

Twitter Health Metrics RFP Submission

Twitter [Blog](#), 3/1/18: *“Our goal in 2018 is to measure and improve the health of public conversation on Twitter. We don’t know all the answers. We do know that this work is critical yet may never be finished. And we cannot do this alone. So we’re asking you to help us define, measure, and evaluate our impact on the public sphere.”*

Full RFP: see [Appendix](#), p. 3.

These are responses to the 2 substantive Sections.

1. *“Propose and define a health metric Twitter could use to measure itself.”*

Proposed health metric: Twitter Verified Constituent accounts

Extend the Twitter Verified account model to also disclose the account owner’s political constituency at 5-8 levels of jurisdiction, verified by a trusted 3rd party identity reputation provider:

- State
- Congressional District
- State Senate (Upper) District
- State House (Lower) District
- County
- City (if applicable)
- City council district (if applicable)
- Unified school district (if applicable)

The default designation would NOT disclose the account owner’s name and address, just the account’s 5-8 political jurisdictions.

Health metric definition: Geolocal tribalism.

Tribalism is inherent in humans. Before the 21st century, all tribalism (and politics) was local. Social media has elevated ideological tribalism above geography, unmoored from the multiplicity of values present in any geographical community. The Twitter hashtag has evolved into the essence of ideological tribalism when used as a filter, allowing narrow, often

provisional, ideas to aggregate, separated from the moderating influences of neighborhoods and communities.

2: *“How would Twitter capture, measure & evaluate this health metric?”*

Capture:

1. Jurisdictional data will be captured using the GEOvoter API, hashed and placed on the HyperLedger blockchain.
2. Twitter has non-exclusive, defined access to the verified constituent’s credential token.
3. Verification ensures the account is owned by an accredited human with abundant metadata, contained in the VC’s identity token:

- **Twitter.US verified constituent token:**

```
"@brittb#PZMD081818-16620,1|http://newgov.net/demomap/index.html?&x_min=-77.1018087937648&x_max=-77.06747651837325&y_min=38.97409424002034&y_max=38.99077412232904|1514309206517|1520019016|1551555037"
```

- **Components:** (“|” separator)

- Twitter handle “@brittb”
- PowerZone: “#PZMD081818-16620,1”
- @brittb’s Power Zone map URL, like:
`http://newgov.net/demomap/index.html?&x_min=-77.1018087937648&x_max=-77.06747651837325&y_min=38.97409424002034&y_max=38.99077412232904`
- Unix epoch at first verification: “1520019016”
- Unix epoch at latest verification: “1551555037”
- &Hashed Key SHA-256 (SHA2) - 64 characters
“4ee65c4456602c21cd146c9e469791e4171526f13f93fe88422e444f12b744c5”

Measure:

Collectively, a verified politician’s verified constituents constitute a backchannel speaking more loudly and forcefully than the politician’s assertions.

Comments

After scrubbing bots and bad actors, there remains a vast population of Twitter users who are neither bots, nor banned, nor blue-checked and whose follower count is the only relevance indicator available for government representatives using Twitter to engage in conversations about policy. A system in which follower count equals influence shapes representative engagement to the power-law distribution rather than to the boundaries of the district they were elected to represent. This diverts representation toward social media influence at the expense of constituent influence.

We propose a system to verify US users at scale, correlate the user to all their political districts from National to local, and place an indicator of their verified status in their profile creating Verified Constituents, or VC's. District codes are then enabled in Twitter's filter and search functions so that candidates, representatives, and constituents can connect easily, effectively, and at scale. This realigns representation to reflect the interests of the constituency rather than follower counts.

We anticipate an immediate result in which the Twitter engagement of representatives will focus increasingly within their own districts. Given a sufficiently large population of VCs in a district, this hypothesis can be tested by correlating the representative's Twitter engagement to VC's in their district and tracking over time. To the extent that engagement within the district increases over time, we could say that Twitter will have made a positive impact on the "representative" part of our representative democracy.

Many secondary results would follow as the system matured. For example, the Twitter content of VC's becomes an auditable public record of constituent feedback which then drives greater transparency and accountability. Constituent tweets would be a more reliable source of feedback than office staff phone logs in which constituency is loosely verified, if at all. Many other possibilities open up if Twitter becomes a system of record for constituent feedback. For example, representative voting records can be compared to VC feedback on Twitter to determine the degree to which these correlate. The incentive arising from improved accountability and transparency should over time increase the alignment of representatives to constituents and tracking this metric over time should provide a basis to confirm or refute the hypothesis.

Perhaps the greatest impact may be greater overall participation in our democracy. Many voters don't know who their representatives are at the national level, let alone for their school district or city council. Candidates are

even less well known among their potential constituents. With Verified Constituency, Twitter can become their source for that information. Using the VC system, voters can tell who their candidates or sitting reps are. They can target tweets to specific local and national districts to engage with their candidates and reps. We anticipate that raising the level of constituent engagement throughout the year will result in a sustained improvement in overall participation. This too can be tested by correlating VC engagement metrics to real-world results such as voter registration, voter turnout, voter approval ratings, turnout at local council or board meetings, etc.

