Appendix

American Twitter users have ideological differences of opinion about the War in Ukraine

August 1, 2024

Contents

A Extended Discussion of Methods and Materials

A.1 Data Collection

We collect the universe of tweets discussing the war in Ukraine using a broad keyword search via the Twitter Application Programming Interface (API). We download tweets posted during February 24–28, 2022, and contain at least one keyword (Ukraine, Ukraina, Ukrainian, Ukrainians, Kyiv, or Kiev), resulting in about 6.5 million tweets collected. We chose this period because it marks the first 5 days of the war, including Russia's initial invasion. This strategy allows us to examine public responses to the war as the opinions were still forming. We pre-process the text of the tweets by removing hyperlinks, mentions, extra space, and new lines. Then, we use *cld3*, a language identification package released by Google to detect the language of tweets and select only American English tweets. Using this method, we retain an American English corpus containing about 4.4 million tweets.

Measuring User Ideology

Next, we estimate the political leaning of each tweet by estimating the ideology of the user who posted it. Each tweet in our sample contains meta information that includes usernames. We estimate the political ideology of Twitter users based on who they follow. We begin by creating a list of 75 conservative and 75 liberal media accounts, including political commentators, talk show hosts, and journalists. We limit this list to popular accounts with a blue checkmark verification status on Twitter with more than 100,000 but fewer than 1 million followers. The data collection took place before the Twitter policy change that made the blue checkmark purchasable. This restriction ensures that our list contains mainstream accounts with clear ideologies. We evaluate the ideology of users in our data by matching the number of media accounts they follow and taking the difference between the number of conservative- and liberal-leaning accounts they follow. Formally, we define a user's partisan score as $\frac{L-C}{L+C}$, where L and C represent the number of liberal- or conservative-leaning media accounts they follow.

We classify users with a score of less than -0.5 as conservative, and those greater than 0.5 as liberal. We consider users with scores between -0.5 and 0.5 as moderate. We then connect each tweet to the user's ideology and retain only tweets for which we could identify the users' ideology. This process gives us 1.8 million tweets (almost half of the English corpus) that we use in the main analysis. Because the list of media accounts are U.S. based, users who follow them and thus are included in our analysis also are likely to be Americans or people who are connected to this country. As a further validation check of whether these accounts belong to Americans or not, we fit a structural topic model on a random sample of tweets included in our in the main analysis (in sample) and excluded (out of sample). We find that in-sample tweets have a higher proportion of topics related to American politics and domestic concerns than out-sample tweets.

This method exploits established homophily theory in social network analysis (*1*), which describes individuals' tendency to associate with others who have similar values or social status. Recent work finds evidence to support homophily theory in the realm of social media by showing that people tend to recognize and follow those with similar political preferences (*2*). As US Congress members remain united so far while various media outlets adopt different framings of the Ukraine issue, it maximizes our chance to observe ideological polarization by measuring users' ideology according to the types of news accounts they follow rather than politicians' Twitter handles (*3*).

We validate this measure in three ways. To begin with, we collect elected US officials' Twitter handles and use our list of media accounts to estimate their party affiliations. Assuming that Republicans are conservative and Democrats are liberal, we find that elected officials' party affiliations perfectly match the ideological types of the news accounts they follow. Democratic officials follow more liberal-leaning accounts than conservative-leaning accounts; Republican officials behave the opposite.

Next, we download the universe of tweets that include one of two ideological salient hashtags during the run-up to the 2022 U.S. midterm elections (January-November 2022): (1) #voteprochoice, a liberal hashtag used by liberals to encourage fellow liberals to vote for pro-choice candidates in the 2022 midterm elections; and (2) #voteprolife, a conservative hashtag used by conservatives to encouraging fellow conservatives to vote for pro-life candidates in the 2022 midterm elections. These hashtags focus on an ideologically salient issue in a costly way by using the word "vote." A conservative ideologue, for example, would be reticent to use the hashtag #voteprochoice since it denotes electoral support for a liberal position. We estimate the ideology of each tweet that either hashtag using the method we outline in the paper. This validation shows us that our measure performs as expected. Of the 14,313 tweets that use the #voteprochoice hashtag, our measure can identify the ideology of almost 60% of tweets. Among them, 8,157 tweets (97.9%) are by accounts coded as liberal. Similarly, of the 5,134 tweets that use the #voteprolife hashtag, our measure can identify the ideology of about 50% of tweets. Among them, 2,253 tweets (94.6%) are by accounts that code as conservative.

