EU Twitter diplomacy –
a text-mining analysis

Lewin Schmitt, Predoctoral Fellow at Institut Barcelona d'Estudis Internacionals (IBEI)

and PhD Candidate at Universitat Pompeu Fabra (UPF)

lewin.schmitt@upf.edu

https://orcid.org/0000-0001-6439-0333

For discussion at 2021 UACES GF: 07 June 2021

Abstract. This paper uses text mining and sentiment analysis of Twitter posts to explore the EU's diplomatic communication practices and to measure public opinion on foreign affairs. Building on an original dataset of almost one million Tweets from the past five years, this analysis reveals differences in public perceptions of the EU's relationship with China, India, and Russia. Attitudes are most positive in the case of the EU-India relationship, followed by EU-China and EU-Russia. Furthermore, the paper examines hundreds of official EU Twitter accounts regarding their communications on diplomatic relations with these countries. A main finding is that the EU talks about its diplomatic relations in more positive terms than the wider public, though this verbal politeness effect is less pronounced in the case of EU-Russia relations.

Keywords: EU foreign policy; Twitter diplomacy; text mining; public opinion

1. Introduction

Since the end of the Cold War, the European Union (EU) has consolidated itself as a global actor sui generis (Bretherton and Vogler, 1999; Smith, 2014) that today is widely considered one of the most influential international powers (Bindi, 2012; Bradford, 2012; Moravcsik, 2017). Subsequent reforms have transferred more and more foreign policy competences from the Member States to the EU. The creation of the European External Action Service (EEAS) in 2010 has given the EU its own diplomatic service. Led by the High Representative for Foreign Affairs and Security Policy, the EEAS serves as the EU's quasi-foreign ministry. Although the bloc's foreign policy-making may be obscure and complex, its collective voice gives it significant power in international trade and regulatory matters.

To better understand the EU's external relations with three key counterparts from the BRICS pool of major emerging countries (China, India, and Russia), this analysis employs large-scale text mining of Twitter data. While Twitter is not fully representative of the general public in demographic or ideological terms, its study nevertheless allows for important conclusions regarding patterns or shifts in public perceptions.[1] It can also be useful to extract insights about relative levels of attention, salient topics, and associated attitudes. Furthermore, since many opinion leaders, journalists, and multiplicators engage in policy discussions via Twitter, it serves both as amplifier and sentinel of trends in the foreign policy community as well as in the wider population. Lastly, by specifically studying external communications from a political actor's official Twitter accounts, one can draw conclusions on the underlying policy priorities, diplomatic practices, and strategic outlooks. These accounts may be organisational (e.g., the European Commission's *@EU_Commission* handle) or personal – comprising leading politicians as well as designated spokespersons and ambassadors.

Feeding into the emerging literature of Twitter-based opinion mining, this study builds on an original dataset of 927,120 English-language tweets talking about the EU's relationship with China, India, and Russia. In addition, the dataset includes all tweets discussing relations to these three countries from a comprehensive list of over 400 official EU Twitter accounts. The resulting dataset is examined through national

---

[1] For a more detailed discussion of the limitations of this study's approach, consider section 4.3.

language processing (NLP) and linear regression analysis. This multi-layered method provides answers to pertinent questions such as: are there differences in tonality when users and official accounts speak about each of these three relationships, and how do they evolve over time? What were the defining topics and in which way did key events affect public opinion on a relationship? How do the EU's official Twitter accounts talk about diplomatic affairs and which topics do they prioritise?

2. The EU's external relations with China, India, and Russia

The three countries selected for this study are all from the pool of major emerging countries, often coined BRICS. They are amongst the world's largest and most powerful nations and thus play a key role in shaping international relations and geopolitics. By many accounts, China has reached or is about to reach the status of a superpower. It is also the EU's second-largest trade partner. India is a major regional power and the world's largest democracy. Russia, as the heir of the former Soviet Union, fashions itself a global power and, due to its military muscle, is still essential for global security questions. As the EU's biggest Eastern neighbour – and given the historical ties between Russia and many EU Member States – the EU-Russia relations have always received a lot of attention.

Overall, all three are high on the EU's foreign policy agenda and recognized as strategic partnerships. India historically commanded less attention, as it is considered the least challenging of the three partners from a Brussels perspective. In the absence of substantive disagreements, hostilities, or major conflicting interests, the relationship is described in fluffy terms: As per the European External Action Service (EEAS), India and the EU are "long-standing partners", "committed to dynamic dialogue in all areas of mutual interest as major actors in their own regions and as global players on the world stage" (EEAS, 2021b).

The picture is less enthusiastic when looking at the complex relationships the EU has with China and Russia. While the fast-growing Chinese market was long seen as an opportunity for the EU's exporting economies, the past years have given way to a sobering realisation of the many challenges associated with China's rise. Concerns over unequal market access and unfair trade practices, cyber espionage, or the expansion of Chinese influence in the EU's immediate neighbourhood have led to a marked

reassessment of the EU-China relationship. The EU's 2019 Strategic Outlook acknowledged this shift by taking a more assertive stance, defining China simultaneously as "a competitor, negotiating partner, and systemic rival" (EEAS, 2021a). The latter highlights the tense state of bilateral relations.

The EU's relationship with Russia has also deteriorated considerably over the past years. While it has never been an easy one, the 2014 illegal annexation of Crimea and the violent conflict in Eastern Ukraine have caused lasting damage to bilateral relations. Yet, as the EU's largest neighbour and due to historic, cultural, and trade links, "Russia remains a natural partner for the EU and a strategic player combating the regional and global challenges" (EEAS, 2021c). The legal basis is the 1997 Partnership and Cooperation Agreement (PCA), complemented by many sectorial agreements. However, following recent conflicts and political events, some of the policy dialogues and cooperation mechanisms are halted, and sanctions have been adopted by both sides.

Based on this overview of the state of bilateral relations, one would expect the sentiment of tweets to be more positive when it comes to the EU-India relationship compared to the EU-China and EU-Russia relationships. As the empirical section shows, this is indeed the case for both the EU's official accounts and the wider Twitter userbase. Beyond that, there are many nuances and differences that help draw a detailed portrait of the state of the EU's bilateral relations and the way it communicates about them through its Twitter presence.

3. Literature review

But first, this section situates the study within the various strands of literature it engages with, and presents related research. On the one side, this paper explores questions pertaining to the study of public diplomacy. Here, it builds on and feeds into work on the role of social media and the impact of Twitter diplomacy in international relations. On the other side, its methods are related to the recent, but fast-growing scholarship that uses social media data to study real-world phenomena. Here, it connects with efforts to extract information about networks and interactions as well as to measure public opinions based on Twitter data.

3.1 Twiplomacy: the role of social media in international relations and public diplomacy

Much has been written on the important role of public diplomacy in statecraft and international relations (Fitzpatrick, Fullerton and Kendrick, 2013; Dodd and Collins, 2017). Public diplomacy (sometimes referred to as nation branding or strategic communications) means the "power to influence the global discourse" (Collins, DeWitt and LeFebvre, 2019) and communicating to "foreign publics in order to create a receptive environment for foreign policy goals and the promotion of national interest" (Manor and Segev, 2015, p. 93). With the advent of social media platforms, a new form of public diplomacy, digital diplomacy, has emerged. Digital diplomacy is multi-faceted, encompassing everything from negotiations about digital policy files, to the introduction of digital tools in traditional diplomatic practices, to the use of social media for diplomatic purposes (Hocking and Melissen, 2015; Adesina and Summers, 2017; Krzyżanowski, 2020), most notably as a strategic communication tool. This work explores the latter, often referred to as Twitter diplomacy, hashtag diplomacy or Twiplomacy.

Twiplomacy essentially functions in two modes, high-stakes and strategic communications: In high-stakes situations, world leaders, diplomats or governmental agencies communicate to domestic or foreign audiences over Twitter with an immediate, significant diplomatic impact (Wang, 2019). The ominous use of Twitter by former President Trump has made international relationships more fragile, as its directness circumvented many long-held diplomatic norms (Šimunjak and Caliandro, 2019). The fast and unfiltered pace of Twitter interactions can also result in less sway for experts within the governmental bureaucracy, enabling a more top-down (and often erratic) approach to foreign policy-making.

On the strategic communications side, Twiplomacy enables states and international organizations to conduct nation-branding and improve public relations, quickly and effortlessly reaching large foreign audiences. An early and innovative example was the Swedish government's use of the @*Sweden* Twitter handle. It was handed to a different average citizen each week who could then curate it to their liking (Mickoleit, 2014). The EU has also built up a strong presence on various social media platforms – especially on

Twitter. As of early 2021, its registry of official Twitter accounts listed almost 400 handles of high-level political leaders, spokespersons, and institutional entities with a combined followership of over 6 million (European Union, 2021).[2] And in countries where Twitter is not accessible, the EU has gained visibility through other platforms, such as Weibo in China (Bjola and Jiang, 2015).

The role of Twitter communication in international relations and foreign affairs has been the subject of previous studies. Since 2012, BCW's annual "Twiplomacy Study" looks at how heads of state and government, foreign ministers and international organizations use social media channels (Burson Cohn & Wolfe, 2020). They find that in 2020 the governments and leaders of 189 countries had an official presence on Twitter. In addition, the heads of state and government of 163 countries and 132 foreign ministers maintain personal accounts on Twitter. As the use of social media, and in particular Twitter, has become so widespread, the scholarly interest in studying international relations and the conduct of diplomacy through these platforms has grown accordingly (Šimunjak and Caliandro, 2020).

