



Social Media Analytics for Digital Advocacy Campaigns: Five Common Challenges

Discussion Paper

April 2016

ACKNOWLEDGMENTS

We would like to recognize the able assistance of Luping Wang in researching and writing of this paper. We are grateful to Matthew Leveque, Kjerstin Thorson and Aimei Yang of the USC Annenberg School for their insight and guidance. Thanks to Lauren Lee White, Michael Arda Karakash, Erica McNamara and Sohaela Amiri for their invaluable support. We would also like to offer special thanks to the North America Advocacy division of Global Affairs Canada, for their contributions to the framing and review of this project. This work was made possible by financial support from Global Affairs Canada.

INTRODUCTION

There has been extensive discussion, among both the practitioner and scholar communities, on public diplomacy in an increasingly digitized environment. Researchers have found that digital technologies like social media and big data have not only reinforced and expanded public diplomacy discourses,¹ but also become a driving force in change management in the structure and process of diplomatic work.² There is broad consensus that digital communication is central to a nation's efforts to engage with foreign publics, and that the general goal of such public diplomacy endeavors is to bring about actions performed/decisions made by foreign actors which are favorable to the national interest.

Digital technologies generate an unprecedented amount of data, allowing us to examine communication efforts in closer detail and to understand the extent to which they have an impact on the stated goals. For the most part, we know whom we engage (outside of general awareness campaigns, which rely heavily on non-targeted broadcasting), the number of times we engage, and what messages we deliver within the context of those engagements (outputs). We know whether an issue was resolved in the national interest (ultimate outcome), or not. It is the intervening period with which we struggle in terms of measurement. In other words, how can we bridge the measurement gap between advocacy operations (outputs) and ultimate outcomes?

Indeed, this measurement challenge is not unique to public diplomacy. It is understandable that a direct linkage between public diplomacy engagement with an actor and that actor's decision on a given issue is virtually impossible to demonstrate, given the multiple variables and the interaction of the variables that could have played a role in producing the actor's decision. To date, we have been operating on the premise that certainty in such a complex system, where human action stands at the intersection of technology, policy and politics, is out of reach; and as such, the most that can be aspired to is exploitation of any and all data which can serve to bolster our confidence that our advocacy operations are contributing to success in terms of ultimate outcome (or even immediate and intermediate outcomes).

How can we bridge
the measurement
gap between
advocacy operations
(outputs) and ultimate
outcomes?

The key question then becomes: what means and measures offer us the most efficient and effective means of increasing our confidence that our activities contributed to success or failure, with respect to immediate, intermediate and ultimate outcomes? This discussion paper seeks to reflect on and refine current practices in tracking and analyzing digital advocacy campaigns. We aim to advance a more rigorous, structured approach to social media analytics and their role in strategic planning.

To that end, we focus on the following five key areas in an attempt to gain increased confidence:

1. Search Parameters: Information and understanding derived from enterprise-level social media analytics tools are only as useful as the comprehensiveness of the search parameters used. The more precise we can be, the more confident we can be in the resulting analytics. However, with added precision comes unwieldy and time-consuming search strings. What approaches might we take to ensure the most precise search strings without unreasonable levels of time and effort?

2. Social Media Share of Voice: Defining share of voice as the ratio of posts attributed to our own accounts within total volume of conversation/posts retrieved based on a search string minus, what is the value of an increase or decrease over time? One could argue that an increase in share of voice demonstrates more control over a conversation, but is that desirable? Shouldn't the goal over time be to have a shrinking share of voice, with volume, positive sentiment and evidence of our messaging increasing?

3. The Who: Reams of quantitative data of all types can be collected. However, it is the qualitative "who" that matters. One million posts can be returned as a result of a search parameter, but the only posts that truly matter for our purposes are those associated with influencers/decision makers in relation to our advocacy priorities. Aside from manual analysis of every single hit and categorization as to influencer/non-influencer, how can we more easily separate the wheat from the chaff?