Finally, we also conduct a survey on Amazon MTurk to validate our measurement strategy. In the survey, we collect information on subjects' political participation and Twitter usernames. This data shows that subjects' voting records and ideologies are correlated with the types of news accounts they follow on social media. Subjects who voted for Biden in 2020 or defined themselves as Democrats usually follow more liberal-leaning accounts than conservative-leaning accounts on our list.

A.2 Geographic Location

Our sample is likely broadly representative of American political discourse on Twitter for three reasons. First, we use American English to refine our sample. Individuals who post in American English participate in the broader American online political discourse on a given subject, even if they are not voters. Second, by limiting our analysis to those accounts that post in American English *and* follow at least one political account, we are limiting our sample to users that are at the very least participating in the political conversations about Ukraine, even if they are not doing so on an active or permanent basis. Finally, as a robustness check, we download and analyze the self-reported geographic location of all the tweets in our sample. 0.06% of our sample (2,648 Tweets)—a tiny fraction of our overall sample—did not originate in the United States, according to this user-reported metric. However, we still retain these users because, as discussed, these individuals may still participate in relevant American political discourse, even if they are not located in America. Some of these users may be traveling abroad overseas. Others may be living oversees permanently. Indeed, there are nearly 3 million eligible American voters living abroad. Regardless, we keep these tweets in the sample because the content of the tweet was a part of the broader political conversation we are interested in analyzing.

A.3 Clustering Tweets

We apply Top2Vec, an unsupervised clustering algorithm, to classify our large collection of social media posts (*4*). This method is especially well suited to analyzing short political texts (*5*). Compared to other topic modeling methods like latent Dirichlet allocation (*6*),our method does not require pre-determination of the number of topics or tokenization and stemming of the documents. Instead, the model retains the order of words when learning fixed-length distributed vector representations of documents (*7*) with neural networks. When the learning is done, documents that are semantically similar are supposed to be close in the vector space. The model assumes that semantically similar documents indicate an underlying topic. Therefore, it maps these document vectors to a lower-dimensional space using Uniform Manifold Approximation and Projection (*8*) and automatically find dense areas in that space using a Hierarchical Density-Based Clustering technique (*9*). Documents in the same dense areas are assigned to the same cluster.

A.4 Aggregate Topic Labeling

Recall that the Top2Vec model groups together tweets with similar words and meanings, which generates a large number of clusters. Figure A1 presents information about the top 50 clusters of tweets discussed by different ideological groups and all users.[1] We find that these clusters can be further aggregated into a few categories. For example, tweets using words such as "lord," "amen," "Jesus," and "God" were grouped together, and upon further analysis, it was determined that such tweets were generally sent by people expressing their sympathy and support for Ukraine and the Ukrainian people. The procedure of aggregate topic labeling and validation is described below:

1. We review a significant number of tweets from the 100 largest clusters and identify 7 meaningful categories (which we call substantive topics) related to the discussion surrounding the war in Ukraine, including Ukraine support, Russia support, domestic politics, misinformation, etc.

2. We train two RAs to aggregate all clusters into substantive topics. Each tweet has a matching score indicating how far it is from the center of its assigned cluster. The higher the score, the more representative the tweet is of the whole group. They read the ten most representative tweets from each cluster and determine which category that topic belongs to.

3. If none is applicable, we exclude that cluster from our analysis. This results in dropping about 34% of the tweets from the substantive topic analysis.

4. We compare two RAs' labels and retain those they both agree on.

5. For clusters for which two RAs choose different labels, the three authors code them again and adopt the topic coding a majority of the authors agree upon.

6. Since it indicates valence, we validate the Ukraine support measure by hand-coding 500 randomly selected tweets of those our NLP model classified as supportive of Ukraine. We find that our NLP model accurately predicted the valence of 90.8% of the tweets.

