media at the think tank Demos where he is Research Director. He's a Senior Fellow at the Institute for Strategic Dialogue, a Visiting Research Fellow at King's College London, an Associate of the Imperial War Museum, a member of the Global Initiative Against Transnational Organised Crime, and a member of the Challenging Pseudoscience group at the Royal Institution. He presents programmes for BBC's flagship technology show Click and is the author of The Death of the Gods: the new Global Power Grab (PenguinRandomHouse), which won the Transmission Prize in 2019.

Chris Inskip is a partner at CASM Technology. He is also currently a doctoral candidate at the University of Sussex researching frameworks for online audience segmentation through representation learning and network analysis. Alongside David, he conducted the modelling used in this report.

Dr Oliver Marsh is a researcher at CASM Technology. Previously, he helped create the counter-disinformation Rapid Response Unit in Number 10 Downing Street and the Cabinet Office, and was Head of European Data Adequacy Assessments at DCMS. He is the founder of The Data Skills Consultancy.

Dr Francesca Arcostanzo is a researcher at CASM Technology and a Senior Digital Research Methods Lead within the Institute for Strategic Dialogue's Digital Research Unit. She holds a PhD in Public opinion, Political communication and Electoral behaviour from the University of Milan, a MA in Government and Public communication, and a MSc in Intelligence & ICT. She has 7+ years of experience in digital research, with a focus on issue politicization, digital polarization and election campaigns. Prior to joining CASM, she was a Digital Analytics Specialist in the Web and Digital Division of the European Central Bank, monitoring and analysing conversations, disinformation and threats around the ECB and its policies.

Professor David Weir is a co-founder and Partner at CASM Technology. He is a professor of Computer Science at the University of Sussex and since 1982 has been involved in the research and development of Natural Language Processing, including in the theory of natural language grammar formalisms and parsing and unsupervised distributional semantics. He has been on the editorial board of the field's leading journal Computational Linguistics, and an area chair for the field's key conferences on several occasions.