4. Sentiment Analysis: Current sentiment/natural language processing algorithms are often unreliable, leading to inaccurate aggregate statistics for search parameters. Manual survey sampling of the total volume of results is suggested to develop a more accurate assessment of sentiment. However, large search result pools (e.g., 1 million) are common—what is the minimum sampling size that would be necessary (least effort) to gain a reliable assessment?

5. Demographics: Demographic data are only available for those users that self-identify (e.g., gender, location, age, etc.), leading to unreliable aggregate statistics for search parameters. How can we increase the accuracy of demographic identification without the need for manual investigations of every hit?

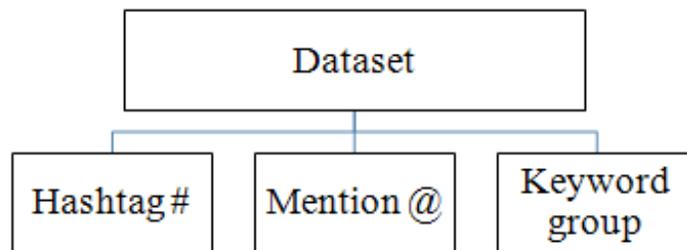
To help us explore these questions, we focus on digital advocacy campaigns and use the social media platform Twitter as a case in point. With 320 million monthly active users (as of December 2015), Twitter is one of the most popular digital content platforms. In this discussion paper, "advocacy" refers specifically to policy advocacy initiated by governments. A "campaign" is defined as a planned engagement program for a specific duration of time in a defined geographic region. In this context, "digital advocacy campaign" refers to an engagement effort initiated and implemented by a government targeting foreign publics on a specific policy matter through Twitter. In addressing the above questions, we draw on both academic research on related topics and our practical experience in analyzing big data sets.

Currently, there are multiple free, enterprise-level social media analytical software programs.

Popular free tools include HootSuite, Tweroid and Simply Measured. These free tools can provide basic functions, such as trend tracking, keyword search, and audience analysis. However, they do not provide customized filters for further analysis, which is a must for digital advocacy research. The discussion in this paper is mostly based on the paid tools Radian 6 and Crimson Hexagon. These two software programs allow the user to set up a personalized profile and conduct content analysis of a given campaign. Users also have full access to Twitter content and can export data for further coding. Moreover, paid tools are not limited to 3,200 historical tweets, which is a big disadvantage of free APIs (application program interface).

1. SEARCH PARAMETERS

To ensure accurate search results, we need to design precise search parameters on our chosen topic. Free or paid social media analytical tools are used for downloading the data. Before opening the analytical software, a three-dimensional dataset research parameter is recommended:



The big dataset is divided into three different groups: hashtags, mentions, and keyword groups. Hashtags are effective when generating group actions because they can gather people with the same interests. As Twitter has become an international social network, its important role in political conversation has also increased. For instance, hashtags have been used by scholars when researching activism.³ A universal hashtag can bring people with the same political demands together, thus increasing overall impact. Hashtags have been effectively used in social movements, such as the Arab Spring and the Occupy Movement. However, some scholars think that hashtag activism is a form of “slacktivism,” because it is about what people say instead of what people do. Slacktivism is often used to describe online actions that aim for political or social changes but with minimal efforts to engage, for instance, signing an online petition or joining in a social media discussion.

A mention is a tweet that contains another user’s Twitter name, or “handle,” preceded by the “@” symbol. Replying to a tweet is also considered a mention. Twitter collects all these messages and shows users in the notifications tab. Users can also include multiple usernames in one tweet. Specifically, Twitter mentions come in three categories: a) a mention of someone you follow, b) a mention of someone who is following you, and c) a mention of someone without any follower/followed relationship. Scholars are currently using big data approaches to figure out how the mention function accelerates the progress of online activism, for instance.⁴ In social movement research, key mentioned usernames include the

initiator of a movement, celebrities involved, most retweeted usernames, influencers, etc. In the context of this discussion, to ensure the wholeness of a research dataset, we need to conduct background research to identify key mentioned usernames before setting up search parameters.