Looking specifically at the EU and its Member States, Kenna (2011) studies the way that social media is used in the EU's public diplomacy toolbox and recommends a clear social media strategy. Since then, the usage of Twitter by EU institutions and political leaders has increased significantly. Kuźniar and Filimoniuk (2017) analyse Twitter communication strategies and efficiencies of European ministers of foreign affairs and their ministries' Twitter channels.

Overall, EU leaders have exhibited a somewhat more restrained handling of their Twitter diplomacy than some of their global counterparts. Therefore, this analysis is less concerned with high-stakes Twitter communications, which would require a qualitative analysis of key tweets and conversations. Inquiries on the EU's strategic communications, on the other hand, are more concerned with continuous messaging and thus ideally suited for a quantitative, large-n text mining analysis. Through such an

---

[2] Many users follow more than one EU account, hence the number of unique followers is supposedly somewhat lower. Nevertheless, the European Commission's main account alone has over 1.4 million (unique) followers. And the three most influential people for foreign policy in the previous Commission, then-Commission President Juncker, then-HRVP Mogherini and then-Council President Tusk boast another 2.5 million followers.

analysis, this paper's contributions are two-fold. First, it measures public perceptions regarding the different bilateral relationships as expressed on Twitter. Second, it investigates the usage of the social media platform Twitter by the EU as a diplomatic tool. For the former, it examines views and attitudes expressed on Twitter as a proxy to measure public opinion towards the EU's relationships with China, India, and Russia. For the latter, it zooms into the EU's external communications, analysing specifically what and how its official Twitter handles speak about these bilateral relations, thereby shedding light on the EU's strategic communications and public diplomacy practices.

3.2 Using Twitter data to investigate diplomatic practices

This work also engages with two strands of scholarly work which both use Twitter data as an empirical base. One is concerned with how Twitter can be used to measure public opinion, and mostly relies on analysing the content of tweets. Another is interested in studying public diplomacy through Twitter data. Here, in addition to content analysis, researchers have used network theory to map connections, interactions, and clusters of Twitter users.

Using Twitter data to investigate international relations and public diplomacy is a relatively young, but rapidly evolving field. Sobel, Riffe and Hester (2016) conduct a content analysis of eight American embassies' Twitter feeds to explore inconsistencies between the embassies' use of Twitter and the State Department mission. Dodd & Collins (2017) examine 41 embassies' Twitter accounts and conduct content analysis to understand their public diplomacy practices. Palit (2018) identifies key characteristics of India's digital communication and explores the contribution of digital communication to India's international stature. Sevin and Ingenhoff (2018) explore the ways in which government-run or funded Twitter accounts of Australia, Belgium, New Zealand, and Switzerland engage in nation branding, proposing a model to assess the impacts of public diplomacy through social media analysis. Collins, DeWitt and LeFebvre (2019) develop methods and techniques to understand the process, reach, and impact of digital diplomacy via Twitter.

Many of the above-mentioned studies are either theoretical, conceptual work or, if empirical, employ mainly qualitative methods. However, the nature of the study object lends itself to quantitative, computational analysis. Social media generates vast amounts

of data and thanks to advanced techniques in natural language processing (NLP), researchers can use that data to generate many interesting insights (see the Methods section for a more detailed explanation). This gave way to a strand of international relations research sometimes coined "computational international relations" (Unver, 2018). Generally, research of this type uses one or a combination of tools such as data mining, NLP, automated text analysis, web scraping, geospatial analysis, or machine learning, to study international relations.

A number of studies have used network analysis and topic modelling to study various aspects of Twitter-based digital diplomacy (Park, Chung and Park, 2019; Ingenhoff, Calamai and Sevin, 2021). In perhaps the broadest such study to date, Sevin and Manor (2019) use social network analysis to discover similarities between traditional embassy networks and Twitter links among ministries and ministers of foreign affairs of 130 countries. Also focused on the US, O'Boyle (2019) explore topics, tones, similarities and differences in tweets during state diplomatic visits in India and the United States. Dubey, Chawla and Kaur (2017) deploy text mining and sentiment analysis to study tweets by two leading Indian political diplomats, exploring how foreign service members are perceived by and interact with online audiences. Šimunjak and Caliandro (2020) examine the ways in which and reasons why EU Member States communicated about Brexit by tracking their UK-based embassy's Twitter activities. They find that "the framing of Brexit on Twitter was deliberate and strategic" and conclude that "Twitter is seen by diplomats as a tool conducive to meeting public diplomacy's aim of relationship-building, but not one to be used for advocacy and influencing interpretation of controversial Brexit issues."

Another application of computational international relations that is also used in this paper is Twitter-based sentiment analysis. It is described more thoroughly in the methods section, but the basic idea is to assign sentiment scores to tweets based on their content, usually on a negativity-to-positivity scale or according to certain emotional categories. Analysis of such sentiment scores can be a useful proxy for real-world phenomena, inter alia for public opinion towards a given topic, as discussed in the next section.

3.3 Using Twitter data to measure public opinion

Such harvesting of Twitter data to measure public opinion is a relatively novel research methodology with many advantages (Klašnja *et al.*, 2017). Notably, it has applications in the area of public health, most recently to track public mood shifts during the COVID-19 pandemic (Sansone *et al.*, 2019; Boon-Itt and Skunkan, 2020; De Caro, 2020; Dyer and Kolic, 2020; Tavoschi *et al.*, 2020). Bian et al. (2016) mined Twitter to investigate the public's perception of the Internet of Things. Validating the reliability of such an approach, several studies have compared trends in Twitter data with traditional public opinion surveys (O'Connor, 2010; van Klingeren, Trilling and Möller, 2020). And Bollen, Mao, and Zeng (2011) found that measurements of collective mood states derived from Twitter feeds are correlated with stock market indexes. Twitter data has also been used to predict electoral outcomes (Gayo-Avello, 2013; Beauchamp, 2017) or to explore gender-based differences in citizens' communications with politicians (Beltran *et al.*, 2021). Yet, the methodology bears inherent challenges that need to be acknowledged. Accordingly, the external validity of some of these studies remains contested in the literature.

Overall, while still in its infancy, Twitter-based study of public opinion can provide novel and reliable insights, but restrictions and limitations of the method need to be considered carefully. First, the Twitter userbase is not representative of the general public and suffers from self-selection bias, resulting in an overrepresentation of younger, more educated, male voices. These biases are even stronger amongst users that engage in political discussions, who were found to have additional demographic and ideological biases (Barberá and Rivero, 2015; Bode and Dalrymple, 2016). Second, structural factors, such as unequal access to digital tools and services (for instance, Twitter is not accessible from within China), and certain online-specific communication traits may impact results. Third, linguistic methods such as NLP rarely work across languages, usually restricting analysis to a body of tweets written in the same language. Lastly, and especially relevant to online discussions on contentious global issues, the use by state and non-state actors of automated bots and coordinated communication campaigns may distort topics and sentiments. The implications of these limitations on the present study are discussed in the methods section together with certain technical challenges. Despite these caveats, a careful and well-calibrated study can generate valuable and fairly accurate insights about perceptions of Twitter users, which – given

their role as amplifiers and leaders of political discourse – can serve as an approximation to the wider public's perceptions (Anstead and O'Loughlin, 2015; McGregor, 2019; Gaisbauer *et al.*, 2020).

4. Methods

The following section describes the methodology and techniques employed in this analysis. First, it briefly discusses the selection of cases from a technical point of view. Then, it presents the data retrieval and processing workflow. Lastly, it discusses the text-analysis methods, notably sentiment and emotion analysis, and the statistical methods.

4.1 Case selection

This study focuses on the EU's bilateral relationships with three key countries: China, India and Russia. The scope of future analysis could certainly be extended to other important partners such as the US or UK. But in light of practical constraints (especially finite computational power and impractical handling of extremely large datasets), the three cases were selected based on the following considerations: 1) the geopolitical importance of these relationships makes their study especially relevant; 2) certain similarities and contrasts in the diplomatic status of each of the three relationships allow for suitable comparisons; 3) the variation of structural characteristics between cases allow to control for systematic measurement errors.

Concretely, there are several methodological advantages of restricting the analysis to the three selected countries: Each of them displays a structural trait that is not present in the other two cases (e.g., Twitter is banned in China, but not in India and Russia; use of the English language is widespread in India, but much less so in China and Russia; while India and China are roughly equal in population size, Russia is significantly smaller). Should any of these differences introduce systematic bias (e.g., in terms of the overall volume of tweets gathered by the analysis), this would immediately become obvious when compared to the other two cases.

4.2 Data retrieval and processing workflow

Figure 1 depicts the mining and analysis workflow with which this study was conducted. The four main steps are: 1) collecting and pre-processing English-speaking tweets from the last five years that mention one of the three relationships; 2) structuring and enriching the data through the use of the R software and additional packages; 3) pre-processing the tweets' text so that it is usable for NLP methods; 4) data analysis and visualisation using ggplot2. This section describes each step in more detail, before discussing the limitations of the approach. Less technically minded readers might want to skip straight to the presentation of the findings (next section).

STEP 1 (python)	STEP 2 (R + packages)	STEP 3	STEP 4
Scraping relevant tweets **Python scraping script** gathers English-language tweets (excl. retweets) from 2016-2020 that: • mention keywords of the three bilateral relationships and either have 1 or more likes • or come from the list of official EU accounts and mention any of the three target countries	**Structuring data, pre-processing text** Inter alia, **dplyr**, **lubridate**, **stringr**, and **tidytext** packages: • remove stopwords, punctuation and special characters • transform to lowercase • convert emojis to words • tokenization • unify synonymous words • No stemming, lemmatization	**Assigning sentiment and emotion scores** **sentimentr** package: • assigns polarity score (from -5 negative to +5 positive) to words and computes a tweet's overall sentiment (unweighted mean score) • incorporates valence shifters and adversative conjunction • Repeat for emotion scores • Cross-validated results with IFANN sentiment lexicon	**Descriptive statistics and regressions** **ggplot2** package for visualisations **tidytext** and **stringr** packages to extract topics and keywords (manually filtering entries without meaningful information) **QuantPsyc** package for linear regression analysis

Figure 1: Data collection and processing workflow.
Source for all figures: own elaborations.

