

The World Social Media Survey (WSMS)

Memo Prepared for Stanford Global Populisms Conference

Valeriya Mechkova (University of Gothenburg)
Daniel Pemstein (North Dakota State University)
Brigitte Seim (University of North Carolina, Chapel Hill)
Steven Wilson (University of Nevada, Reno)

Introduction

This memo gives a high-level overview of the World Social Media Survey – a project led by the co-PIs, in collaboration with, and with initial funding from Facebook. The project uses the infrastructure of the V-Dem Institute to collect data for the initial World Social Media Survey in January of 2019 (to capture data from 2000 to 2018) with an update in January of 2020 (to capture 2019 data). The survey entails a set of ~35 new indicators of polarization and politicization of social media, misinformation campaigns and coordinated information operations, and foreign influence in and monitoring of domestic politics, measured at a country-year level across 180 countries. The V-Dem data team will process this survey using the standard V-Dem measurement modeling and quality control processes, using the V-Dem Institute infrastructure for collection and aggregating expert-based data on democracy, reaching out to a network of over 3,200 scholars from more than 180 countries. The full data set will be made publicly available in the spring of 2019.

V-Dem and Expert Coding

V-Dem has built up a world-leading research infrastructure for collection and aggregating expert-based data on democracy, and is the largest data provider of its kind. The existing Varieties of Democracy (V-Dem) infrastructure consists of a custom-built relational database, a set of web-based survey tools and administrative interfaces, a public website including online analysis tools, with a second layer of custom-designed intranet interfaces for administration and management, a network of over 3,200 scholars from more than 180 countries, and six staff at the V-Dem Institute with unique competencies. Together, this infrastructure allows V-Dem to produce 350+ unique disaggregated indicators, five distinct indices of democracy, and 47 indices measuring various components of democracy and governance. After quality control, and aggregation, V-Dem makes 18mn data points publicly available for download and online analysis. Today, the data set covers 178 countries from 1900 to 2017.

V-Dem has developed innovative methods for aggregating expert judgments in a way that produces valid and reliable estimates of difficult-to-observe concepts. For example, while it may be easy to code whether a country is holding elections, evaluating whether there was some kind of electoral misconduct is a difficult task. Therefore, V-Dem's data collection processes are designed to pay close attention to potential sources of error. Five independent country experts code all evaluative

indicators. These experts are typically academics originally from or with extensive experience in the country they are coding. We have gathered the knowledge of over 3,000 experts in total, allowing for inter-coder reliability tests and additional tests that allow for the detection of systematic biases.

These biases can stem from the fact that judgments and thresholds may vary significantly across experts and cases, especially when they come from different contexts and backgrounds. Moreover, because even equally knowledgeable experts may disagree, it is imperative to capture and report potential measurement error. We address these issues using both cutting-edge theory and methods.

Pemstein (one of the co-PIs on the WSMS) et al. (2018) have developed a custom-made Bayesian Item-Response Theory (IRT) measurement model. The IRT models we use allow for the possibility that experts have different thresholds for their ratings and perceptions for what concepts mean. We have encouraged country experts to code other countries and cases than their original case, allowing us to compare individual experts' thresholds, and facilitating for cross-country comparability. We also employ anchoring vignettes to further improve the estimates of expert-level parameters. Anchoring vignettes are descriptions of hypothetical cases that provide all the necessary information to answer a given question, providing a great deal of information about how individual experts understand question scales.

Finally, the model aggregates each set of ratings for a particular country/indicator/year, taking coder confidence (an expression of each coder's subjective sense of uncertainty) into account. For each point estimate we also calculate an overall certainty statistic that reflects expert disagreement.

As recognition for its high quality, V-Dem received the most prestigious award for comparative datasets in political science: the Lijphart/Przeworski/Verba Best Dataset Award presented from the American Political Science Association's Comparative Politics Section in 2016.¹

Implementation of the World Social Media Survey

The WSMS is a newly designed expert-coded survey, built to capture thirty-five new indicators measuring the politicization of social media, misinformation campaigns and coordinated information operations, and foreign influence in and monitoring of domestic politics via the Internet. Other than a handful of multiple selection and free-response style questions, all questions were presented with Likert scales, included a full set of anchoring vignettes, and were offered in six languages (English, in addition to French, Spanish, Portuguese, Arabic, and Russian).