When people do not use hashtags or mention others' names, they tend to have keywords or keyword groups in the tweet content, which can be a single word or a phrase. The keyword function can be used as a filter to select the content closely related to the research. For example, during COP 21 (2015 United Nations Climate Change Conference) in Paris, when people couldn't find a relevant hashtag or key influencer immediately, they would use the keyword "climate" or the keyword group "climate change" in their tweets. Keyword group searches are commonly used in academic research to complete datasets for further analysis.⁵

Some scholars think that hashtag activism is a form of "slacktivism," because it is about what people say instead of what people do.

Generally speaking, when we set campaign objectives, they need to be clear and specific, i.e., quantifiable, within a time frame, for a given geographic region. From the perspective of online monitoring, a defined goal also means a small range of detailed research stings, which will increase the accuracy of the research.

In Twitter analytics, before conducting large-scale data collection, we need to make sure we have collected major hashtags, mentions and keyword groups used in the conversation. These can be official hashtags promoted by key actors, or trending topics, which are frequently updated in a sidebar on Twitter. As the goal is to extract social media insights efficiently, we suggest using a preset keyword/topic. For example, if we decide to zoom in on climate change-related social media actions, the keyword "climate" could be used as the first filter. Then we look at top mentions, top hashtags and keyword groups in the climate-related conversation. With the help of social media analysis software, this process is relatively efficient.

We also need to be careful about setting the time range of the search parameters. It is not sufficient to collect data on the day of the campaign; data also must be collected on the days before and after to track possible changing trends. Take the examples of studying the Arab Spring movement and the People's Climate March. As the former was a social movement with a longer duration, researchers examined the Twitter data for several months to capture the big picture; while in the case of the People's Climate March, a one-day event in New York City, it was sufficient that researchers only looked at the tweets the day before, the day and the day after for analysis. In a word, the time range we choose for search parameters should be consistent with our research goal.

Current free Twitter monitoring tools on the market collect data by using the official Twitter API. However, these free tools have a limited number of tweets they can extract. Enterprise software is needed for any expansive social media analysis. Here is a comparison of three

such programs based on their pros and cons:

Software	Company information	Pros	Cons
Radian 6	A social media analytic tool from Marketing Cloud Suite, under the parent company Salesforce	- Easy to use - Twitter content download - Basic content analysis (most retweeted username, topic trend, etc.)	- Analysts need to apply more coding scripts to do further hashtag and network analysis.
Crimson Hexagon	A social media analytics company based out of Boston, Massachusetts	- Data visualization - Advanced sentiment analysis - Customized analysis tool	- Analysts cannot download Twitter data.
DiscoverText	Affiliated with Texifter LLC, a company targeting social data capture, historical Twitter access, and text analysis	- Multiple items such as hashtags, mentions, retweets in one spreadsheet. - Good for academic research.	- No data visualization. - Download limits to 15,000 per day - time and cost consuming.

Based on the comparison table above, we recommend a combined usage of Radian 6 and Crimson Hexagon. Radian 6 is useful to set up a topic profile and extract raw data from the Twitter database, while Crimson Hexagon is useful to analyze the content and trends of a given research topic. This mixed method gives us not only data visualization of how a topic develops over time, but also allows us to use a big data approach to conduct deep analysis on social networks and strategic communication.

2. SOCIAL MEDIA SHARE OF VOICE

Traditionally, share of voice is a concept associated with the advertising industry. The concept and practice are premised on the existence of competitors and competition. It calculates the percentage of a certain brand's advertising presence in the gross advertising market. For digital advocacy campaigns, it is critical to understand share of voice in social media through monitoring and analyzing the competition landscape.