The search queries used to scrape relevant tweets via a python script[3] were restricted to exclude retweets and to only target tweets that Twitter's built-in classification engine labelled as "English-language". The start and end of the query were set to catch all tweets published between 01.01.2016 and 31.12.2020. To limit the scope to a technically manageable volume and to provide a minimum threshold of tweet relevance (especially useful to exclude bots), a tweet had to have generated at least one like. In terms of tweet content, the query was defined to capture tweets that either directly address a relationship (through a hashtag such as "#EUIndia") or that simultaneously matched an EU-keyword list and one from the partner countries. These keyword lists comprised variations of the country name (e.g., "China", "Chinese", and "PRC") as well as

---

[3] Snscrape: https://github.com/JustAnotherArchivist/snscrape

widespread synonymously used names (e.g., "Kremlin", "Moscow") and key political figures and institutions as well as their Twitter handles (e.g., "Modi", "PMOoffice"). For the full queries, see Appendix A.

Next, a list of 416 official EU Twitter accounts was created, including accounts of EU institutions Commission departments, or delegations (e.g., *@EU_Commission*, *@eeas_eu*, *@Europarl_EN*) as well as leading politicians and spokespersons (e.g., *@FedericaMog*, *@extspoxeu*). In addition to currently active politicians, it also includes the handles of members of the previous College of Commissioners. All their tweets of the past 5 years that made mention to any of the three countries (or variations thereof, as per the above-mentioned keyword lists) were scraped. For a more detailed explanation, see Appendix B.

After merging the datasets and dealing with duplicates, this process resulted in a body of 927,120 tweets. In the second step, the data was structured and the text pre-processed to enable NLP-based analysis, following standard text-as-data procedures. In particular, the tweets' content was cleaned by removing stop words (e.g., "the", "and", "for"), punctuation and special characters; transforming all text to lowercase; translating emojis into words; tokenizing the tweets (i.e., breaking it up into individual words) and computing various word frequency statistics.

Step 3 consisted of a sentiment analysis based on Rinker's sentimentr package (Rinker, 2019). Commonly used lexical approaches assign individual words a value on a scale from -5 to +5 depending on their negative/positive connotation from a lookup in a pre-defined dictionary.[4] The word values are then weighted relative to the overall lengths of the tweet. Sentimentr augments this process by taking into account valence shifters (negators, amplifiers (intensifiers), de-amplifiers (downtoners), and adversative conjunctions). Even with such improvements, these methods should not be used to assess sentiment of individual tweets. Rather, they work well when aggregated over a large corpus of data. Hence, for the following analysis, monthly means were used (unless stated otherwise). These monthly aggregations give sufficiently large samples producing reliable and robust results. To cross-validate the results and ensure their

---

[4] The curious reader can find a selection of tweets and their assigned sentiment scores in Appendix C.

robustness against sampling artefacts, sentiment was also computed based on the IFANN dictionary embedded in the tidytext package (Silge and Robinson, 2016) and across various sample sizes and combinations. The results showed remarkable robustness throughout. The process was repeated for the emotion analysis, in which words are associated with one or several of eight emotion categories (anger, anticipation, disgust, fear, joy, sadness, surprise, trust).

In the final step, several simple linear regressions were fitted on the data. For that, dummy variables were constructed either for the account type (official EU account vs other) or the relationship (EU-Russia or not, EU-India or not, EU-China or not). The regressions were run on monthly mean sentiment of the targeted samples. To extract topics and keywords, word frequencies for the targeted samples were computed. In a reiterative process, words void of information were filtered out to produce lists of meaningful, substantive terms. Furthermore, certain synonymous terms were unified based on a manual review (e.g., so that "Ukrainian" and "Ukraine" would be counted together, or "IPR" and "intellectual property").

4.3 Limitations of the study

As acknowledged previously, the approach exhibits several constraints and limitations, which need to be considered carefully when interpreting the results and especially when making inferences about public attitudes. The first set of challenges relates to the representativeness of the data. For one, the Twitter userbase cannot be deemed a representative sample of the wider population and at best can serve as a proxy.[5] Yet, some studies indicate that ideological differences are mostly driven by demographic distortions and disappear when controlling for these (Mellon and Prosser, 2017). Furthermore, Twitter's demographic composition (younger and more male) may not be too far away from the similarly distorted part of the population that is most vocal when it comes to discussing political issues (Barbera and Rivero, 2014). While generalizability is somewhat limited by these biases, the study still serves as a powerful sentinel for attitudes of Twitter users – which are an interesting study object in their

---

[5] For more on Twitter's self-selection bias and the resulting demographic and ideological distortions, consider the following findings: US and UK Twitter users have shown to be younger and wealthier than their respective general publics. In the US, they are also more likely Democrat-leaning, while UK Twitter users are better educated than the average citizen. In both cases, Twitter users are disproportionately members of elites (Blank, 2017; Mellon and Prosser, 2017; Wojcik and Hughes, 2019).

own: for one, they may reflect trends and patterns of the wider population; moreover, many opinion leaders and multiplicators (journalists, politicians, influencers) are using social media to shape conversations, set agendas, and to inform their own views – all of which inevitably feeds back into the offline world.

A second factor limiting the representativeness of the present sample is the restriction to English-language Tweets for methodological purposes. This excludes many relevant Tweets that were written in other languages. The impact is somewhat mitigated by many social media users choosing English as their primary communication language regardless – a trend probably even more pronounced amongst observers and practitioners of the foreign policy community.

Thirdly, one should be cautious of the presence of bots, unauthentic accounts, and coordinated communication campaigns on Twitter, which are especially active in social media discussions related to contentious foreign policy topics. However, they inevitably form a part of the Twitter discourse. Hence, it makes sense to also keep them in the study.

Lastly, the script used to scrape the Twitter data is not an official API and may hence suffer from sporadic omissions. While careful cross-checks have been conducted and a manual sample validation has shown that indeed the full set of queried Tweets has been downloaded, the sheer size of the project means that there may always be singular emissions. However, there are no grounds to believe that those would follow a non-random distribution. Thus, it should not cause any systematic measurement errors.

For the second part of the study, which only looks at official EU accounts, all these representativeness concerns should be less relevant. While not all EU politicians may be present on Twitter, it seems unlikely that there is any correlation of foreign policy views and the choice of being on Twitter. For institutional accounts falling under the EU's corporate communication umbrella, this consideration does not matter at all.

A second set of challenges pertains to the analysis of the collected data: sentiment scoring is prone to misclassify some text, so it cannot be used to reliably assess an individual Tweet. Reasons include the use of irony and sarcasm (users expressing negative sentiments using positive words or vice versa), word ambiguity (words can have different meanings depending on the context), or multipolarity of a given Tweet (expressing both positive and negative opinions towards different actors). However, this study leverages the law of large numbers,

meaning that the misclassification error over- and underrates to a similar extend, hence cancelling each other out on average. Conceptually, the main pitfall of lexicon-based sentiment scoring is probably the sometimes unclear link between the tonality of expressions and the writer's real attitudes on the wider topic of interest. As an illustration, consider Tweets relating to the COVID pandemic, which often include negatively ranked words such as "death". Yet, this does not necessarily mean that the writer has a negative attitude towards the EU's relationship with a given partner, since they could also express grief, condolences, or solidarity. However, here again one caveat is that these measurement errors may cancel each other out to a large extend. Furthermore, the concurrences of predominantly negative terms combined with a specific country will undoubtedly have an impact on public sentiment towards that country, even if the link is not too direct and straightforward.

While the study may allow for inference about the tonality and salience of Twitter discussions regarding the EU's external relations, one has to be careful when drawing conclusions about attitudes of the general public. These caveats are less significant in the chapters exploring the EU's own Twitter diplomacy.

5. Findings

This multi-faceted research method allows for various novel insights into attitudes towards three of the EU's bilateral relationships as well as the way that the EU communicates about these countries. This section presents the findings, starting with a look at the levels and evolution of overall interest in each of the three relationships. Then, the results of sentiment analysis are plotted to visualise shifting patterns and trends in public opinion. The significance of the difference in attitudes towards the three relationships is confirmed by fitting the sentiment scores to regression models. Then, the analysis zooms into the use of Twitter by the EU's official Twitter accounts. It shows that the EU's Twitter accounts talk more positively about the relationships, although again interesting variation can be found. The study reveals significant differences in sentiment towards the three relationships. Lastly, top keywords and topics are extracted using text mining techniques. It brings to light the evolution of focus by the EU over time. It also zooms in to explore which words are most associated with strongly positive/negative tweets, allowing to draw conclusions over politically salient and possibly contentious topics.