[1]We label each of the top 50 clusters by manually examining representative tweets.

Figure A1: Proportion of Top 50 Clusters of Tweets about Ukraine-Russian War by Groups of Users

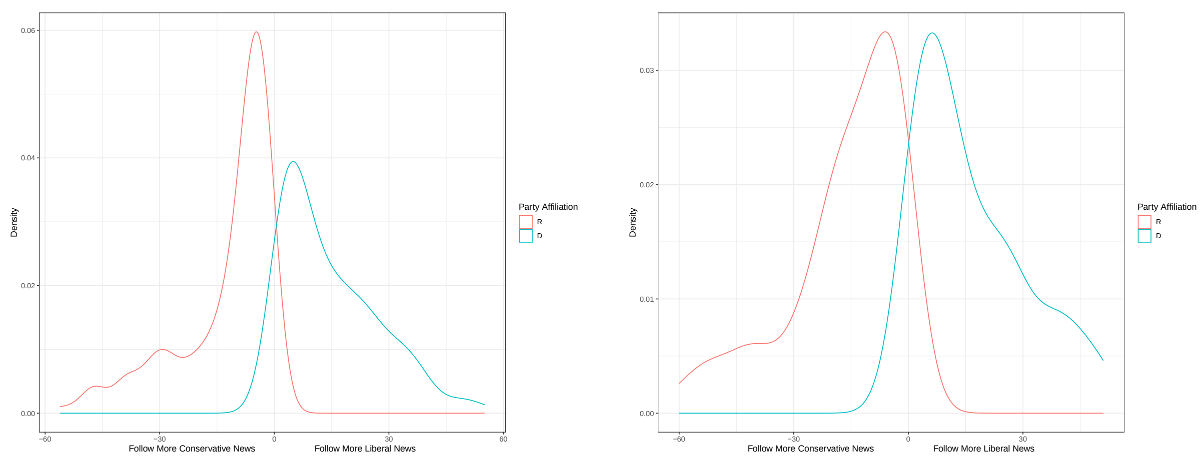
B Validation of Ideology Measure

B.1 Politician Social Media Accounts

To validate our measure of user ideology, we examine the relationship between the lists of ideological accounts we use to construct the measure and the accounts that partisan politicians follow on Twitter. We compile a list of 587 politicians including members of the 117th US Congress, cabinet officials, and some former Senators whose party affiliation is known. Among them, 336 official accounts and 128 political accounts have posted tweets about the Ukraine war during the time window in our data collection. Density plots in Figure A2 shows that politicians' party affiliations are closely associated with the types of news accounts they follow. Republican officials follow more conservative accounts while their democratic counterparts follow more liberal accounts.

Figure A2: Distribution of Differences between Types of News Accounts Followed by Politicians

(a) Politicians (Official Accounts) (b) Politicians (Political Accounts)



B.2 Survey Validation

In addition, we design a survey implemented on Amazon Turk, which received IRB approval from the Washington University in St. Louis IRB. The questionnaire is available in Section G. In the survey, we ask 725 subjects' political participation and 442 of them provided their Twitter usernames. Of these, 69 users at least follow one of the accounts in our lists of 75 conservative and 75 liberal political commentators, talk show hosts, and journalists. We are only able to validate the authenticity of 41 of the 69 users and so caution against overinterpreting this validation check. However, we include this in the Appendix for transparency.

We ask respondents whether they think of themselves as Democrats, Republicans, or Independents and construct a seven-point measure ranging from -3 to 3 of their self-reported partisanship attachment. We also construct a measure of their ideology based on the social networks they follow, as we do in the main manuscript. Despite the small sample size, we find that self-reported partisanship predicts ideology ($\beta_1 = 0.5938$, $SE = 0.3472$). We also ask respondents whether they voted for Joe Biden and Donald Trump in the 2020 election. We find a strong relationship between who respondents voted for and our measure of ideology ($\beta_1 = 5.197$, $SE = 2.371$).