The analysis of social media share of voice leads to the following sub-questions: Is there any competition concerning the campaign? Are people getting information from an official channel rather than other channels? How many comments are generated from followers? How many new followers were attained during the campaign? Identifying the power of controlled accounts is the first step in understanding the share of voice in a bigger conversation.

When it comes to social media, we suggest first designing search parameters in order to collect a complete dataset to research share of voice. We also suggest doing background

research to identify competitors/competition before the campaign begins, and monitoring their social media performance during the campaign period. Next, compare the number of impressions generated from our content (including retweets, comments and likes) to competitors' impressions.

If a digital advocacy campaign is effective, the number of tweets, the size of the conversation and the total impressions will increase when compared with those standards previously. The increasing number of tweets and total impressions also indicate a higher level of engagement. In the case of a conversation about climate change and clean energy, the competitor is not necessarily another organization, but rather comments and content contrary to the stated campaign objective. To maximize the impact of the campaign, it is crucial to evaluate the competition landscape. To do so, we suggest using sentiment analysis to evaluate the content of the online conversation.

To clarify and deconstruct the share of voice question, we identify three possible scenarios:

Scenario 1: *The conversation size increases, the percentage of positive content increases, and the campaign initiator's share shrinks.* Take the example of a government-led climate change initiative. As the purpose of the campaign would be to raise public awareness on the urgency of climate change, the share of voice by other parties should ideally be greater than the proportion of government accounts as a result of the campaign.

Scenario 2: *The conversation size increases, the percentage of positive content increases, but the government share also increases.* In this scenario, even though the positive share increases, the government and official channels remain dominant in the conversation. A large proportion of content coming from the official voice also means that there is not an expanding online community focusing on the issue.

Scenario 3: *The conversation size increases, the percentage of positive content decreases.* In that situation, the decrease in the share of voice of positive content indicates that the campaign is counterproductive. No matter whether the government's share goes up or down, the digital advocacy campaign hasn't worked.

There are more possible scenarios, however, these would seem to be the most common. It should also be noted that the desirability of each scenario is dictated by the objectives of the campaign, along with environmental context.

3. THE WHO

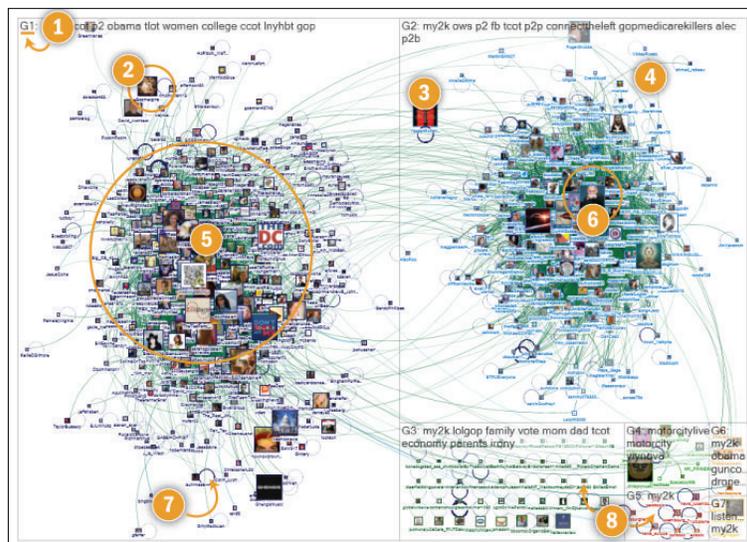
Let us suppose we have a dataset of 1 million tweets. Within these tweets are several networks (clusters). Each network has its own central actor and key topics. Locating "the who" means locating those central actors and understanding their interactions. Understanding how and why people go into any given interest group and the connections

between different clusters will contribute to knowledge of the relationships among “whos.”