5.1 Different patterns of interest

Figure 2 shows the monthly volume of tweets falling into each of the three relationships. Since the overall output on Twitter was relatively constant since 2015 (The GDELT Project, 2019), the growth in tweets talking about each of the three bilateral relations is notable. Beyond a general upwards trend at different absolute levels, some spikes in the EU-Russia and EU-China relationships stand out: In March and July 2018, Twitter users generated a lot of content talking about the EU-Russia relationship, which can be traced to two defining events. First, the attempted poisoning of former Russian military intelligence officer Sergei Skripal on UK soil; second, the geopolitical tensions related to the situation in Ukraine blew up again in July that same year, when the EU prolonged economic sanctions, while President Trump sided with Putin and called the EU a "foe" (Roth, Helmore and Pengelly, 2018). Meanwhile, the EU-China relationship has experienced a dramatic increase of attention in early 2020, when Tweet volume more than tripled. This is clearly connected with the spread of the coronavirus, and while still higher than pre-pandemic levels, the monthly volume has since returned to somewhat more normal heights. There are no such notable spikes on the EU-India curve. Together with the lower level of overall attention (roughly half the number of tweets than for the other two relations), this suggests that over the past five years the EU-India relationship
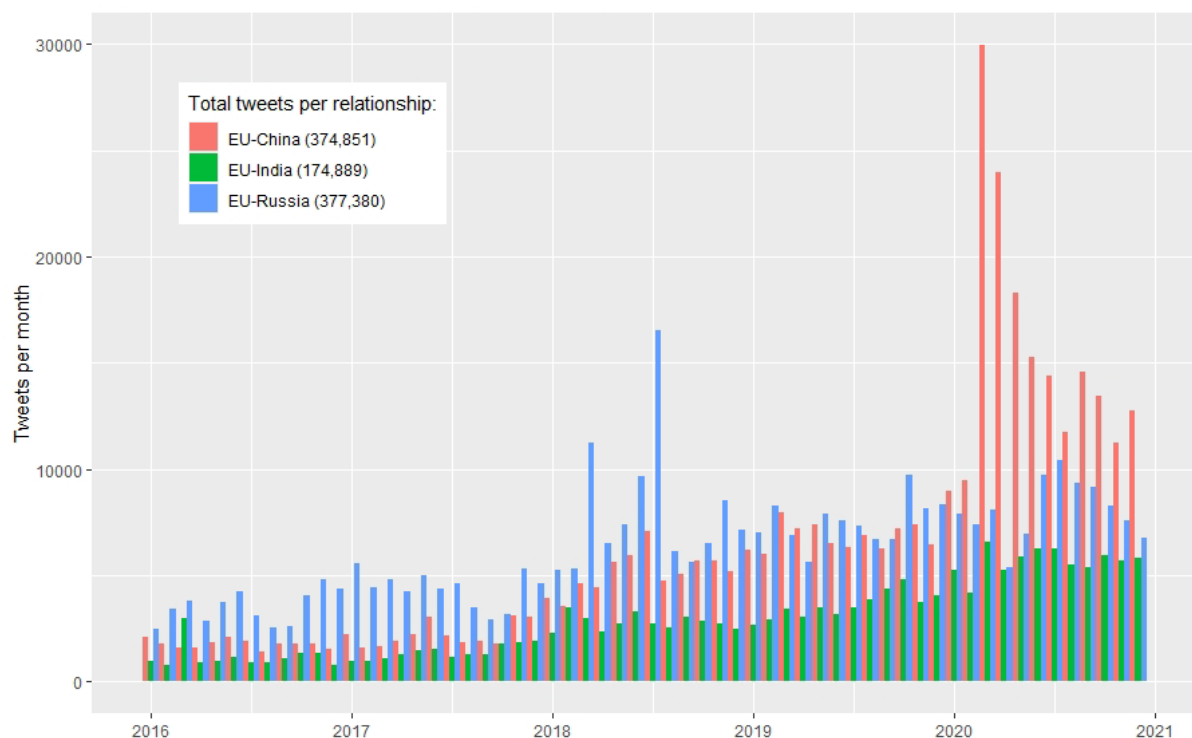

Figure 2: Tweet volume over time, by relationship

16

is less salient to Twitter users, while the EU-China and EU-Russia relationships generate roughly the same level of interest. However, while the interest in EU-Russia relations was initially higher, EU-China has caught up around 2019 and clearly outperformed the others in 2020. At the same time, the monthly rate of tweets relating to the EU-India relationship has risen and almost caught up with tweets talking about the EU-Russia relationship.

5.2 Public perceptions of the state of the EU's bilateral relations

Next, I use tweet sentiment as a proxy for public opinion towards the EU's bilateral relationships. The distribution of sentiment across all tweets follows a slightly positively shifted normal distribution, with a standard deviation of 0.266 and a mean of 0.022. This means that virtually all tweets are scored between -1 and +1. To see a sample of tweets and their corresponding sentiment scores, see Appendix C.

|	Intercept	EU-China	EU-India	EU-Russia
EU-China	0.050		+0.024*	-0.058*
EU-India	0.075	-0.024*		-0.082*
EU-Russia	-0.008	+0.058*	+0.082*	

$^*p < 0.001; R^2 = 0.6433$

Table 1: Sentiment coefficients and significance threshold for combinations of the regression model. Table interpreted as follows: sentiment of column X is higher/lower (cell value and sign) than row Y. Intercept of row Y indicates mean sentiment score.

To see whether there are significant differences in how Twitter users talk about any of the three relationships, a simple linear regression (SLR) model was fitted on the data. For this, a dummy indicating the relationship served as the independent variable (IV), while the sentiment score was the dependent variable (DV). The results from the different combinations are reported in table 1. In all cases, the differences in sentiment are statistically significant at a p=0.001 level with R^2=0.6433. The regressions show that users talk most positively about the EU-India relationship (with an average of 0.075, as indicated by the intercept level), followed by EU-China (0.050) and then EU-Russia (-0.008), which users seem to refer to in the most negative terms.[6]

Looking at the evolution of sentiments over time gives a more nuanced view. Figure 3 plots average monthly sentiments for each of the relationships (blue dots) and draws a

---

[6] To cross-validate the sentiment classification, I repeated the procedure with other dictionary-based sentiment and emotion analysis methods, which all gave the same picture.
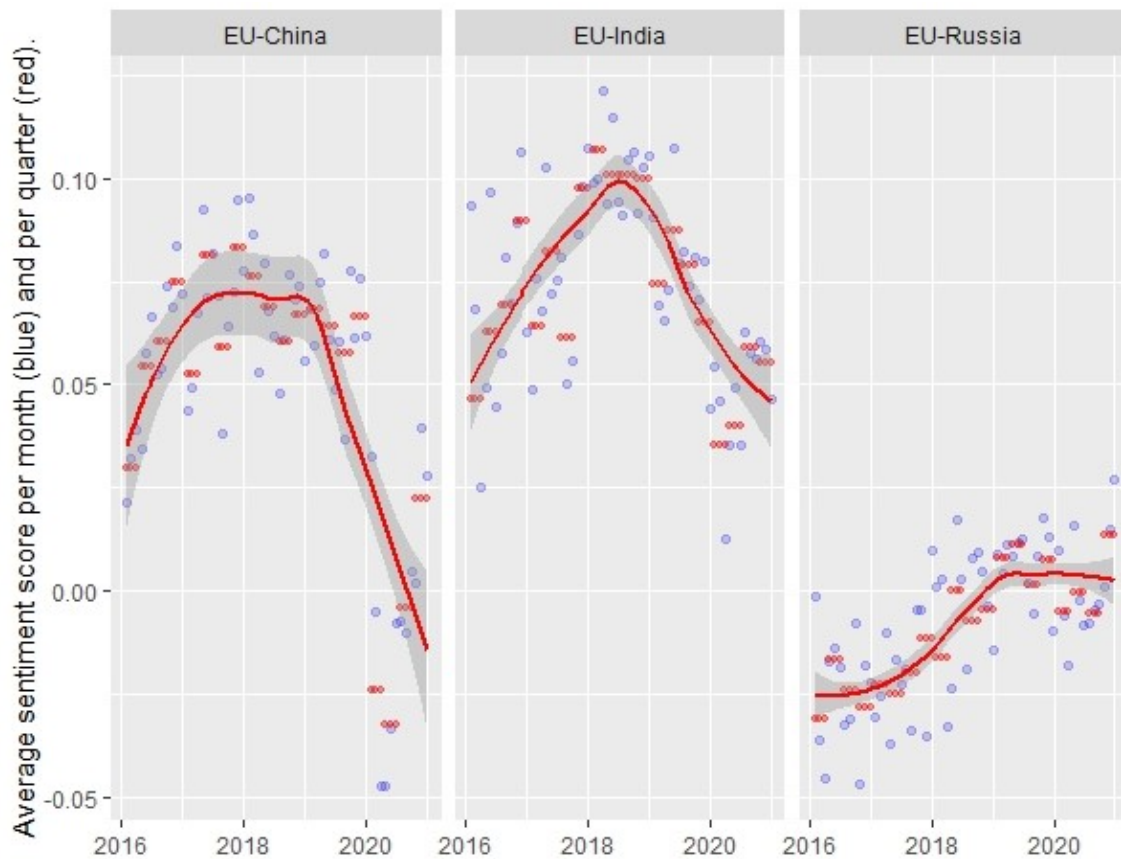
Figure 3: Mean sentiment over time, by relationship

red line connecting average quarterly (i.e., January-March, April-June, etc.) observations (red dots). By widening the interval to a quarterly basis, the means are computed based on a larger sample, which reduces fluctuation and gives a clearer picture of long-term trends. Consistent with the findings from table 1, figure 3 shows lower values for the EU-Russia relationship compared to the other two. But it also brings to light another remarkable development: the strong fall of sentiment in the EU-China relationship since the beginning of 2020, which is driven by tweets related to the emergence and spread of the coronavirus. This is consistent with trends from traditional surveys, which found that unfavourable views of China reached historic highs throughout 2020, including in Europe (Silver, Devlin and Huang, 2020). Furthermore, a curious bump appears in the EU-India relationship, which – while relatively high overall – peaked in 2018. Manually reviewing some of that year's most positive tweets gives us an idea of what has driven

this positive development: In the top 10 are political statements,[7] but also many associated with the popular Korean boy band BTS, whose large fanbase is very active on social media. In 2018, the band celebrated successes on the Indian market, which was reflected in euphoric tweets by its fanbase[8], probably nudging up the average sentiment observed for the EU-India relationship that year.

