CARL MILLER

CHINA'S DIGITAL DIPLOMACY

Reading between the data lines...Carl Miller reports on what was found when bespoke algorithms analysed over 100,000 messages posted by Chinese diplomatic social media accounts.

Across an average week in 2021, hundreds of Chinese diplomatic voices made themselves heard online, posting thousands of messages and provoking hundreds of thousands of reactions, challenges, questions, re-shares and responses. It was often consul generals rather than their more senior ambassadorial colleagues that led the conversation, a new generation of digitally savvy — so-called 'Wolf Warrior' — diplomats, more assertively pushing back against foreign criticism of China.

At the beginning of last year, BBC Monitoring (BBCM) and the CASM Technology set out to study this Chinese public diplomacy as it was happening across social media platforms. The point was to combine BBCM's deep linguistic expertise with CASM's social media research technology to build a research system that was both linguistically and politically sensitive, but also able to operate across the vast expanses of data that social media platforms routinely create.

MULTI-LINGUAL MACHINE LEARNING

For six months, BBCM language teams worked with CASM's technologists and their artificial intelligence research environment (Method52) to train a system of bespoke algorithms that could automatically analyse the messaging of China's diplomatic accounts across the four languages they most often used: French, Arabic, Spanish and English.

Creating a unified framework of themes across all four languages proved to be a significant definitional challenge, requiring a great deal of iterative engagement between each of the four language teams involved. To ensure the framework was reliably applied across all researchers, a small booklet was eventually produced detailing the criteria for inclusion in any theme.

In total, 34 algorithms were trained, most specific to the languages and themes being studied. Narrower themes of more specific language (e.g., COVID-19) tended to be more amenable to rapid training while broader, more linguistically diverse themes (e.g., 'China's culture and people') posed more formidable challenges to the machine learning. Eventually, these models performed with an accuracy of around 80% overall, calculated by comparing classifier outcomes with those of a human on roughly 2,500 randomly selected Tweets and Facebook posts.
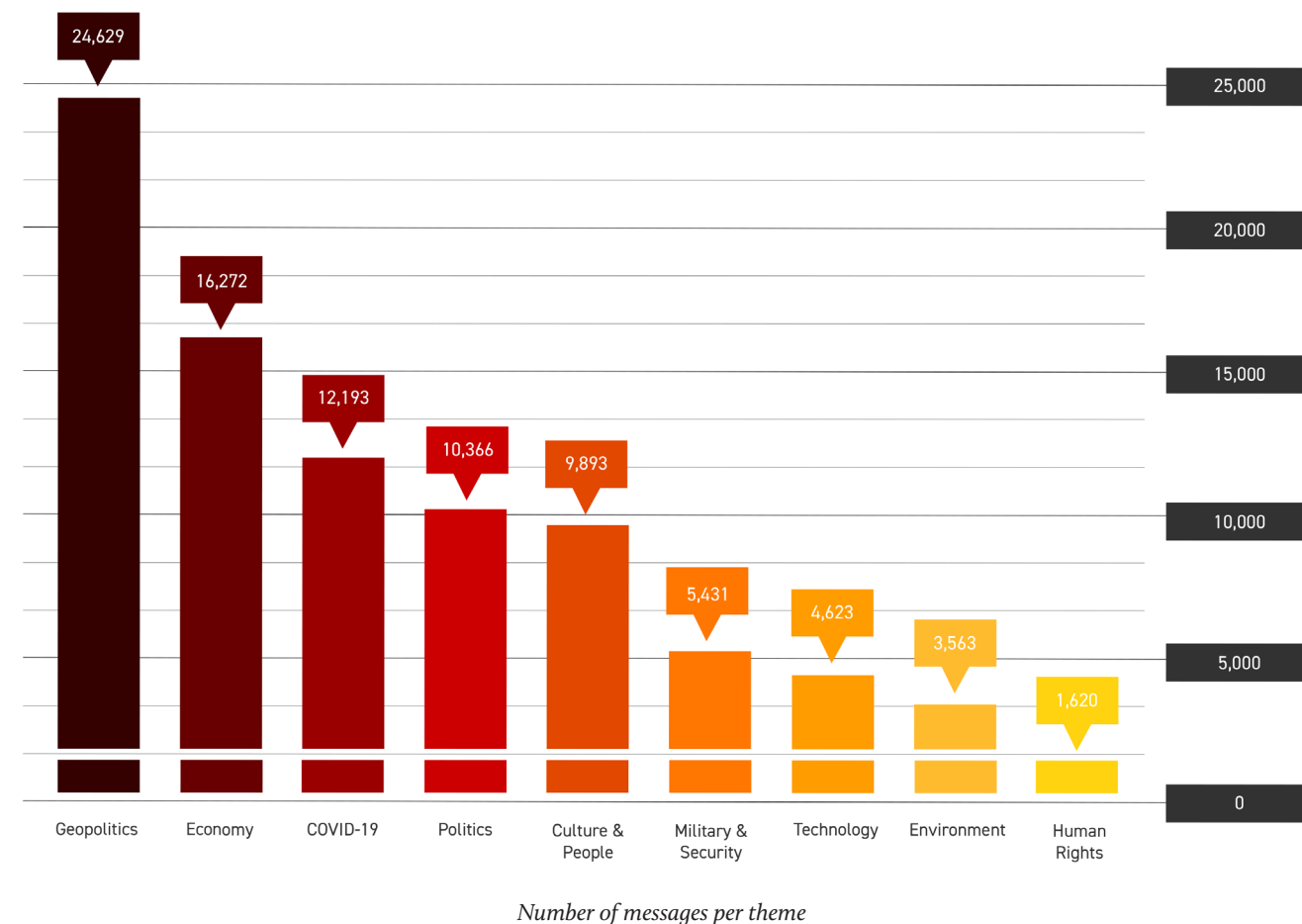
> "Social media platforms are places where China, amongst many other states, are seeking to increase reach and influence watching publics around the world.