In social science research, especially cases where there is large-scale collective action, it is important to categorize different actor types (by using customized labels) and understand their group attributes. In most cases, a codebook is thus designed to facilitate the understanding of the existing categories.⁶ There are several common labels, such as mainstream media organizations, mainstream new media organizations, non-media organizations, mainstream media employees, bloggers, activists, political actors, celebrities, researchers and others.⁷ Obviously, the codebook varies based on topic. The aforementioned general categories can also get more specific to ensure the quality of coding. In the research area of climate change, there is a preset codebook, which covers major labels (see Appendix).

The coding process requires one or more trained coders. There are various methods to test and ensure inter-coder reliability. Practically, in event-based social media management, it is more efficient to manually code top usernames and find out the primary influencer’s account in that conversation. Software coding is not recommended because of the complicated process of determining the actor type. Radian 6 can provide basic data, such as most retweeted usernames, most popular hashtags and most retweeted comments. It is also a good way to understand who is responsible for spreading information on Twitter. Categorizing actor types shows us the big picture of how the movement develops on social media.

Besides the data from Radian 6, we can use the network analysis tool NodeXL to identify key influencers in the conversation. We assume that the influencer is either the most-mentioned user or one who brings users together. NodeXL is a data visualization plug-in for Excel which can map such networks.⁸ After we identify key users and add customized labels to them, the next step is to examine the online network around them. The basic cluster function enables us to see how many clusters there are in the conversation and who the central actors are. We can also visualize how different clusters connect with each other. This is helpful in seeing how information flows from one topic to another (see the graphs below for illustration).



4. SENTIMENT ANALYSIS

Sentiment analysis of social media content aims to determine the content's attitude. Sentiment can be positive, negative or neutral. Current computational natural language processing offers decent accuracy in detecting sentiment, but still needs a lot of work to satisfy research requirements. What's more, manual sampling method becomes time-consuming when the dataset is large. One way to address the challenge is to sample big data for conducting sentiment analysis.

The first step is to use our given search parameters to collect the data. Then use the formula below to determine the sample size:

$$SS = \frac{Z^2 * (p) * (1-p)}{c^2}$$

Confidence level and confidence interval are two key factors in the calculation.⁹ If the confidence level is 95% and confidence interval is 5, and we have 100,000 tweets, according to the formula, we only need 383 sample tweets to do sentiment analysis, which will significantly reduce the workload.

After we determine the sample size, we create a sampling plan. There are three major methods for this:

1. Simple random sampling. This is the easiest way to do sampling. Analysts use the Excel function "RAND()" to generate a random sample from the data (username). Each username is chosen completely by chance and has an equal probability of being selected in the result. However, random sampling may create representation bias.
2. Stratified random sampling. A stratified random sampling separates the data into mutually exclusive sets by members' shared characteristics, and then draws simple random samples from each sets. It can provide greater precision than random sampling.
3. Cluster sampling. A cluster sample is a "natural" simple random sample of clusters of elements. Cluster sampling generally provides less precision than either simple random sampling or stratified sampling.

In the current context, we recommend using stratified random sampling as the main sampling method. Let's suppose that each month we collect a dataset of tweets. We would then separate the data by month and by ratio. If 300 is the sample size according to the

formula and in January, February and March, we have the number of tweets 100, 200 and 300 respectively, the ratio of sample of these three months will be 1:2:3 and the number of sampled tweets would be 50:100:150.