Figure 4 disaggregates these findings even further. By looking at emotion (instead of sentiment), one can trace the evolution of annual average emotion values for each of the relationships. There are large similarities and parallel trends between the relationships,
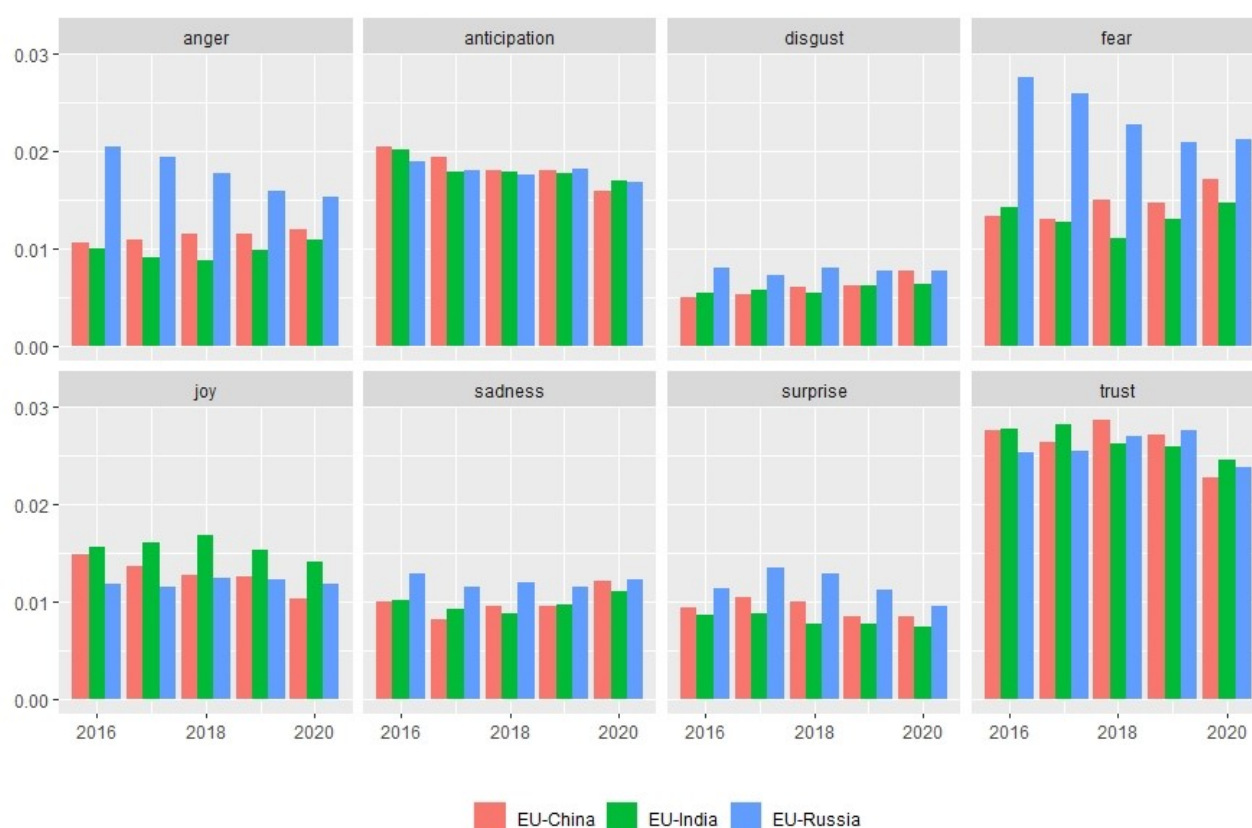

Figure 4: Mean emotions over time, by relationship

---

[7] For example: "*thank you hon'able anurag thakur, president of indo-eu parliamentary friendship group, for very interesting & fruitful meeting. parliamentary diplomacy is an important element of eu-india strategic partnership. looking forward to further contacts. @ianuragthakur @eu_in_india*" or "*@perunchithranar european union's con-federal framework is the right constitutional model for the nations of india to truly realize the promise of freedom & independence promised on 15th august 1947, while maintaining an open market & shared defense & external affairs. #stophindiimposition*"

[8] For instance, the top 10 most positive tweets of that year include adoring references to band members: "*@bts_europe yes he's the most warm hearted, most caring, mos lovable, hard working and supportive, the cutest mochi, jimin i love the way he care for us, love the way he talks haha actually i love his voice, cute voice #meet_bts #meet_jimin @bts_twt*". Because the fan groups are internationally connected, they refer to each other (as, in this case, through the tag of "bts_europe", a European BTS-fan page), which leads to the tweets being included in the automatic sampling process of this study.

suggesting that there could be structural changes in the patterns of how Twitter users speak or which groups of users engage in the discussion. But then there are also marked differences: The strength of emotion in tweets associated with "anger" and "fear" displayed a slight increase in the case of EU-China, remained pretty much constant in the case of EU-India, and declined strongly in the case of EU-Russia. Yet, given the high initial level of "anger" and "fear" in tweets talking about the EU-Russia relationship, despite the drop they continue to be more prevalent than in tweets about the other two relationships. The difference in "anger" and "fear" levels most likely also explains the lower overall sentiment for the EU-Russia relationship, as discussed above.

5.3 Politeness over politics? The EU's diplomatic Twitter language

Besides serving as a sentinel to gauge overall public perceptions towards the three relationships, disaggregated sentiment analysis also allows us to compare the tone of tweets composed by official EU accounts with that of the wider Twitter userbase. For several reasons, one would expect a "verbal politeness effect", i.e., the EU's official Twitter accounts employ more positive (or less negative) language than the wider Twitter userbase.[9] After all, diplomatic norms command a certain vocabulary that leans towards positive terms and coated language, even in the face of underlying issues. Furthermore, trolls and users with strong ideological views likely have a negative impact on the sentiment score of the wider Twitter userbase, thus making the EU's communications even more positive in comparison. As Rasmussen (2009) discussed, the EU's external communication leans on two main approaches: providing information to foreign audiences – which ought to display an objective and neutral tone – and projecting a narrative – by telling success stories and highlighting the positive outcomes of EU action. Negative messaging occurs only rarely, when the EU issues concerns or reservations about certain world events that run counter to its values or human rights.

Are the expectations of a verbal politeness effect confirmed by the present data? Indeed, the sub-set of 8,342 tweets by EU accounts is scored more positively than the other tweets: a mean monthly sentiment of 0.187 compared to 0.020, a difference of +0.167. Fitting a simple linear regression (SLR) model with an "EU account" dummy as IV and

---

[9] On the use of verbal politeness in diplomacy, see Chilton (1990) and Yapparova, Ageeva and Pavol (2019).

	average_monthly
eu_account	0.160*
	(0.009)
Constant	0.023*
	(0.006)
Observations	120
R^2	0.747
Adjusted R^2	0.745
Residual Std. Error	0.047 (df = 118)
F Statistic	349.031* (df = 1; 118)

*$p < 0.01$

Table 2: Simple linear regression on the effect of account type on sentiment

sentiment score as DV confirms the statistical significance of the difference in sentiment between each group's tweets (p=0.001, R^2= 0.747, see table 2). This shows that, consistent with expectations, the EU uses significantly more positive language than the wider Twitter userbase.

So far, the analysis of the EU's verbal politeness effect considered sentiment of tweets across all three relations. However, one could also expect to observe differences in the attitudes expressed in each of the three relationships individually. Table 3 shows the average sentiment for each of the three relationships over the past 5 years, disaggregated by whether the tweets came from an EU official account or not. It stands out that the positive shift in the EU's communications (delta in table 3) is much stronger in the cases of the EU-China and EU-India relationships, and only marginal in the case of the EU-Russia relationship. In other words: when talking about Russia, the language of EU Twitter accounts is not only more subdued but also much more aligned with the wider Twitter userbase. This is also confirmed by SLRs that test the significance of account type (official EU account or not) for each of the relationships: the politeness effect is statistically significant in all three cases but R^2 is much lower in the EU-Russia case, indicating that less of the variation can be explained with the SLR model.

	EU-China	EU-India	EU-Russia
Wider public	0.048	0.071	-0.008
EU official accounts	0.199	0.256	0.030
$\Delta EU - non\text{-}EU$	+0.151*	+0.185*	+0.038*
SLR	$R^2 = 0.7117$	$R^2 = 0.5922$	$R^2 = 0.1336$

*$p < 0.001$

Table 3: Average monthly sentiment by relationship and account type, differences and SLR results

Overall, the verbal politeness effect is around three times stronger in the case of the EU-China and EU-India relationships than in the EU-Russia relationship. Together with the much lower baseline for EU-Russian tweets from official EU accounts (0.030, compared to 0.199 for EU-China and 0.256 for EU-India), this suggests that the EU is less shy speaking critically or negatively about its relations with Russia.

5.4 Salient topics in the EU's bilateral relations

Having explored *how* the EU is talking about the different relationships, the next section looks at *what* it is talking about. Text mining methods such as relative word frequency allow to extract useful information about the shifting focus of topics that the EU's Twitter accounts prioritise.