Here, we report on the output of this architecture: a window on over 100,000 separate messages sent by 393 Confucius Institutes, ambassadors, consular officials, and accounts from China's foreign ministry on Facebook and Twitter from the start of 2021 to the end of September that year.

The picture it paints is one of clear and reasonably stable global and strategic trends, but, as we'll see, also of important, sometimes dramatic, variation across time, theme, region and language.

GLOBAL PATTERNS

Beijing uses its network of diplomats around the world as the main way of getting its message out. Overall, they sent 102,883 messages which, generally, found an audience. In total, their messages (on Twitter) were retweeted 899,391 times across the period of study; an average of 12.4 times per Tweet. They also received a total of 5,883,361 'likes'; an average of 64.9 likes per message on Twitter and 38.7 on Facebook. Baselining this level of engagement is difficult because it is influenced by a number

Number of messages per theme

of factors; the followers of the messengers, the time when the messages are sent, the kind of messages that they are and the socio-culture mores of the audiences to the messages. However, it does represent a fairly significant audience in absolute terms.

THE THEME

Two thirds of China's messaging fell into one of the nine overall themes:

1.   Geopolitics (26.3%)

More messaging was on Geopolitics than any other theme. This covered the factors, events and themes that governs and structures China's relationship with the world. This included any announcement, meeting or issue covering any of China's bilateral relationships. Especially key here was the China-USA relationship, Taiwan and Hong Kong.

2.   The Economy (17.4%)

This second most popular theme included commentary regarding the production and consumption of goods and services and the supply of money as they relate to China. This covered Chinese economic development, reform, e-commerce, finance, taxation, marketing and advertising, transport infrastructure, trade, energy, mining, agriculture and industry. It also included specific economic programmes and projects, especially the Belt and Road Initiative.