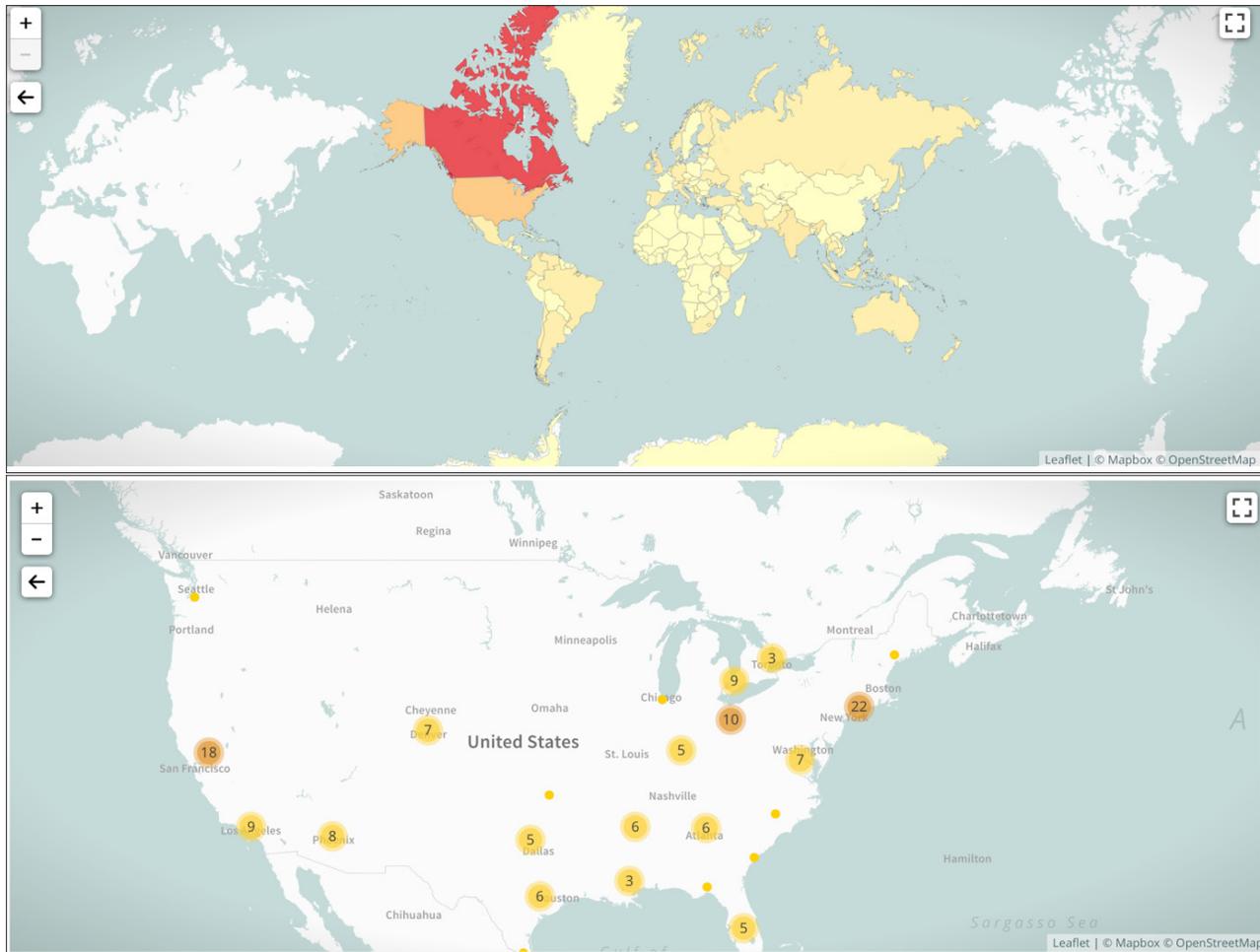
We now have the data sample for sentiment analysis. Typically, we would then do manual coding to determine sentiment. To ensure accuracy, a second coder is needed for inter-coder reliability. Sentiment is categorized as positive, negative or neutral. The advantage of manual coding is that a trained coder can determine the sentiment in the content more accurately given the contextual nature of language, though the process can be time-consuming. Current social media software can also do automatic sentiment analysis, which relies on a preset coding algorithm. Auto-Sentiment utilizes pre-defined sentiment categories like positive, neutral and negative. It uses a vast set of training tweets (over 500,000) that were human-coded.¹⁰ Then it uses this labelled content to calculate the frequency distribution of each key word present in those tweets across the positive, negative and neutral categories. These frequency distributions are used to construct a model that analyzes each new post and classifies it by sentiment. In the latest version of Crimson Hexagon, coders can further train the software by telling it the sentiment of the content. The more it gets trained, the more accurate the results become. This machine learning approach is especially helpful when there is a large sample to analyze.