¹ V-Dem infrastructure, data collection, research, collaboration and outreach is/has been funded by a collection of research foundations and international sources including the European Commission/DEVCO, the World Bank, Ministry of Foreign Affairs, Sweden, Danish International Development Agency, Canadian International Development Agency; the Research Council of Norway/NORAD, the Mo Ibrahim Foundation, the B-Team, International IDEA, The European Research Council, the Research Councils of Sweden, Norway, and Denmark, the Riksbankens Jubileumsfond, and the M&M Wallenberg and the K&A Wallenberg foundations. Co-funding has been provided by the Vice Chancellor, the Dean of Social Sciences, and Department of Political Science at University of Gothenburg (UGOT).

Initial data collection concluded in January of 2019 for 180 countries from 2000 to 2018, and the V-Dem data team is currently processing using the standard V-Dem measurement modeling and quality control processes. Funding has already been secured for an additional round of coding in January of 2019. The indicators cover five sub-domains, detailed below, along with the question text for each question (though not including the full set of Likert scale choices for each, given space constraints). Most questions were asked with a descriptive five-item Likert scale.

Coordinated Information Operations

Social media is increasingly used as a tool of coordinated information operations. These operations can be used by either foreign powers with a vested interest in the political trajectory of the country, or by domestic actors with an incentive to skew information available to the public. These actors use the reach of social media and tools such as “troll armies” to generate and disseminate particular viewpoints or fake news. This portion of the survey captures the involvement of foreign actors in domestic politics via Internet technologies, and the presence and characteristics of either foreign or domestic coordinated information operations. In addition, it captures the capacity of regimes for using such techniques both domestically and abroad.

Indicator	Question Text
Government dissemination of false information domestic	How often do the government and its agents use social media to disseminate misleading viewpoints or false information to influence its own population?
Government dissemination of false information abroad	How often do the government and its agents use social media to disseminate misleading viewpoints or false information to influence citizens of other countries abroad?
Party dissemination of false information domestic	How often do major political parties and candidates for office use social media to disseminate misleading viewpoints or false information to influence their own population?
Party dissemination of false information abroad	How often do major political parties and candidates for office use social media to disseminate misleading viewpoints or false information to influence citizens of other countries abroad?
Foreign governments dissemination of false information	How routinely do foreign governments and their agents use social media to disseminate misleading viewpoints or false information to influence domestic politics in this country?
Foreign governments ads	How routinely do foreign governments and their agents use paid advertisements on social media in order to disseminate misleading viewpoints or false information to influence domestic politics in this country?

Digital Media Freedom

These questions will mode (e.g., filtering, active takedowns, limitation of access), actor (e.g., government, non-state actors), and extent of censorship.

Indicator	Question Text
Government Internet filtering capacity	Independent of whether it actually does so in practice, does the government have the technical capacity to censor information (text, audio, images, or video) on the Internet by filtering (blocking access to certain websites) if it decided to?
Government Internet filtering in practice	How frequently does the government censor political information (text, audio, images, or video) on the Internet by filtering (blocking access to certain websites)?

Government Internet shut down capacity	Independent of whether it actually does so in practice, does the government have the technical capacity to actively shut down domestic access to the Internet if it decided to?
Government Internet shut down in practice	How often does the government shut down domestic access to the Internet?
Government social media shut down in practice	How often does the government shut down access to social media platforms?
Government social media alternatives	How prevalent is the usage of social media platforms that are wholly controlled by either the government or its agents in this country?
Government social media monitoring	How comprehensive is the surveillance of political content in social media by the government or its agents?
Government social media censorship in practice	To what degree does the government censor political content (i.e., deleting or filtering specific posts for political reasons) on social media in practice?
Government cyber security capacity	Does the government have sufficiently technologically skilled staff and resources to mitigate harm from cyber-security threats?
Political parties cyber security capacity	Do the major political parties have sufficiently technologically skilled staff and resources to mitigate harm from cyber security threats?