3.   COVID-19 (13%)

This included its impacts, countermeasures, vaccine development, controversy over the origin of its outbreak, 'COVID diplomacy' and its many social, political and economic implications.
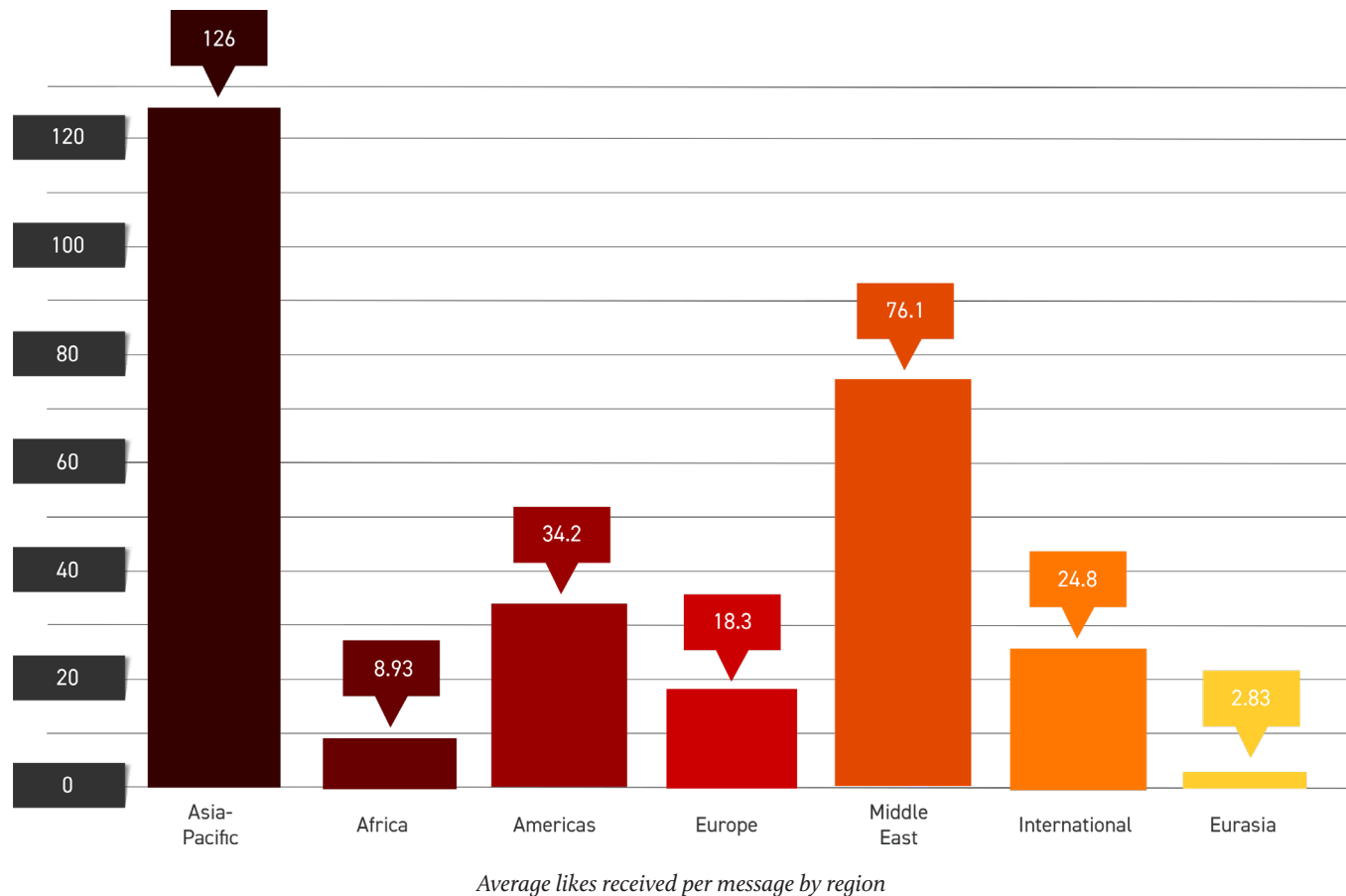
4.   Politics and Society (11.1%)

Messages in this theme were about how political power and influence are distributed and exercised to control, direct or influence events and the actions of people and officials in China. This included messaging related to the Chinese leadership within a domestic context, 'Xi Jinping Thought', corruption, crime, migration, welfare, protest and human rights. It specifically included treatment of ethnic and religious groups such as the Uyghurs of Xinjiang, and inhabitants of Tibet.

5.   Chinese Culture and People (10.6%)

This was a broad theme that spanned China's culture(s), customs, its people(s), activities and events. This includes China's history, its 'food diplomacy', Chinese festivals, sports, the Olympics, Confucius Institutes, foreign students in China, its universities and educational exchanges, and outreach to the global Mandarin-speaking diaspora.

6.   Military and Security (5.8%)

Messages related to the armed, intelligence or domestic security forces. This included armed forces modernisation, nuclear

Average likes received per message by region

weapons, robots, drones, cyber warfare, military exercises, defence diplomacy, policing and counter-terrorism operations.

7. Technology (4.9%)

This theme covered messaging specific technologies, especially the Internet and cyber-security, IP infringement, 5G, Huawei, space exploration, biotech and renewable energy. Also included in this theme were the discussions of technologies regarding security, national economic interest or space and military issues. Any technology related to COVID was excluded.

8. The Environment (3.8%)

The environment was discussed comparatively little, as only the eighth most popular theme. This covered anything relating to climate change, air pollution, environmental deterioration and responses to these challenges that could include policies, technological solutions, and changing attitudes.

9. Human Rights (1.73%)

The least common of the nine themes covered any messaging related to both domestic and international human rights. This included criticism of the West's human rights record, speech, religious and press freedoms, Xinjiang and Tibet, as well as State monitoring surveillance and the social credit system. The other messages tended to fall into a series of smaller and miscellaneous themes or were too event-specific to place into a broader category.

REGIONAL VARIATION

Within the global trends, there was a high degree of variation between China's diplomats based in different parts of the world and in different languages. We present these contrasts, below, as a series of synoptic regional profiles although the reader should note these are only based on our analysis of English, French, Arabic and Spanish and not any other language that accounts from each region might use.

1. Asia Pacific: Key region

The Asia Pacific was the key region where China's digital diplomats were based. It saw sharp increase in message volumes from mid-March onwards and then sustained higher message volumes for the rest of the period of study.