5. DEMOGRAPHICS

Social media demographics, compared with traditional demographics research, are harder to collect because of the scale of online conversation. Demographics include, but are not limited to, geolocation, gender, age, race, education level, etc. In this section, we consider the demographics of key influencers, as well as other participants in a given conversation on social media. How do we determine these demographics? Currently, analysts rely mainly on self-identified user data. Some also use surveys to extrapolate demographic information.

Geolocation is the easiest demographics data to capture. Radian 6 and Crimson Hexagon provide the IP address of each tweet, but analyzing these data is labor intensive. To facilitate the process, Twitter has a dashboard which shows users the demographics of other Twitter users (analytics.twitter.com). It provides followers' languages, interests, consumer behavior, and even mobile carrier. However, these data are only about Twitter users in general. Analyzing the demographics of those who tweet on a certain topic is still not possible.

Below, we further explore the method Crimson Hexagon uses for geolocation analysis. It can zoom in on the geographical distribution of a topic by city and street level in any country around the world, based on the geotags provided. Twitter enabled location data in September 2009. However, only around 1% of all tweets can be tracked with this level of precision. For the remaining 99% of tweets without exact location, the estimation is based on multiple pieces of information, such as user profile, content, time zone and language preference. About 90% of all tweets can be located to the country of origin by using this algorithm. See the example below of a Crimson Hexagon-generated map (by country and by city).



Dataminr is another piece of software that is used to do social media insight analysis. It translates a Twitter feed and other public datasets into signals, which can give analysts the earliest notification on breaking, real-time events on social media. It also embeds geolocation addresses on the map so analysts can view the real time data and make predictions.

In sum, aside from self-identifying data (which many Twitter users choose not to disclose), we can extract demographics from business analysis software by translating IP addresses to geolocation data. Still, certain key demographic and lifestyle information, such as age, interests, and personal values, are hard to track. We can sample a small dataset to estimate the overall demographics, but accuracy remains a big challenge.

SUMMARY

This discussion paper offers structured methodology for addressing some of the key challenges involved in bridging the measurement gap between advocacy operations (outputs) and ultimate outcomes. We use digital advocacy campaigns through Twitter as a case in point. Search parameters, social media share of voice, the who, sentiment analysis and demographics are five key challenges in that process. According to our analyses, the

following best practices have been identified in addressing those challenges:

1. To ensure a comprehensive and accurate dataset, three-dimensional search parameters are recommended, which include relevant hashtags, mentions and keywords, along with the use of multiple monitoring and analysis tools each used according to their respective strengths.
2. Develop a comprehensive view of our assets versus those of competitors before establishing share of voice; and establish in advance which share of voice scenario is most valuable given your objectives and environmental context.
3. Implement an actor-type codebook to categorize influencers, in addition to using social network software to perform cluster analysis and identify those influencers who are hubs connecting individuals and organizations.
4. In order to address the challenge of the error-prone sentiment analysis currently offered by social media monitoring and analysis tools, engage in separate manual or automated sentiment analysis using a statistically representative sampling of the aggregate data pulled in by the tool(s). Stratified random sampling is recommended.
5. Supplement existing demographic data based on user-volunteered information with monitoring software that allows for the extraction of Twitter users' geolocation data by analyzing IP addresses and/or the content issued by those users. Recognizing the potential fallibility of the latter, this data can be used to attempt more demographically targeted messaging.

In conclusion, given the growing need for accountability in public diplomacy work and the continued expansion of digital diplomacy, this paper underscores the importance of taking a rigorous approach to social media analytics to enhance the validity of evaluative analysis for better strategic planning. We focus on five key challenges that came out of practical experience. The recommended practices will form the foundation for assisting practitioners in the exploitation of any and all data which can serve to bolster confidence that advocacy operations are contributing to success in terms of ultimate outcomes.