Proposed: A relevance engine based on verified constituency of Twitter members

- Verified Constituent profiles bear a visible indicator of that status.
- Representatives can opt into the enhanced features based on an account setting, or by becoming a VC themselves. Representative profile bears a visible indicator either way.
- Candidates are equivalent to representatives until elected or their candidacy ends.
- Search/filter on political districts from federal to local. Example: "#healthcare #NC8" returns tweets hashtagged #healthcare from Twitter members known to live in North Carolina's 8th Congressional District. (Exact syntax TBD but all districts are available. See below for details.)
- A new revenue stream from Verified Constituents. (Service is free to representatives to encourage their participation.)
- Possibility of free or tiered verification based on mapping of verified address to socioeconomic status. (Paymentless verification method required and TBD.)

Key Performance Metrics:

New metrics that can be derived from the communication stream among representatives and their constituency include (but are not limited to):

- Twitter KPM: Number of Verified Constituents enrolled ("VC"s).
- Twitter KPM: Number of representatives opted in or enrolled at each level from local to Federal.
- Twitter KPM: District participation at each level broken down as 0, partial, or 100% of potential reps opted in or enrolled.

- Twitter KPM: Engagement as a measure of use of district elements in filters, searches, lists, and moments. This is further sliced by VCs, representatives, and per-district.
- Twitter KPM: Correlate all other KPM's to voter turnout at election time.
- Rep KPM: Absolute numbers of conversations with constituency. Trends up or down over time.
- Rep KPM: Proportion of conversations with constituents compared to total. (High and/or rises over time.)
- Rep KPM: Voting record alignment with public feedback. (High and/or rises over time.)
- Voter KPM: Proportion of political tweets directed to members' own representatives. (Am I talking with the people who represent me?)
- Voter KPM: Representative impact based on number of Tweets to reps that get replies from those reps.
- Both KPM: Overall Twitter message volume within a given district.
- Both KPM: One's own Twitter message volume within any or all one's districts.
- Both KPM: District impact defined as degree to which original tweets generate replies, likes, retweets within the district.

Evaluate:

Overarching Assumption:

Interactions among Twitter-verified human constituents is presumed to be more authentic and “healthy”, as defined by Cortico. Cortico does not address the health of interactions based on the visibility of social media outside of one’s ideological “tribe”, based on physical proximity of participants rather than ideological closeness at a distance.

1. Shared Attention: Is there overlap in what we are talking about?
2. Shared Reality: Are we using the same facts?
3. Variety: Are we exposed to different opinions grounded in shared reality?
4. Receptivity: Are we open, civil, and listening to different opinions?

MIT:

The spread of malicious or accidental misinformation in social media...can have harmful effects on individuals and society.

we developed models for automated verification of rumors (unverified information) that propagate through Twitter.

To predict the veracity of rumors, we identified salient features of rumors by examining three aspects of information spread: I

Linguistic style used to express rumors,
characteristics of people involved in propagating information,
network propagation dynamics.

The predicted veracity of a time series of these features extracted from a rumor (a collection of tweets) is generated using Hidden Markov Models. The verification algorithm was trained and tested on 209 rumors representing 938,806 tweets collected from real-world events, including the 2013 Boston Marathon bombings, the 2014 Ferguson unrest, and the 2014 Ebola epidemic, and many other rumors about various real-world events reported on popular websites that document public rumors. The algorithm was able to correctly predict the veracity of 75% of the rumors faster than any other public source, including journalists and law enforcement officials. The ability to track rumors and predict their outcomes may have practical applications for news consumers, financial markets, journalists, and emergency services, and more generally to help minimize the impact of false information on Twitter.

Quotes

Some hard truths about Twitter's health crisis

TechCrunch, 3/10/18

"Dorsey says the "health" of conversations on his platform is now the company's "number one priority" — more than a decade after he typed that vapid first tweet, "just setting up my twttr", when he presumably had zero idea of all the horrible things humans would end up using his technology for."

"Dorsey says the "health" of conversations on his platform is now the company's "number one priority" — more than a decade after he typed that vapid first tweet, "just setting up my twttr", when he presumably had zero idea of all the horrible things humans would end up using his technology for."

Cortico, 3/1/18

"Our plan is to develop these indicators building on research from the MIT Media Lab and other organizations. We will make the indicators and the underlying methods by which they work publicly available and open for discussion and debate. We seek to analyze public discourse across social media and traditional broadcast media (e.g., talk radio and television)."

For the health indicators to serve the public effectively, we plan to work with social media platforms and other media distributors, content producers, advertisers, and the public to understand how the health of our public spheres is evolving, and what actions can be taken to improve it. We are happy to see Twitter aligned with our approach ([see here](#)) and hope to see other stakeholders engage in the future as well."

APPENDIX

Twitter Health Metrics Proposal Submission

Thursday, 1 March 2018

We're committing to helping increase the collective health, openness, and civility of public conversation around the world, and to hold ourselves publicly accountable toward progress. By measuring our contribution to the overall health of the public conversation, we believe we can more holistically approach and measure our impact on the world for years to come.