While Europe actually had more accounts than the Asia Pacific, messages from accounts in the Asia Pacific also saw on average around twice as many reposts and likes as messages from any other region, suggesting, perhaps, that China's most visible and influential online sources are disproportionately concentrated in this region.

2. Africa: Multi-lingual messaging about COVID-19

Africa was the only region to see significant volumes of messaging in all four languages and a quarter of all messages (over 23,000) were sent from accounts based there. The messaging tended towards COVID-19 as a theme, but attracted

extremely low levels of engagement with an average of just two reposts per message. Differing levels of engagement are explainable through a number of factors, including the specific visibilities of China's diplomats in the region, the prevailing socio-technological norms of the populations living there and any regional trends around technology use that act as a backdrop to all of the behaviours in this report.

3. The Americas: English and Spanish language soft power

In the Americas, China's diplomatic accounts tended to emphasise the soft power topics of the economy and China's people and its society, with the least concentration on geopolitics. More Spanish messages were sent from accounts in this region than anywhere else. In total, the region saw the third-highest number of messages and also the third-most level of engagement with those messages.

4. Europe: Multi-lingual soft power to a less engaged audience

The activity of China's accounts based in Europe were similar to the Americas. In Europe, too, there was a relative preference for soft power topics across English, French and Spanish. Perhaps the greatest distinction was found not in the messaging but the behaviour of the audience, with China's accounts based in Europe seeing roughly half the levels of engagement and amplification as those based in the Americas, an average of 3.46 reposts and 18.3 likes per message.

5. Middle East: Low number of messages provoking a larger response

The Middle East saw a relatively small number of messages that were highly engaged with, second only to the Asia-Pacific region on average. Naturally enough, this region saw more messaging in Arabic than any other, and also a pronounced emphasis on political themes, including sharp spikes of activity not seen in any other region, in March, April and July (detailed more fully in the report). This region posted the highest proportion of 'Human Rights', messaging, albeit still very low in absolute terms; 4.1% of the region's output, compared to an average across regions of 1.77%.

6. Xinjiang: Most commonly mentioned entity

Much like the themes themselves, the entities being mentioned by China's diplomats changed significantly over time. The project used multi-lingual automated Named Entity Recognition technology to identify these entities — peoples, places or organisations — within the messages collected. Many were related to specific events; mentions of 'Xi Jinping', for instance, increased on at least three occasions in February, late April and July, with the latter occasion also met with an uncharacteristically high number of messages mentioning 'Beijing' and the 'Communist Party of China'. This coincided with the 100th anniversary of the founding of the Communist Party of China, an event we observe in greater detail below.

Strikingly, 'Xinjiang' — the Uyghur Autonomous Region — was the most commonly mentioned entity across five of our nine themes ('culture and people', 'the economy', 'geopolitics', 'human rights' and 'military and security'). Volumes of messages mentioning 'Xinjiang' saw a number of sharp increases throughout the report period, most prominently during the first half of 2021. These 'Xinjiang spikes' (as we call them) occurred across February/March, again in April and a third in late May. Each tended to represent vocal opposition from diplomats and embassies to the UK, US, Canada and EU's coordinated sanctions and blacklisting of several officials over alleged human rights abuses in Xinjiang.

CONCLUSION

Social media platforms are places where China, amongst many other states, are seeking to increase reach and influence watching publics around the world. They know that opinions and attitudes can be formed there, and that they are an opportunity to make their case, raise the issues considered to be priorities, and respond to the criticism and messaging of other states.

The consequences of this are, of course, possibly very wide-ranging. China's messaging matters for activists, journalists, and really any of the planners and the strategists who work on the vast variety of different issues and areas around the world that it touches. For the UK, analysis of this messaging can provide insight into the thinking and priorities of China, as well as the prospect for strategic and tactical counter-communications of their own.

Researching geopolitics must suit the modes that the phenomenon itself now takes and this collaboration was as much interested in the method and technology used by the research as the topic itself. It was an attempt to blend together powerful machine learning with human linguistic and subject matter expertise to create an approach that was both sensitive to language and context, but also capable of handling data scales far beyond those of a manual analyst. In doing so, the contribution we hope to make is of an empirical, data-driven system that can provide a window into the way in which governments and others are using social media platforms to project certain narratives and messages around the world.

Geopolitics and influence, perhaps even statecraft itself, is changing. And as it does so, the ways we understand, track, measure and evaluate these phenomena must be just as data-rich as the environments where they now so routinely play out. The full report will be published on BBC Monitoring's website.

Carl Miller is co-founder of CASM Technology, a team of technologists working to develop social media research methods. He is also the research director of the Centre for the Analysis of Social Media at Demos. Find him on Twitter: @carljackmiller