C Validation of Country-Base of Twitter Users

Since our paper explores the ideological division of American social media users regarding their discussion of Russian invasion of Ukraine, we need to restrict the tweets included in our main analysis to be posted by U.S as much as possible. persons. However, there is no direct way to validate the origin of Twitter users as most of them do not report their countries or hide the locations of tweeting due to privacy issues. To indirectly inform the country base of users, we randomly select 10,000 tweets from the whole English corpus and create a variable indicating whether a tweet is included in our main analysis (in-sample tweets). The results in figure A3 show that these in-sample tweets have a higher proportion of topics related to American, Biden, Trump (Topic 2 and Topic 9), compared to out-sample tweets which are excluded in the main analysis, suggesting that our in-sample tweets are more likely to be posted by US persons.

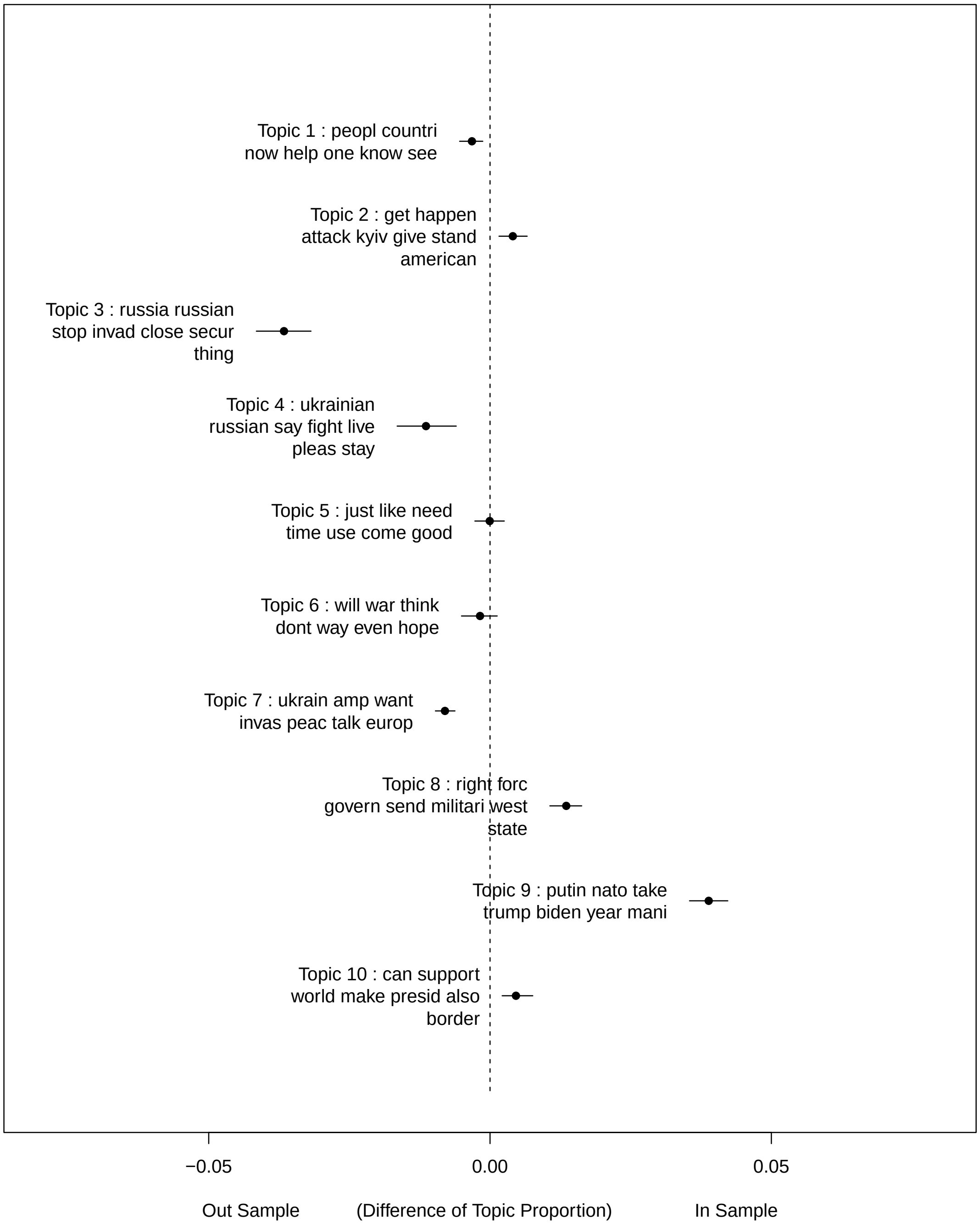


Figure A3: Topical Prevalence Contrast between In-Sample and Out-Sample Tweets

D Labeling Topics

We hired two research assistants who are majored in political science to manually code each cluster of tweets into the substantive topics we categorize from the corpus. We train them with a detailed coding protocol, which are provided in Section E. After receiving their work, we check their agreement rate, which is about 54%. This could be due to the higher number of categories and some ambiguity in the definition of each category. For example, some Tweets mention both Biden and Trump when discussing domestic politics. Therefore, we combine the three categories "Domestic Politics - Biden", "Domestic Politics - Trump", and "Domestic Politics" together. For those clusters of Tweets that two RA disagree with each other in choosing the substantive topic labels, the authors code again and adopt the majority rules.

E Coding Protocol

Background

Nearly two million Tweets regarding the Russian invasion of Ukraine were downloaded from Twitter and categorized according to the messages they conveyed. Under the assumption that Tweets with similar messages would use similar language, a computer algorithm was used to group together Tweets that had words or phrases in common. This process was generally successful in delineating Tweets according to subject matter, and resulted in the identification of over 6,000 clusters. For example, Tweets using words such as "lord," "amen," "Jesus," and "God" were grouped together, and upon further analysis, it was determined that such Tweets were generally sent by people expressing their sympathy and support for Ukraine and the Ukrainian people.