Online Media Polarization

This portion of the survey provides indicators of the level of polarization in discourse in both online and traditional media, probing the extent to which media environments are fractionalized, the extent to which citizens obtain political information from polarized sources, and the extent to which media markets serve particular ideological niches.

Indicator	Question Text
Online media existence	Do people consume domestic online media?
Online media perspectives	Do the major domestic online media outlets represent a wide range of political perspectives?
Online media fractionalization	Do the major domestic online media outlets give a similar presentation of major (political) news?

Social Cleavages

This portion of survey examines the extent to which social cleavages proliferate, are activated, and engender ongoing conflict within states. This exploration includes several questions specific to online, social media discourse, as well as more indirectly related measures of cleaves in society more generally.

Indicator	Question Text
Online harassment of groups	Which groups are targets of hate speech or harassment in online media? (Multiple selection of 10 groups, with free-text entry for other)
Use of social media to organize offline violence	How often do people use social media to organize offline violence?
Average people's use of social media to organize offline action	How often do average people use social media to organize offline political action of any kind?
Elites' use of social media to organize offline action	How often do domestic elites use social media to organize offline political action of any kind?
Party/candidate use of social media in campaigns	To what extent do major political parties and candidates use social media during electoral campaigns to communicate with constituents?
Arrests for political content	If a citizen posts political content online that would run counter to the government and its policies, what is the likelihood that citizen is arrested?

Types of organization through social media	What types of offline political action is most commonly mobilized on social media? (Multiple section of 9 actions, with free-text entry for other)
Polarization of society	How would you characterize the differences of opinions on major political issues in this society?
Political parties hate speech	How often do major political parties use hate speech as part of their rhetoric?

State Internet Regulation Capacity and Approach

States vary dramatically in their capacity to regulate online content. This portion of the survey examines the extent to which the state has the capacity to regulate online content, and the model that the state uses to regulate online content. In particular we ask questions about the extent to which laws allow states to remove content, privacy and data protections provided by law, the extent to which actors can leverage copyright and defamation law to force the removal of online content, and de-facto levels of state intervention in online media.

Indicator	Question Text
Internet legal regulation content	What type of content is covered in the legal framework to regulate Internet?
Privacy protection by law exists	Does a legal framework to protect Internet users' privacy and their data exist?
Privacy protection by law content	What does the legal framework to protect Internet users' privacy and their data stipulate?
Government capacity to regulate online content	Does the government have sufficient staff and resources to regulate Internet content in accordance with existing law?
Government online content regulation approach	Does the government use its own resources and institutions to monitor and regulate online content or does it distribute this regulatory burden to private actors such as Internet service providers?
Defamation protection	Does the legal framework provide protection against defamatory online content, or hate speech?
Abuse of defamation and copyright law by elites	To what extent do elites abuse the legal system (e.g., defamation and copyright law) to censor political speech online?

Preliminary Findings

The full measurement model is still being run to produce the final set of indicators for public release, but we do have some preliminary findings worth reviewing, both in terms of how social media is used for offline mobilization and in terms of information operations by regimes.

Social Media and Mobilization

First, indicators show that with the rise of Internet usage around the world, the new communications platforms have been colonized by offline hatred. We measure this in part with the multiple select question "Which groups are targets of hate speech or harassment in online media?". The most common specific targets are LGBTQ groups and individuals (in 76% of countries), followed by specific ethnic groups (66%), specific religious groups (58%), and women (51%). In only twelve countries does the expert consensus hold that no specific groups are targeted by hate speech or harassment online, mostly countries with among the lowest Internet penetration rates in the world.

In addition, multiple indicators capture dimensions of how the Internet and social media are being used to solve the collective action problem, both for good and for ill. In “What types of offline political action are most commonly mobilized on social media?” we find that online organization seems to be extremely widespread, with the most common offline political action being street protest (organized online in 92% of countries), petition signing (90%), voter turnout efforts (86%), and strikes/labor actions (65%). There are definite geographic patterns in this data. For instance, Guyana is the *only* country in all of Latin America that doesn’t report labor actions organized online.

