

[COVID Information Commons \(CIC\) Research Lightning Talk](#)

Transcript of a Presentation by Amit Sheth (University of South Carolina) & Valerie Shalin (Wright State University), June 2022



Title: [Semantic analysis of social media and new Big Data to understanding COVID-19's impact on mental health, addiction, and gender-based violence](#)

[YouTube Recording with Slides](#)

[June 2022 CIC Webinar Information](#)

Transcript Editor: Lauren Close

---

Transcript

Amit Sheth:

Slide 1

Ok, all right. So we're going to discuss our analysis of public health issues. When we analyze very large amount of social media content, the focus is on mental health and addiction related analysis and what we call social quality of index. The AI Institute of the University of South Carolina and Georgia State University, Wright State University are participants or the members for the team belong to these.

Slide 2

So we all know of the massive impact of this pandemic.

Slide 3

And this has also resulted in a significant impact on depression, anxiety, other mental health issues, on addictions and substance abuse.

Slide 4

So we said, well, can we understand how various policy decisions, social, economic, and public health policies, and the choices that government and policymakers have made affect the well-being of society - members of the society. And our toolbox is knowledge - enhanced social media analysis.

Slide 5

So, we position it as the following: we capture the data between the period of, as you see here, March 2020 and end of July to January 2021. That covered two major COVID-19 waves. We had 25 crawling terms. We collected 12 billion tweets. In addition to this big data, we had a lot of other data to be used. And that is a very important methodical part of the study because it's not just analyzing social media. Data is not enough, you need lot of other relevant information to be able to analyze the data, so we had location related information, we looked at specific subreddits to train our language models, and we had mental health related knowledge bases, DSM-V drug abuse ontology, DAO, DBpedia, Wikidata, UMLS.

Additionally, we had to look at on the ground events. For example, you know, loan relief, household income relief, testing policy, all those things that all those choices that various state leaders made or federal, you know, policymakers made. So we need to also have an understanding of what happened when and where. And there are other events that also we have to understand.

Slide 6

Our focus was on understanding the, you know, content on social media related to mental health. So you can see some examples here.

Slide 7

And we developed an empirical measure we call Social Quality Index. It aggregates mental health components and addiction and substance use disorder components, right? So analyzing the social media data for these components to come up with a Social Quality Index.

Slide 8

There is a pretty comprehensive infrastructure as analysis of this kind is pretty daunting, so you have Twitter data, but you have news articles. I'll quickly mention why your news article - your various knowledge base. Then, you have to do a whole variety of analysis, you can see there in the second vertical box then it goes to language models and topic models, semantic mappings, so deeper understanding. And, you know, our team works on knowledge infused learning, so we have a particular, you know, enhanced deep learning methods that uses knowledge of the domain to better understand the language. From there on, to understand content related to depression, anxiety, addiction, substance use, all in the context of COVID, leading to SQI calculations, there were language model training components, and all those that are not shown here in this picture.

Slide 9

So what are the innovations? We used news to continuously identify new entities of interest into COVID. And in this context, we use location extractions using Geonames Ontology, Open Street Map API and other things. We used multiple vocabularies and knowledge graphs for semantic extraction, we used a specialized subreddit corpus for language model training and topic models. And we trained classifiers to scale the analysis of big data. So we couldn't do the, you know, in our analysis - to analyze such large amounts of data, we have to create these classifiers and then understand depression, addiction, anxiety, and other issues. With that background, I'm going to pass it on to Valerie to explain what we found.

Valerie Shalin:

Slide 10

Thank you, Amit. Let me give you an idea of the kind of analyses that become possible with the kind of capabilities that Amit just described. So here are some sample state graphs of COVID rates and SQI over time. Like Brandon [Johnson], we think time is absolutely the issue here. So COVID rates per capita is on the left, y-axis. And SQI is on the right axis. Up is bad, down is good. Time is on the x-axis and there's a discontinuity there marked with the hash marks. And I'm not expecting you to read these graphs carefully except that you'll certainly notice that there's very little correlation between the Social Quality Index that we've identified and COVID prevalence. There is one place - whoops, go back one slide, just for a second - there is one place that you'll notice a correlation. And that is towards the end, and we think that that is likely reflecting a latent variable. That is the holiday season, that's the winter holiday season, and so probably what's going on is that the winter holidays are increasing both COVID rates and we already know that the winter holidays are very challenging for mental health. So that makes a lot of good sense and it serves as an informal validation of the SQI metric. Next slide.

Amit:

I can't advance, so I don't know what - Lauren, I can't...

Lauren Close:

Let's see, I may be able to share on your behalf.

Florence Hudson:

Did you try just clicking on the slide or using the up or down arrows? Did you try a few different options? Ok, there you go.

Valerie:

Slide 11

Ok, so one of the points that we want to make is that the holidays of course are not the only event that we can now examine. And this is a graph for the state of Washington. And what we're showing you there is the policy changes that we've identified in news aligned with the various periods that we have marked out there with the vertical lines. So our point is that it's now possible to look for patterns relating this Social Quality Index to the implementation of policy by state. Next slide, Amit.

Slide 12

So SQI is a standardized measure. It's independent of state population or amount of social media activity and because it's standardized in that way we can compare states. And we've done - next slide.

Slide 13

Some state clustering to see if we can discern patterns. And you see that Connecticut, Louisiana, New Jersey, Nevada, etc. all seem to be following the same kind of pattern. SQI is bad at the beginning, it gets better, and then gets worse. This is for a truncated high resolution time period. And then a different cluster: Florida, Georgia, Michigan, etc. have an SQI better - SQI worse pattern. So we can reveal some of these patterns over time. Of course, we have to do more longitudinal and time series analyses to pull out the potential causes, but it does look to us like it's the economic issues that are going to be the major factor here. Next slide.

Slide 14

So what is now possible with the tools that the South Carolina team has developed? We have a knowledge driven analysis of social media at an abstract level and it provides a very high resolution data set, both with respect to time and space. It is rapid and it provides separable measures of mental health that are not feasible in survey metrics. We can get lots and lots of data spread out over large portions of space and time and that's the kind of data that you need in order to pull out the confounds and begin to evaluate whether or not your policy decisions were having an effect on your population. So this is a really important research tool that advances our capabilities in science, but more broadly it enables real-time monitoring and preparation for mitigation. And of course it supports policy makers in general. And that is our presentation for you.