Twitter's health will be ***built and measured by*** *how we help encourage more healthy debate, conversations, and critical thinking*; ***conversely***, *abuse, spam and manipulation will detract* from it. We are looking to partner with outside experts to help us identify how we measure the health of Twitter, keep us accountable to share our progress with the world and establish a way forward for the long-term.

This approach is inspired by work from others in the industry. Cortico, a non-profit research organization, has spent time to more deeply understand the concept of measuring conversational health and developed four indicators to measure it: shared attention, shared reality, variety of opinion, and receptivity. We believe that we can identify indicators of conversational health that are even more specific to Twitter and its impact.

We don't have all of the answers and cannot do this alone, but know that the outcome will be stronger when we look to experts around the world for counsel and support. If you're interested in helping us define what health means for Twitter and how we should approach measuring it, please ***submit your proposal through the form below by Friday, April 13.***

To Apply

Applicants should submit a proposal through the online form available below by Friday, April 13, that includes:

Contact information (name, email, organizational details)

Your proposed health metrics, and methods for capturing, measuring, evaluating and reporting on such metrics

Anticipated resource requirements and methodology

Proposed output from your proposal and estimated time needed to capture, measure and evaluate health metrics

Relevant, peer-reviewed, publications and papers

Please make clear in your proposal if this is a joint application with another institution.

Award

Successful applicants will collaborate directly with Twitter's team, receive public data access and ***meaningful funding for their research. Funding will be provided as an unrestricted gift to the proposer's organization(s), to be provided in full at the start of the project.***

Our expectation is that successful projects will produce peer-reviewed, publicly available, open-access research articles and open source software whenever possible.

Timeline

Submissions are due Friday, April 13, and selected applicants will be invited to share further details on their proposals during May and June. We expect to announce the first selected projects in July.

Terms of Agreement

This proposal request in no way constitutes an agreement between Twitter and any applicant. The issuance of this proposal request does not imply that Twitter is making an offer to do business. Twitter reserves the right, in its sole discretion, to cancel this solicitation and proposal request at any time during the process. This proposal request is not and shall not be considered an "agreement to negotiate." Twitter reserves the right to make any award, group of awards, or no award as it determines in its sole discretion. Twitter may, in its discretion, amend, supplement, terminate, modify, negotiate or otherwise change any provision or part of this proposal request, including but not limited to the evaluation criteria, the process used for evaluation, and/or the expected timeline, at any time prior. Any award of this contract will be made to applicant(s) at the sole discretion of Twitter after consideration of the factors stated above. Twitter shall provide only public Twitter data (which shall be used solely for the purposes stated above) and such use shall be subject to additional terms and conditions (including, but not limited to, terms similar to Twitter's Developer Terms). Further terms may be included in a contract with successful applications.

Earlier material:

A premium service for Twitter **Verified Constituents** “**VC**”s:

- Accounts owned by verified and *credentialed* human beings
- US addresses
- Verified as constituents of 6-8 US political jurisdictions*
- Small monthly fee generates home address location.
- Revenue Model: VC’s are Twitter’s customers, not its products.

* Verifying Constituency: A Sovrin Use Case,

A Twitter **Verified Constituent** “**VC**” enjoys:

- More benefits and prestige than a Verified Account
(= verified account + constituency power over specific reps/committees)
- No advertising
- No promoted tweets
- Verified as a constituent in 6-8 political jurisdictions
- Real name & address not exposed.
- Twitter VC credentials updated monthly when credit card is charged.
- Based on account holder’s address.
 - Billing address
 - Delivery address
 - Geolocation
 - Authorized by VC
 - FourSquare, etc.
- A Twitter-issued VC Token is:
 - A hashed credential code (SHA-256 key)
 - Maintained on the HyperLedger Fabric (Linux Foundation).
 - Owned & managed by the VC account owner, not Twitter.
 - Portable
 - Extensible
 - Useful for multiple assertions & privileges.
 - Includes metadata in addition to political jurisdictions.

- After Twitter issues the Token:
 - Twitter maintains no personal data on VC.
 - Has permissioned access to Token.
 - VC can use the data any way she wants.
(based on the HyperLedger provisions)
- Twitter role:
 - Certifies VC's constituency to elected representatives.
 - Links to 51,231 League of Technical Voters jurisdiction maps:
 - 50 States
 - 435 Congressional Districts
 - 2,030 State Senate (Upper) Districts
 - 4,842 State House (Lower) Districts
 - 3,138 Counties
 - 29,267 Cities
 - 554 City council districts (wards) in 37 major cities
 - 10,915 Unified school districts
 - Each jurisdiction map:
 - Displays trending issue hashtags:
 - Used by that jurisdiction's Verified Constituents.
(UI Example: <http://GEOvoter.io>)
 - Optionally displays all issue hashtags in jurisdiction.
 - Can replay hashtag history to show adoption rate.