Even the small numbers of countries reporting no online organization in particular categories do not entirely overlap, as each country tends to have at least some categories of political action organized online. The only exception, perhaps as expected, is North Korea.

Significant violent action is also mobilized online in many contexts. Terrorism is organized online in 10% of countries, and vigilantism in 10% as well. Interestingly, only a third of those cases overlap, indicating that different varieties of violence are organized in different country contexts. In addition, the use of social media in organizing ethnic cleansing or genocide is reported in five countries: Burma/Myanmar, India, Iraq, Saudi Arabia, and South Sudan.

The indicator “how often do people use social media to organize offline violence?” provides some additional perspective, on a three-point Likert scale. Only seven countries in 2018 ranked as frequently with “numerous cases in which people have used social media to organize offline violence”: Iraq, the DRC, Hong Kong, Libya, Maldives, Bangladesh, Syria, and Kosovo. 122 countries ranked in the “sometimes” category, while the remaining countries in which violence organized online was considered rare were split between a handful of highly consolidated democracies with little violence in the first place, wealthy and highly capable autocracies, and countries with low levels of Internet access.

In addition, we distinguish which segment of the population is organizing “offline political action of any kind” with social media in a pair of indicators that separately capture whether average people and domestic elites are doing so. It is telling that in 70 countries “average people” often or frequently do, but only in 37 countries do domestic elites. This distinction at least in part reflects social media’s capacity as a classic “weapon of the weak”, something that can disproportionately empower those with the least power.

On the other hand, it is intriguing to pull out the country cases in which the opposite seems to be the case: that is, countries in which the indicator for elite usage to mobilize significantly exceeds that of “average” person usage. The most significant difference is in the Philippines, reflecting Duterte’s extensive use of social media for populist purposes.

Coordinated Information Operations

Next, we can examine the extent to which information operations using online media are present in countries around the world, both by domestic and foreign actors. One concern about this set of questions was whether domestic online media was widespread enough in most countries to

represent a concern. That is, in less technically inclined countries, it was plausible that the content of online media, in particular social media, was simply a foreign import.

As such, we included an indicator for measuring the extent to which domestically sourced online media is consumed in each country (i.e. “Do people consume domestic online media?”). Not a single country ended up in the “no domestic online media exists” category, and only 17 are in the “limited consumption” category. The rest of the world is either relatively extensive (which is described as “domestic online media consumption is common”) or extensive. Since domestic online media consumption is nearly universally high, questions about the distribution of false information in that sphere are particularly salient.

We have several indicators that map the degree to which false information operations act on social media in each country. First, we ask the degree to which “the government and its agents use social media to disseminate misleading viewpoints or false information to influence its own population.” The worst offenders on this account are Azerbaijan, North Korea, Cuba, Tajikistan, Bahrain, Yemen, Syria, Venezuela, Russia, and Iran.

Second, we also measure the degree to which governments use social media to spread false information to “to influence citizens of other countries abroad.” These are mostly the same offenders as the previously described indicator, with some notable exceptions such as China cracking the top ten.

Finally, we also ask “how routinely do foreign governments and their agents use social media to disseminate misleading viewpoints or false information to influence domestic politics in this country?” Only three countries qualify for the “worst” category in which such action is described as “extremely often” and occurring with regard to “all key political issues”: Bahrain, Taiwan, and Latvia. On the other hand, only 25 countries manage the highest rating, which maps to “never or almost never.” For comparison, the United States, targeted by well-documented Russian information operations, ranks 17th worst in the world by this metric. And tellingly, of the 20 worst countries, eight are former Soviet Bloc or member states.

Conclusion

We are only just getting started processing the enormous amount of data that we collected, but hope to have official time series data ready for public release with a couple of months across all 35 indicators. In addition, the three free-response text questions yielded some 50,000 words of text from experts (~150 pages single spaced) which should yield a trove of additional qualitative information.