Figure 5 shows the ranking of the EU accounts' most-used keywords over time.[10] When talking about the EU-China relationship, official accounts emphasize the economic aspects, especially "trade", "intellectual property", "small- and medium-sized enterprises", "market", and "business." Matching the shifting stance of policy-makers in Brussels during the past years, mentions of "cooperation" have given way to "protectionism." This may indicate that the EU is mostly communicating to a domestic audience that has grown sceptical of the EU's trade relations with China, which are perceived as unfair and disadvantageous by many. However, the analysis also brings to light the absence of contentious issues in the bilateral relationship. Common areas of concern that the EU has with China, such as human rights, espionage, or the treatment
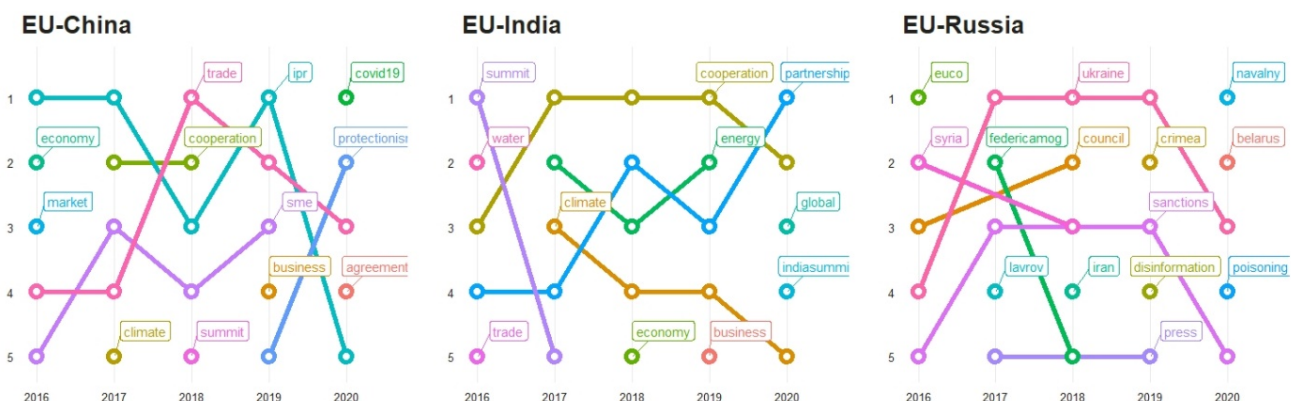
Figure 5: Ranking of most-used keywords by EU accounts over time, by relationship

---

[10] Note: some terms void of information (e.g., "european") and emojis are manually excluded.

of the Uighurs, did not make it into the top ranks, and even "climate" only makes an appearance once in 2017. This could mean that the EU is well aware that Chinese officials may also follow Twitter, and subsequently avoids the diplomatic costs of addressing these concerns publicly.

The EU-Indian relationship gets a lot of soft, friendly language: "cooperation" and "partnership" are ranking high throughout the years. Economic terms such as "trade", "economy", and "business" only make it into the top 5 sporadically. One specific issue of importance seems to be "energy" – which was often mentioned in connection with attributes such as "green" and "sustainable." This also fits into the more prominent use of "climate."

In the case of the EU-Russia relationship, contentious issues are much more prevalent: With top entries covering Belarus, Crimea, disinformation, Iran, the Navalny poisoning, Syria, and Ukraine, it reads like a diplomat's shopping list of foreign policy concerns the EU has with its Eastern neighbour. Clearly, the EU is less concerned with coating differences than it is in the case of China. Instead, it chooses to actively communicate about contentious issues. This may be a consequence of the strategic communications struggle that has tainted the West's relationship with Russia at least since the 2016 election of Trump, in which Moscow was accused of meddling through coordinated social media campaigns (Adams, 2019). Following similar interferences across Europe, including influence campaigns specifically attacking the EU institutions, the EEAS bolstered its dedicated strategic communications unit to combat the spread of disinformation (EEAS, 2018).

An additional observation can be made on the different styles of diplomatic conduct: In the case of EU-Russia, individual politicians (Russian Minister of Foreign Affairs Sergey Lavrov and the EU's former HRVP Federica Mogherini) as well as the European Council are mentioned often. The Foreign Affairs Council as the body overseeing the EU's sanctions policy also ranked highly throughout the years, though did not make it into the top 5. In the cases of EU-China and EU-India, however, political figures do not rank highly, while references to summits appear frequently. Such bi- or multilateral summits have long been the EU's top choice for fostering international cooperation. The fact that

the EU and Russia have not held such a summit since 2014 – the year of the Crimea annexation – highlights the deep and lasting damage this has had on diplomatic affairs.

The last step looks at key terms associated with the most positive and negative tweets from EU accounts regarding each of the three relationships. For table 4, both the top and bottom 10% of the EU's tweets are extracted based on their sentiment score. Of these samples, it lists the six most common terms for each relationship.[11]

	Negative	Positive
EU-China	trade, covid19, rights, issues, economic, market	trade, IPR, protection, cooperation, partnership, agreement
EU-India	ocean, cooperation, summit, people, issues, floods	partnership, energy, sustainable, strategic, clean, climate
EU-Russia	Ukraine, Crimea, disinformation, propaganda, release, sanctions	Ukraine, energy, support, council, security, gas

Table 4: Most common terms in the EU's most positive and most negative tweets, by relationship

In the case of the EU-China relationship, trade-related aspects are very present both in the most positive and most negative tweets, underlining the complex weighting of advantages and disadvantages of the relationship. Not surprisingly, friendly diplomatic terms such as "cooperation", "partnership", and "agreement" occur frequently in positive tweets. Somewhat unexpectedly, "protection" (concerning trade) is also present. Possibly, this is a consequence of the EU attempting to frame itself as a protector of European interests for a domestic audience increasingly sceptical of the bilateral trade relationship. Amongst the most negative tweets, "rights" (concerning both intellectual property and human rights) is mentioned often, suggesting discontent about these issues. Also, the coronavirus has clearly left its mark. Note that this does not necessarily imply a deterioration of bilateral relations. It may just as well reflect an artefact caused by the sentiment measurement method: tweets talking about COVID-19 may be more likely to include terms such as "death" or "killed", which are not necessarily evaluative. But even when used in an objective, purely informative way, the method would attribute a lower sentiment to those scores, hence the interpretation commands caution.

---

[11] Again, terms that are not informative are manually removed.

The positive tweets on the EU-India relationship – in line with expectations – pay lip services to the close and friendly links the EU has with India ("strategic" and "partnership"). In addition, there is a lot of emphasis on "clean energy", "sustainability" and "climate", suggesting that these green topics are framed as an opportunity for EU-India relations. Regarding the negative tweets, "oceans" ranks first, somewhat unexpectedly. A manual review of these tweets reveals them to express condolences and grief after natural disasters striking India – matching the appearance of "floods". Other negative keywords ("cooperation", "summit", "people") also do not offer plausible interpretations of possible strains. This may be because in the case of the EU-India relationship even the most negative tweets are still quite favourable.

The EU-Russia relationship once more shows very contentious issues: "Ukraine" and "sanctions" appear both in the most positive and most negative tweets. This could be because the EU is communicating to different audiences and with different purposes. On the one hand, it uses positive language to express its support of and solidarity with the pro-EU protestors in Ukraine, while using negative language to condemn Russian interference and threatening or announcing sanctions. This would also explain the occurrence of "Crimea" in the negative set. Furthermore, it communicates to a wider, domestic as well as foreign, audience to push back against "disinformation" and "propaganda" from the Kremlin. On the positive side, "energy", "gas", and "security" appear often, highlighting areas in which cooperation is ongoing. "Support" and "council" are in juxtaposition to Russia (for instance, when the EU expresses its support to Ukraine) and should be interpreted carefully.

3. Conclusions

Building on an original dataset of almost one million Tweets talking about the EU's bilateral relationships with China, India, and Russia, this study contributes several novel insights into public attitudes towards these relationships as well as the way that the EU communications about these countries.

In the first place, there was a strong increase in public interest for all three relationships. Singular geopolitical events were shown to drive up the volume of tweets. This was especially visible in the case of the Skripal poisoning and the outbreak of the coronavirus. Over time, attention for EU-Russia affairs has been overtaken by EU-China, with EU-India also catching up quickly.

Sentiment analysis of the tweets showed attitudes of Twitter users to be most positive in the case of the EU-India relationship, followed by China and, lastly, Russia. The same pattern was found when looking at tweets from official EU accounts. However, there was also a marked and statistically significant verbal politeness effect, meaning that the EU talked more positively about its diplomatic relations. While this should not be surprising, it is worth noticing that the effect is much more pronounced in the cases of the EU-India and EU-China relationship, and much less so for EU-Russia affairs. This suggests that the EU is more willing to use firm and resolute language and call out concerns in the latter case. This is in line with the way many observers would describe the strained relationship between the EU and Russia.

The analysis further showed the most salient topics for each of the three relationships. Economic and trade-related aspects dominate the EU-China relation. Substantive issues of disagreement, such as the human rights situation or the treatment of the Uyghurs, does not feature amongst the most used keywords. The EU-India relationship is characterised by the use of more positive and cooperative terms. But disaggregating into words associated with the most positive tweets reveals more: green energy and sustainability are mentioned frequently, hinting at the EU's attempt to frame a positive narrative around these issues. In the case of EU-Russia relations, many contentious issues appear prominently. The EU does not shy away from addressing major foreign policy conflicts such as Syria, Ukraine, Iran and sanctions. It also frequently calls out Russian disinformation and propaganda. On the positive side are areas of ongoing cooperation (although certainly not free of tensions), such as gas and security.