In the spreadsheet produced from the computer algorithm, each row of the spreadsheet contained one column with 50 common words or phrases found in the Tweets in the given cluster. Each row also contained 10 of the Tweets in the cluster to get a sense of the nature, tone, and more precise subject matter of the cluster. Using these two pieces of information – the 50 common words or phrases and 10 sample Tweets – the Tweets were then categorized. The goal was to narrow the large number of clusters into 10 subcategories.

Categories

Given the need to narrow down from over 6,000 clusters to a more manageable size, Tweets were categorized relatively broadly, according to their general subject matter and nature. Throughout the coding process, clusters were assigned to the following categories:

- Anti-Media: This category includes Tweets expressing anger regarding the media and press and the coverage of the Ukraine invasion. Though these Tweets are typically critical of the American "mainstream media," this category also includes Tweets that are critical of foreign outlets like the BBC and Al Jazeera.
- Domestic Politics - Trump: This category includes Tweets regarding former US President Donald Trump, his family, and his staff. These Tweets mention either actions taken by President Trump, or more frequently, Tweeters' opinions and thoughts regarding President Trump.
- Domestic Politics - Biden: This category includes Tweets regarding current US President Joseph Biden, his family, and his staff. These Tweets mention either actions taken by President Biden, or more frequently, Tweeters' opinions and thoughts regarding President Biden.
- Domestic Politics: This category includes Tweets regarding American domestic politics that did not reference either current US President Joseph Biden or former US President Donald Trump. Amongst other domestic policies, these Tweets included discussions regarding election integrity, vaccinations, inflation, and gun violence. This category also includes Tweets with references to other notable American politicians, such as Senators and Representatives. This category includes Tweets in which the Tweeters are asking their followers a question regarding and asking them to respond with their thoughts. This category includes Tweets regarding the economic impact of the Ukrainian crisis on the American economy.
- Foreign Policy: This category includes Tweets regarding countries other than the United States. Generally, such Tweets pertained to the foreign policy actions of these countries, although occasionally, such Tweets included references to domestic grievances in these countries, as well.
- Misinformation: This category includes Tweets that broadly have no basis in reality or truth and revolve around online conspiracy theories. This category includes Tweets that contain information that is false, or Tweets that are discussing information that they know to be untrue. Examples include, but are not limited to, discussions of "deep state," "mafia" involvement, "biolabs," the "Azoz batallion," "trafficking," "globalists," "cash

cow", as well as speculation about a "Ghost of Kyiv," a Ukrainian pilot who is said to have shot down as many as 40 Russian planes.