APPENDIX: Sample Codebook - Actor Types in an Environmental Social Media Campaign.¹¹

1. Mainstream media organizations, network and print. Formal organizational accounts of media organizations that have both a digital and non-digital presence, e.g. *Time*, *the Washington Post*, *VICE*.
2. Online-only media organizations: blogs, news portals, or journalistic entities that exist solely online, e.g. *Huffington Post*.
3. Journalists working for either mainstream or online-only media organizations, but who post under their own accounts.
4. Advocacy groups: the formal accounts of organizations devoted to a certain cause or goal, e.g. 350.org, the Sierra Club.
5. Advocacy media: organizations that are either issue-based around environmentalism, or that take an explicit political stand. Also includes the media wing/outlet of climate advocacy groups.
6. Activists: individuals who self-identify in their profile as an activist, work at an activist organization, have an advocacy group logo on their profile, or who post mostly about activist topics to capture the attention of others.
7. Political and governmental actors: the accounts of individuals or organizations who are known primarily for their relationship to government.
8. Celebrities: individuals who are famous for reasons unrelated to technology, politics, or activism; who are well-known enough to have their own Wikipedia page, e.g., Mark Ruffalo, Leonardo DiCaprio.
9. Bloggers: individuals who identify as a blogger and who appear to post regularly to an established blog.
10. "Regular" people who do not belong to the afore-mentioned categories.
11. Other: bots, parody accounts, accounts whose nature can't be determined. We coded (11a) for "TV show" or "movie" on Facebook.

NOTES

1. Zaharna, R. S., Arsenault, A., & Fisher, A. (2014). *Relational, networked and collaborative approaches to public diplomacy: The connective mindshift*. Routledge.
2. Bjola, C., & Holmes, M. (2015). *Digital Diplomacy: Theory and Practice*. Routledge
3. Gerbaudo, P. (2012). *Tweets and the streets: Social media and contemporary activism*. Pluto Press.
4. Yang, J., & Counts, S. (2010). Predicting the Speed, Scale, and Range of Information Diffusion in Twitter. *ICWSM, 10*, 355-358.
5. Howard, P. N., Duffy, A., Freelon, D., Hussain, M. M., Mari, W., & Mazaid, M. (2011). Opening closed regimes: what was the role of social media during the Arab Spring?. *Available at SSRN 2595096*.
6. Lewis-Beck, M., Bryman, A. E., & Liao, T. F. (2003). *The Sage encyclopedia of social science research methods*. Sage Publications.
7. Lotan, G., Graeff, E., Ananny, M., Gaffney, D., & Pearce, I. (2011). The Arab Spring | the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International journal of communication, 5*, 31.
8. <http://social-dynamics.org/twitter-network-data/>
9. The confidence interval (also called margin of error) is the plus-or-minus figure usually reported in newspaper or television opinion poll results. For example, if you use a confidence interval of 4 and 47% percent of your sample picks an answer, you can be sure that if you had asked the question of the entire relevant population between 43% (47-4) and 51% (47+4) would have picked that answer. The confidence is expressed as a percentage and represents how often the true percentage of the population who would pick an answer lies within the confidence interval. The 95% confidence level means you can be 95% certain; the 99% confidence level means you can be 99% certain. Most researchers use the 95% confidence level: <http://www.surveysystem.com/sscalc.htm>
10. <https://crimsonhexagon.zendesk.com/hc/en-us/articles/206452683-Workshop-Optimal-Uses-of-Auto-Sentiment>
11. Thorson, K., Edgerly, S., Kliger-Vilenchik, N., Xu, Y., & Wang, L. (2016). Seeking visibility in a big tent: Digital communication and the People's Climate March. *International Journal of Communication*.