Taken together, this provides a rich and nuanced portrait of the EU's bilateral relations with three key partners. Further research could expand the scope to cover more partners. Certain aspects could also be investigated with even greater detail and explanatory power by fine-tuning the methods. In all this, the limitations of the research approach need to be carefully considered and accounted for as much as possible. For one, the Twitter userbase is not representative of the wider public, so conclusions about public opinion need to be drawn with great caution. Second, the current approach is limited to English-speaking tweets only. It would be of great interest to also study the sentiment and keywords prevailing in tweets written in the partners' language.

Bibliography

Adams, A. (2019) 'Here's What We Know So Far About Russia's 2016 Meddling', *Time*, 18 April. Available at: https://time.com/5565991/russia-influence-2016-election/ (Accessed: 27 February 2021).

Adesina, O. S. and Summers, J. (2017) 'Foreign policy in an era of digital diplomacy', *Cogent Social Sciences*, 3(1), p. 1297175. doi: 10.1080/23311886.2017.1297175.

Anstead, N. and O'Loughlin, B. (2015) 'Social Media Analysis and Public Opinion: The 2010 Uk General Election', *Journal of Computer-Mediated Communication*, 20(2), pp. 204–220. doi: 10.1111/jcc4.12102.

Barbera, P. and Rivero, G. (2014) 'Understanding the Political Representativeness of Twitter Users', *Social Science Computer Review*, 33. doi: 10.1177/0894439314558836.

Barberá, P. and Rivero, G. (2015) 'Understanding the Political Representativeness of Twitter Users', *Social Science Computer Review*, 33(6), pp. 712–729. doi: 10.1177/0894439314558836.

Beauchamp, N. (2017) 'Predicting and Interpolating State-Level Polls Using Twitter Textual Data', *American Journal of Political Science*, 61(2), pp. 490–503. doi: https://doi.org/10.1111/ajps.12274.

Beltran, J. *et al.* (2021) 'Male and female politicians on Twitter: A machine learning approach', *European Journal of Political Research*, 60(1), pp. 239–251. doi: https://doi.org/10.1111/1475-6765.12392.

Bian, J. *et al.* (2016) 'Mining Twitter to Assess the Public Perception of the "Internet of Things"', *PLOS ONE*, 11(7), p. e0158450. doi: 10.1371/journal.pone.0158450.

Bindi, F. (ed.) (2012) *The foreign policy of the European Union: assessing Europe's role in the world*. 2. ed. Washington, D.C: Brookings Institution Press.

Bjola, C. and Jiang, L. (2015) 'Social media and public diplomacy: a comparative analysis of the digital diplomatic strategies of the EU, US and Japan in China', in Bjola, C. and Holmes, M. (eds) *Digital diplomacy: theory and practice*. London ; New York: Routledge (Routledge new diplomacy studies).

Blank, G. (2017) 'The Digital Divide Among Twitter Users and Its Implications for Social Research', *Social Science Computer Review*, 35(6), pp. 679–697. doi: 10.1177/0894439316671698.

Bode, L. and Dalrymple, K. E. (2016) 'Politics in 140 Characters or Less: Campaign Communication, Network Interaction, and Political Participation on Twitter', *Journal of Political Marketing*, 15(4), pp. 311–332. doi: 10.1080/15377857.2014.959686.

Bollen, J., Mao, H. and Zeng, X. (2011) 'Twitter mood predicts the stock market', *Journal of Computational Science*, 2(1), pp. 1–8. doi: 10.1016/j.jocs.2010.12.007.

Boon-Itt, S. and Skunkan, Y. (2020) 'Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study', *JMIR Public Health and Surveillance*, 6(4). doi: 10.2196/21978.

Bradford, A. (2012) *The Brussels Effect*. SSRN Scholarly Paper ID 2770634. Rochester, NY: Social Science Research Network. Available at: https://papers.ssrn.com/abstract=2770634 (Accessed: 19 January 2021).

Bretherton, C. and Vogler, J. (1999) *The European Union as a global actor*. London ; New York: Routledge.

Burson Cohn & Wolfe (2020) *Twiplomacy Study 2020 - Twiplomacy*. Available at: https://twiplomacy.com/blog/twiplomacy-study-2020/ (Accessed: 23 February 2021).

Chilton, P. (1990) 'Politeness, politics and diplomacy on JSTOR', *Discourse & Society*, 1(2), pp. 201–224.

Collins, S., DeWitt, J. and LeFebvre, R. (2019) 'Hashtag diplomacy: twitter as a tool for engaging in public diplomacy and promoting US foreign policy', *Place Branding and Public Diplomacy*, 15. doi: 10.1057/s41254-019-00119-5.

De Caro, W. (2020) 'Infodemia and COVID-19: a text mining analysis', *European Journal of Public Health*, 30(ckaa165.065). doi: 10.1093/eurpub/ckaa165.065.

Dodd, M. D. and Collins, S. J. (2017) 'Public relations message strategies and public diplomacy 2.0: An empirical analysis using Central-Eastern European and Western Embassy Twitter accounts', *Public Relations Review*, 43(2), pp. 417–425. doi: 10.1016/j.pubrev.2017.02.004.

Dubey, G., Chawla, S. and Kaur, K. (2017) 'Social media opinion analysis for Indian political diplomats', in *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence. 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence (Confluence)*, Noida, India: IEEE, pp. 681–686. doi: 10.1109/CONFLUENCE.2017.7943238.

Dyer, J. and Kolic, B. (2020) 'Public risk perception and emotion on Twitter during the Covid-19 pandemic', *Applied Network Science*, 5(1), pp. 1–32. doi: 10.1007/s41109-020-00334-7.

EEAS (2018) *Questions and Answers about the East StratCom Task Force, EEAS - European External Action Service - European Commission*. Available at: https://eeas.europa.eu/headquarters/headquarters-Homepage/2116/questions-and-answers-about-east-stratcom-task-force_en (Accessed: 27 February 2021).

EEAS (2021a) *China and the EU, EEAS - European External Action Service - European Commission*. Available at: https://eeas.europa.eu/delegations/china/15394/china-and-eu_en (Accessed: 26 February 2021).

EEAS (2021b) *India and the EU, EEAS - European External Action Service - European Commission*. Available at: https://eeas.europa.eu/delegations/india/670/india-and-eu_en (Accessed: 26 February 2021).

EEAS (2021c) *The European Union and the Russian Federation*, *EEAS - European External Action Service - European Commission*. Available at: https://eeas.europa.eu/delegations/russia/35939/european-union-and-russian-federation_en (Accessed: 26 February 2021).

European Union (2021) *European Union - Social networks*, *European Union*. Available at: https://europa.eu/european-union/contact/social-networks_en (Accessed: 19 February 2021).

Fitzpatrick, K., Fullerton, J. and Kendrick, A. (2013) 'Public Relations and Public Diplomacy: Conceptual and Practical Connections', *Public Relations Journal*, 7(4), p. 21.

Gaisbauer, F. *et al.* (2020) 'How Twitter affects the perception of public opinion: Two case studies', *arXiv:2009.01666 [physics]*. doi: 10.13140/RG.2.2.32140.95363.

Gayo-Avello, D. (2013) 'A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data', *Social Science Computer Review*, 31(6), pp. 649–679. doi: 10.1177/0894439313493979.

Hocking, B. and Melissen, J. (2015) *Digital Diplomacy in the Digital Age*.

Ingenhoff, D., Calamai, G. and Sevin, E. (2021) 'Key Influencers in Public Diplomacy 2.0: A Country-Based Social Network Analysis', *Social Media + Society*, 7(1), p. 205630512098105. doi: 10.1177/2056305120981053.

Kenna, M. (2011) 'Social Media: Following EU Public Diplomacy and Friending MENA', p. 4.

Klašnja, M. *et al.* (2017) *Measuring Public Opinion with Social Media Data*. Edited by L. R. Atkeson and R. M. Alvarez. Oxford University Press. doi: 10.1093/oxfordhb/9780190213299.013.3.

van Klingeren, M., Trilling, D. and Möller, J. (2020) 'Public opinion on Twitter? How vote choice and arguments on Twitter comply with patterns in survey data, evidence from the 2016 Ukraine referendum in the Netherlands', *Acta Politica*. doi: 10.1057/s41269-020-00160-w.

Krzyżanowski, M. (2020) 'Digital Diplomacy or Political Communication? Exploring Social Media in The EU Institutions from a Critical Discourse Perspective', in Bjola, C. and Zaiotti, R. (eds) *Digital diplomacy and international organisations: autonomy, legitimacy and contestation*. London ; New York: Routledge, Taylor & Francis Group (Routledge new diplomacy studies).

Kuźniar, E. and Filimoniuk, N. (2017) 'E-Diplomacy on Twitter. International Comparison of Strategies and Effectivity', *Social Communication*, 3(2), pp. 34–41. doi: 10.1515/sc-2017-0010.

Manor, I. and Segev, E. (2015) 'America's Selfie: How the US Portrays Itself on its Social Media Accounts', in.

McGregor, S. C. (2019) 'Social media as public opinion: How journalists use social media to represent public opinion', *Journalism*, 20(8), pp. 1070–1086. doi: 10.1177/1464884919845458.

Mellon, J. and Prosser, C. (2017) 'Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users':, *Research & Politics*. doi: 10.1177/2053168017720008.

Mickoleit, A. (2014) *Social Media Use by Governments: A Policy Primer to Discuss Trends, Identify Policy Opportunities and Guide Decision Makers*. OECD Working Papers on Public Governance 26. doi: 10.1787/5jxrcmghmk0s-en.

Moravcsik, A. (2017) 'Europe Is Still a Superpower', *Foreign Policy*, 13 April. Available at: https://foreignpolicy.com/2017/04/13/europe-is-still-a-superpower/ (Accessed: 23 February 2021).