- News: This category includes Tweets that are purely newsworthy in nature and are absent of any opinion or speculation. This category includes Tweets regarding the military actions of the Ukrainian and/or Russian armies. Tweets in this category generally pertained to troop movements, bombardments, and other actions. Some of these include tweets with the hashtag #breaking.

- Random: This category includes Tweets which either (1) were similar in their content, but not remotely relevant to the Russia-Ukraine conflict, (2) dissimilar in their content to other Tweets that were analyzed, and thus do not fit together into any particular category, or (3) gibberish or foul language that is not relevant to the topic.

- Russia Discourse: This category includes Tweets discussing Russia, Russia's government, and its President Vladimir Putin. Generally, such Tweets are explicitly supportive of Russia's invasion of Ukraine. It may also include discussion of Luhansk and Donetsk, which many users claim is connected to Russia's motivation for invading. It would also include any mention of shelling of Donbass. This category includes Tweets in which the Tweeters perceive hypocritical behavior by the United States and the West in their actions pertaining to Ukraine. Generally, these Tweets were critical of the US and the West. It also includes mention of NATO enlargement, NATO expansion, NATO moving eastward.

- Ukraine Support: This category includes Tweets regarding Tweeter's support for Ukraine, the Ukrainian people, the Ukrainian government, its President Volodymyr Zelenskyy, and its military. It also includes any tweet in which the the user expresses a special family connection to Ukraine. Generally, these Tweets express the wish for peace in Ukraine. This category however also includes Tweets expressing anger towards the Russian government, the Russian military, or Russian President Vladimir Putin regarding the invasion of Ukraine. It should also include any mention of a no-fly-zone over Ukraine. Anything that includes the #StandwithUkraine hashtag.

Sample Tweets

- **Anti-Media**

1. The same #MSM that lied to you for two years about covid, is now asking you to believe them about Ukraine and Russia.

2. The media has lied to US about, Iraq Afghanistan Libya Syria Vietnam North Korea Iran Venezuela Guatemala Honduras Haiti Cuba Panama Nicaragua but somehow they're telling the truth about Ukraine? Carry on.

3. Can we really believe any of the mainstream news coming out of the Ukraine/Russian war?

4. Time to stand with America, Ukraine, and NATO. Time to cut off treasonous Fox News.

- **Domestic Politics**

1. Why is Ukraine's border more important than our southern border?

2. Notice how Ukrainian patriots are defending their capital instead of attacking it.

3. What pronouns do we use for the invasion of Ukraine?

4. The GOP is a Donald Trump-led violence cult. They have "unified messaging" and "never question the leader," like Hitler's Germany and all cults. They can't have human reactions to Ukraine or 1/6 or anything, it's all just hate of Dems and denial of their own violence and crime.

- **Domestic Politics - Biden**

1. I'm just so glad Biden is president and Trump lost. Remember when he was going to withhold military equipment in a quid-pro-quo from Ukraine for dirt against the Bidens? Trump would have been too confused for this... Hell, he thinks we're there now in Ukraine (Amphibian).

2. Hey, remember when our sitting VP, who oversaw foreign policy in Ukraine, & then his son Hunter was appointed to the Board of Directors for Burisma? Hunter had ZERO energy or international business experience. Board of directors 5yrs for a multi-billion dollar company. Sure.

3. I "still cannot believe" Biden has caused this invasion of Ukraine through his fecklessness and incompetence! People are going to die and their blood is on Biden's hands

4. Biden is reason for Russia invading Ukraine. Weak leadership, failed Afghan withdrawal, war on fossil fuels in US while approving Russian pipelines! Putin takes advantage of weak leadership! Invaded Georgia/Crimea under Obama, zero consequences, did zero under Trump, now Ukraine

- **Domestic Politics - Trump**

1. If Trump was still president Russia wouldn't have invaded Ukraine

2. Trump would have already handed over Ukraine to his buddy Putin!

3. Trump wanted out of NATO, he wanted reduced forces in Europe, he met in private with Putin, he took translator records of the meeting, he accepted a soccerball microchipped by Putin, he denied military aid to Ukraine, he took Putin's side over US Intel agencies

4. A reminder that Trump withheld $ 400 million in military aid to Ukraine just 91 minutes after the "perfect phone call" where he asked Zelenskyy to "do us a favor." And not a single Republican in the House and just 1 in the Senate (Mitt Romney) held him accountable in impeachment.

