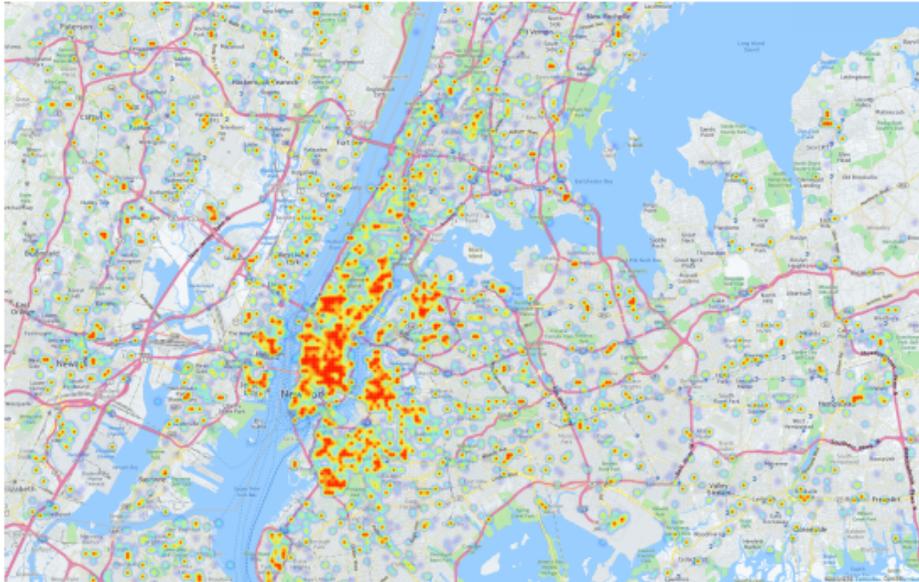


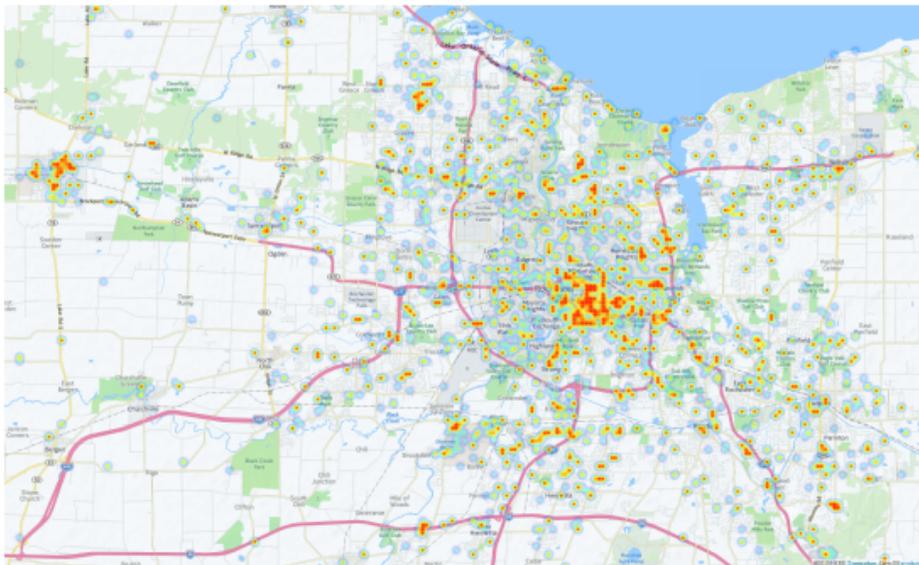
# **An algorithm that figures out if a tweet was sent while drinking**

March 18 2016, by Bob Yirka

---



(a) NYC



(b) Monroe

Heat maps of user-drinking-now tweets showing unusual drinking zones. In NYC, the drinking hot spots are Lower Manhattan and it's surroundings whereas in Monroe County they are Downtown Rochester (center) and the city of Brockport (left). Credit: arXiv:1603.03181 [cs.AI]

(Tech Xplore)—A team of researchers with the University of Rochester has developed a machine-learning algorithm that is able to determine if a tweet was sent while someone was under the influence of alcohol. The team also describes in their paper uploaded to the pre-press server *arXiv*, how they used some of the same data to improve location identification of user tweets.

Sending text messages while [drinking](#) has become a popular pastime, and now as the popularity of Twitter grows, more and more people are sending messages to the world at large while under the influence. In this new effort, the [researchers](#) used workers on Amazon's Mechanical Turk to help them develop a [learning algorithm](#) that could distinguish between ordinary tweets and those made by people who are drinking or have been drinking for some time.

The group started by collecting geotagged tweets, sending thousands of them to Mechanical Turk, where human workers were asked to state whether they thought a given message was alcohol related, and then if they thought the author had been drinking and if so if they had been doing it while tweeting the message. That gave the researchers enough data to feed to a [machine-learning algorithm](#) trained to look for the same answers.

Once the team was satisfied that their algorithm was reasonably accurate they moved on to improving it to answer questions about where the person was while drinking—at home, at work, at a nightclub? To help the machine answer such questions, the researchers filtered tweets for words or phrases that indicated location, like "at home" or "in front of the telly" and sent them once again to Mechanical Turk workers who were asked to judge whether the [tweets](#) were from someone's home, or not. That data was then fed to the algorithm which then cross referenced drinking and location, offering a reasonable estimate of where a given tweeter was while imbibing (with an estimated accuracy of 80 percent

within 1000 meters). And that inevitably led to the creation of maps with dots showing the locations of everyone drinking at a given time, based on 100x100 meter grids.

The researchers note that their data and maps were for New York City and Monroe County only, but that was enough to show trends, such as people in the city appeared to drink more at home than those in Monroe County. They suggest their [algorithm](#) could be used for demographic purposes, or even perhaps to help police figure out where to set up random testing stations to reduce alcohol related driving accidents.

**More information:** Inferring Fine-grained Details on User Activities and Home Location from Social Media: Detecting Drinking-While-Tweeting Patterns in Communities, arXiv:1603.03181 [cs.AI]  
[arxiv.org/abs/1603.03181](http://arxiv.org/abs/1603.03181)

### **Abstract**

Nearly all previous work on geo-locating latent states and activities from social media confounds general discussions about activities, self-reports of users participating in those activities at times in the past or future, and self-reports made at the immediate time and place the activity occurs. Activities, such as alcohol consumption, may occur at different places and types of places, and it is important not only to detect the local regions where these activities occur, but also to analyze the degree of participation in them by local residents. In this paper, we develop new machine learning based methods for fine-grained localization of activities and home locations from Twitter data. We apply these methods to discover and compare alcohol consumption patterns in a large urban area, New York City, and a more suburban and rural area, Monroe County. We find positive correlations between the rate of alcohol consumption reported among a community's Twitter users and the density of alcohol outlets, demonstrating that the degree of correlation varies significantly between urban and suburban areas. While our

experiments are focused on alcohol use, our methods for locating homes and distinguishing temporally-specific self-reports are applicable to a broad range of behaviors and latent states.

© 2016 Tech Xplore

Citation: An algorithm that figures out if a tweet was sent while drinking (2016, March 18)  
retrieved 19 April 2024 from <https://techxplore.com/news/2016-03-algorithm-figures-tweet.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.