O'Boyle, J. (2019) 'Twitter diplomacy between India and the United States: Agenda-building analysis of tweets during presidential state visits', *Global Media and Communication*, 15(1), pp. 121–134. doi: 10.1177/1742766518818859.

O'Connor, B. (2010) 'From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series', p. 8.

Palit, P. S. (2018) 'India's Use of Social Media in Public Diplomacy', *Rising Powers Quarterly*, 3(3), pp. 151–171.

Park, S., Chung, D. and Park, H. W. (2019) 'Analytical framework for evaluating digital diplomacy using network analysis and topic modeling: Comparing South Korea and Japan', *Information Processing & Management*, 56(4), pp. 1468–1483. doi: 10.1016/j.ipm.2018.10.021.

Rasmussen, S. B. (2009) 'Discourse Analysis of EU Public Diplomacy Messages and Practices', *DISCUSSION PAPERS IN DIPLOMACY*.

Rinker, T. W. (2019) *sentimentr: Calculate Text Polarity Sentiment*. Buffalo, New York. Available at: http://github.com/trinker/sentimentr (Accessed: 27 February 2021).

Roth, A., Helmore, E. and Pengelly, M. (2018) 'Trump calls European Union a "foe" – ahead of Russia and China', *The Guardian*, 15 July. Available at: http://www.theguardian.com/us-news/2018/jul/15/donald-trump-vladimir-putin-helsinki-russia-indictments (Accessed: 24 February 2021).

Sansone, A. *et al.* (2019) 'The Sentiment Analysis of Tweets as a New Tool to Measure Public Perception of Male Erectile and Ejaculatory Dysfunctions', *Sexual Medicine*, 7(4), pp. 464–471. doi: 10.1016/j.esxm.2019.07.001.

Sevin, E. and Ingenhoff, D. (2018) 'Public Diplomacy on Social Media: Analyzing Networks and Content', *International Journal of Communication*, 12.

Sevin, E. and Manor, I. (2019) 'From Embassy Ties to Twitter Links: Comparing Offline and Online Diplomatic Networks', *Policy & Internet*, 11(3), pp. 324–343. doi: https://doi.org/10.1002/poi3.199.

Silge, J. and Robinson, D. (2016) 'tidytext: Text Mining and Analysis Using Tidy Data Principles in R', *The Journal of Open Source Software*, 1(3), p. 37. doi: 10.21105/joss.00037.

Silver, L., Devlin, K. and Huang, C. (2020) *Unfavorable Views of China Reach Historic Highs in Many Countries*. Pew Research Center. Available at: https://www.pewresearch.org/global/2020/10/06/unfavorable-views-of-china-reach-historic-highs-in-many-countries/ (Accessed: 26 February 2021).

Šimunjak, M. and Caliandro, A. (2019) 'Twiplomacy in the age of Donald Trump: Is the diplomatic code changing?', *The Information Society*, 35(1), pp. 13–25. doi: 10.1080/01972243.2018.1542646.

Šimunjak, M. and Caliandro, A. (2020) 'Framing #Brexit on Twitter: The EU 27's lesson in message discipline?', *The British Journal of Politics and International Relations*, 22(3), pp. 439–459. doi: 10.1177/1369148120923583.

Smith, K. E. (2014) *European Union foreign policy in a changing world*. 3rd edition. Cambridge Malden: Polity.

Sobel, M., Riffe, D. and Hester, J. B. (2016) 'Twitter Diplomacy? A Content Analysis of Eight U.S. Embassies' Twitter Feeds', *The Journal of Social Media in Society*, 5(2), pp. 75–107.

Tavoschi, L. *et al.* (2020) 'Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy', *Human Vaccines & Immunotherapeutics*, 16(5), pp. 1062–1069. doi: 10.1080/21645515.2020.1714311.

The GDELT Project (2019) *Visualizing Eight Years Of Twitter's Evolution: 2012-2019*. The GDELT Project. Available at: https://blog.gdeltproject.org/visualizing-eight-years-of-twitters-evolution-2012-2019/ (Accessed: 27 February 2021).

Unver, H. A. (2018) 'Computational International Relations', p. 20.

Wang, C. (2019) *Twitter Diplomacy: Preventing Twitter Wars from Escalating into Real Wars*, *Belfer Center for Science and International Affairs*. Available at: https://www.belfercenter.org/publication/twitter-diplomacy-preventing-twitter-wars-escalating-real-wars (Accessed: 21 February 2021).

Wojcik, S. and Hughes, A. (2019) *Sizing Up Twitter Users*. Pew Research Center. Available at: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ (Accessed: 9 June 2021).

Yapparova, V., Ageeva, J. and Pavol, A. (2019) 'Verbal Politeness as an Important Tool of Diplomacy', *Journal of Politics and Law*, 12, p. 57. doi: 10.5539/jpl.v12n5p57.

Appendix

Appendix A: Search queries

Note that capitalization and hashtag variants of the keywords are also captured. As explained in the methods section, in addition to meeting the following content criteria, tweets also had to meet certain other criteria (English language, not being a retweet, having one or more likes).

- For the **EU-China relationship**, a tweet had to mention either "EUChina", "ChinaEU" or a combination of China, Chinese, PRC, Beijing, Jinping, Xijinping, Keqiang, CCP, ChinaEUmission, or MFA_China and EU, Europeanunion, European Union, Europeancommission, European Commission, EU_commission, EEAS, Brussels, Europe, European, JunckerEU, vonderleyen, FedericaMog or JosepBorrellF.
- For the **EU-India relationship**, a tweet had to mention either "EUindia", "IndiaEU" or a combination of India, Indian, Pmoindia, Narendramodi, Modi, Indembassybru, Indiandiplomacy, MEAIndia, Santjha and EU, Europeanunion, European Union, Europeancommission, European Commission, EU_commission, EEAS, Brussels, Europe, European, JunckerEU, vonderleyen, FedericaMog or JosepBorrellF.
- For the **EU-Russia relationship**, a tweet had to mention either "EURussia" or "RussiaEU", or a combination of Russia, Russian, Kremlin, Putin, Kremlinrussia_E, GovernmentRF, MedvedevrussiaE and EU, Europeanunion, European Union, Europeancommission, European Commission, EU_commission, EEAS, Brussels, Europe, European, JunckerEU, vonderleyen, FedericaMog or JosepBorrellF.
- For the **tweets from official EU accounts**, a tweet had to come from an official EU account (see Appendix B) and mention one of the following keywords: EUChina, ChinaEU, China, Chinese, PRC, Beijing, Jinping, Xijinping, CCP, ChinaEUmission, MFA_China, Keqiang, EUIndia, IndiaEU, India, Indian, PMOIndia, Narendramodi, Modi, Indembassybru, Indiandiplomacy, MEAIndia, Santjha, EURussia, RussiaEU, Russia, Russian, Kremlin, Putin, KremlinRussia_E, GovernmentRF, MedvedevRussiaE.

Appendix B: List of official EU accounts

The official accounts were scraped from the EU's website, accessed on February 19, 2021: https://europa.eu/european-union/contact/social-networks_en This gave me 382 Twitter handles. In addition, I manually added 22 missing accounts of those Commissioners from the Juncker-Commission (2014-2019), which were in office during the period of inquiry. I also included 12 accounts from key EU officials and foreign policy spokespersons that were in office during the period of inquiry. This resulted in a final list of 416 EU accounts.

Appendix C: Sample tweets and sentiment scores

Below are some randomly drawn examples from tweets with the highest and lowest assigned sentiment value (score in brackets):

- *"This is a brilliant & very, very helpful work for all European citizens about the triangle between Europe, the us & china & its future! (attention 20 p. But very worth reading). Congrats to @cer_ianbond 👏 @sophiabesch 👏@leoschuette👏"* **(+1.8874805)**
- *"Russia can strengthen its geopolitical positioning in Europe in some respects by seeking to cooperate more with Germany, its most important European partner. @dmitritrenin explains the importance of this strategic partnership"* **(+1.4422306)**
- *"ugo astuto, ambassador, delegation of the European union to India shared his thoughts on how the equal role of girls in the society can ensure a more equitable, inclusive and sustainable future. Watch this space to learn more about the #girlsgetequalchallenge"* **(+1.3218321)**
- *"I welcome chinas willingness to join covax, we are all in this together multilateralism is key to reaching our global goal of access to vaccines everywhere for everyone who needs them we look forward to working with china and other partners on this"* **(0.3670083)**
- *"brussels and china need to work together to boost demand not legislate against dumping"* **(0.1532065)**
- *"i find it crazy how the eu would give the china more economic power in europe given china recent bullying of australia if biden is smart hell emphasize our allies in the pacific instead of the ones in Europe"* **(-0.316227766)**
- *"europe once saw xi china as a hedge against trump not anymore"* **(-0.316227766)**
- *"@rt_com and Russia is an example what for?! Warmongering, murders, corruption, crimes, endless stupid lies, desinformation, occupation of European countries, killing civilians, breaking relations with the civilised world, more murders, more lies.."* **(-1.8031223)**
- *"fuck vodka, fuck putin, fuck winter, fuck those stupid fucking hats, fuck invading small eastern European democracies, fuck. You."* **(-1.6403225)**
- *"i wish more people would relentlessly remind everyone of russian interference in british/european/american politics."* **(-1.2950000)**
- *"This confusion, lack of unity and obvious weakness on the European side stimulates Moscow and Beijing to double down with their efforts: more intimidation, more disinformation, more economic blackmail -- as this seems to work."* **(-1.2244999)**