- **Foreign Policy**

1. EU.USA! Russian military will destroy Ukraine's internet and communication networks EU and the US should provide free internet and satellite communication services to Ukrainians Real-time Russian military murder through YouTube. The whole world should be able to see it in real

2. Earlier today, the Islamic Emirate of Afghanistan (the Taliban) called for peace in Ukraine and restraint by both parties. They ask "for both parties to resolve the crisis through dialogue and peaceful means"

3. I'm not worried about where he'll go next I'm worried about Ukraine and Putin is exactly the same as Osama Bin Laden and we need to take him out just like we did Osama bin Laden. Full stop

4. Its a good comparison both Russian and Turkish military are invaders Turkey is an invader in Iraq, Syria, and Cyprus.while Russia is an invader in Ukraine! No difference!

- **Misinformation**

1. The whole thing is a huge Psy-op. The ghost of Kiev? Come on people. Really

2. I stand with Ukraine people NOT the Ukraine Deep State Cabal!

3. To all the Putin bashers, FYI: Vlad is cleaning house in the Ukraine. Getting rid of: the Khazarian maffia, Soros Orgs and outposts, Bio Weapon Labs (7 out of 11 already destroyed), the Clinton Money Laundering Machine and more... #Ukraine #Ukraina #WEF #GreatReset #Putin

4. This is about WAY more than Russia and Ukraine. This is about taking out the money behind the elites/cabal/DS Governments that have lied to us for 100 years. This is the start of the fall of the DS / Cabal.

- **News**

1. #FIFA and #UEFA suspended #Russian national teams and clubs from all competitions. (UEFA) #Ukraine

2. #BREAKING UN votes to hold emergency General Assembly session on Ukraine. UNSC meet on holding UNGA session on Ukraine Voting: Yes 11 Against 1 (Russia) Abstentions 3 (India, China, UAE) The resolution is adopted to hold emergency #UNGA session.

3. #BREAKING: Russia ready to talk if Ukraine army 'lays down arms': foreign minister Lavrov –

4. Ursula Von Der Leyen, President of the European Commission, tells Euronews that the EU wants Ukraine in the bloc and that "they're one of us." #RussiaUkraineConflict

- **Random**

1. Now everyone is a Ukraine expert.

2. This made me laugh really hard.

3. Get lost Russian bot

4. WTF are you talking about? It's Ukraine being invaded.

- **Russia Discourse**

1. Ukraine is a proxy of US and NATO And they are radicalizing Ukrainian to fight with Russian

2. Russia is pretty clearly trying to avoid civilian casualties as much as possible, while Ukraine is going full Endsieg

3. They are trying to liberate Ukraine from freedom.

4. Ukrainian government is corrupt. I think Putin knows n exposing them

- **Ukraine Support**

1. "No one is as brave as the people of Ukraine. The bravery I'm seeing is outstanding. Just outstanding"

2. #IStandWithUkraine My family, friends, and colleagues stand with Ukraine.

3. Give Ukraine all the Stingers, Javelins and Tows. Give Ukraine all the Stingers, Javelins and Tows. Give Ukraine all the Stingers, Javelins and Tows. Give Ukraine all the Stingers, Javelins and Tows. Give Ukraine all the Stingers, Javelins and Tows.

4. President Zelensky will be remembered in history as a true hero together with the brave Ukrainian people.

F Validation of Topic Valence

One important hypothesis in our paper is that liberals are more likely than conservatives to express support for Ukraine. To ensure that the model correctly captures the sentiment of pro-Ukraine and groups them together, we randomly sample 500 tweets from those labeled as "Ukraine Support". Then, we assigned a RA to manually check if each tweet actually shows support for Ukraine. The RA could code each tweet as "Yes", "No", and "Unclear". The coding results are provided in Table A1. According to these statistics, our model achieves 84% precision rate. Precision can be seen as a measure of quality, and higher precision means that an algorithm returns more relevant results than irrelevant ones.

Table A1: Validation of Pro-Ukraine Attitude

Coding	No. of tweets	Metric
Yes	420 (84%)	True Positive
No	46 (9.2%)	False Positive
Unclear	34 (6.8%)	False Positive
Total	500	

G Survey Questionnaire

INTRODUCTION:

Thank you for agreeing to take this brief questionnaire about social media and politics. We will ask you some questions about who you are and your general political ideology.

To begin with, what is your Twitter username?

____

Section I: Demographics

1. How old are you?

____

2. What gender do you identify as?

a. Male

b. Female

c. Other __________

d. Prefer not to answer

3. What is your race or ethnicity (select all that apply)?

a. White

b. Black or African American

c. American Indian or Alaska Native

d. Asian

e. Native Hawaiian or Other Pacific Islander

f. Some Other Race

g. Prefer not to say

4. What is the highest degree or level of education you have completed?"

a. Some High School

b. High School

c. Bachelor's Degree

d. Master's Degree

e. Ph.D. or higher

f. Trade School

g. Prefer not to say

Section II: Political Activity

5. Generally speaking, do you usually think of yourself as a Republican, a Democrat, Independent, or something else?

[if Democrat:]

6. Would you say you are a strong Democrat or not very strong Democrat?

a. Strong

b. Not very strong

[if Republican:]

7. Would you say you are a strong Republican or not very strong Republican?

a. Strong

b. Not very strong

[if Independent:]

8. Do you think of yourself as more similar to Republicans than Democrats, more similar to Democrats than Republicans, or equally similar to Republicans and Democrats?

a. More similar to Republicans

b. More similar to Democrats

c. Equally similar to Republicans and Democrats

9. Did you vote in the 2020 Presidential election?

a. Yes

b. No

c. Prefer not to answer

[if no, proceed to 13]

10. Who did you vote for?

a. Joe Biden

b. Donald Trump

c. Someone else

11. Did you donate to any Presidential campaign or political organization in the 2020 election?

a. Yes

b. No

[if no, proceed to 15]

12. Which Presidential candidate did the campaign or organization to which you donated support?

a. Joe Biden

b. Donald Trump

c. Someone else

13. Did you volunteer for any Presidential campaign or political organization in the 2020 election?

a. Yes

b. No

[if no, end survey]

14. Which Presidential candidate did the campaign or organization with which you volunteered support?

a. Joe Biden

b. Donald Trump

c. Someone else

Thank you for your participation in the survey.

[Survey finished, respondents receive information about collecting payment]

References

1. M. McPherson, L. Smith-Lovin, J. M. Cook, Birds of a feather: Homophily in social networks. *Annual review of sociology* **27**, 415 (2001). _eprint: https://www.annualreviews.org/doi/pdf/10.1146/annurev.soc.27.1.415.

2. M. Mosleh, C. Martel, D. Eckles, D. G. Rand, Shared partisanship dramatically increases social tie formation in a Twitter field experiment. *Proceedings of the National Academy of Sciences* **118**, e2022761118 (2021).

3. D. Demszky, others, Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *arXiv:1904.01596* (2019). _eprint: https://arxiv.org/pdf/1904.01596.pdf.

4. D. Angelov, Top2vec: Distributed representations of topics. *arXiv:2008.09470* (2020). _eprint: https://arxiv.org/pdf/2008.09470.pdf.

5. G. Lin, W. G. Nomikos, New Methods and Procedures for Measuring Public Opinion with Social Media Data. *SocArXiv* .

6. D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation. *Journal of machine Learning research* **3**, 993 (2003). _eprint: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

7. Q. Le, M. Tomas, Distributed representations of sentences and documents. *International conference on machine learning, PMLR* **32**, 1188 (2014). _eprint: http://proceedings.mlr.press/v32/le14.pdf.

8. L. McInnes, J. Healy, J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv:1802.03426* (2018).

9. L. McInnes, J. Healy, S. Astels, hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software* **2**, 205 (2017).