



# Socio-spatial Self-organizing Maps: Using Social Media to Assess Relevant Geographies for Exposure to Social Processes

KUNAL RELIA, New York University, USA

MOHAMMAD AKBARI, New York University, USA

DUSTIN DUNCAN, New York University School of Medicine, USA

RUMI CHUNARA, New York University, USA

Social media offers a unique window into attitudes like racism and homophobia, exposure to which are important, hard to measure and understudied social determinants of health. However, individual geo-located observations from social media are noisy and geographically inconsistent. Existing areas by which exposures are measured, like Zip codes, average over irrelevant administratively-defined boundaries. Hence, in order to enable studies of online social environmental measures like attitudes on social media and their possible relationship to health outcomes, first there is a need for a method to define the collective, underlying degree of social media attitudes by region. To address this, we create the Socio-spatial-Self organizing map, “SS-SOM” pipeline to best identify regions by their latent social attitude from Twitter posts. SS-SOMs use neural embedding for text-classification, and augment traditional SOMs to generate a controlled number of non-overlapping, topologically-constrained and topically-similar clusters. We find that not only are SS-SOMs robust to missing data, the exposure of a cohort of men who are susceptible to multiple racism and homophobia-linked health outcomes, changes by up to 42% using SS-SOM measures as compared to using Zip code-based measures.

CCS Concepts: • **Computing methodologies** → **Cluster analysis**; **Neural networks**; *Information extraction*; *Classification and regression trees*; • **Applied computing** → *Sociology*;

Keywords: Clustering; racism; homophobia; self-organizing maps

## ACM Reference Format:

Kunal Relia, Mohammad Akbari, Dustin Duncan, and Rumi Chunara. 2018. Socio-spatial Self-organizing Maps: Using Social Media to Assess Relevant Geographies for Exposure to Social Processes. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 2, CSCW, Article 145 (November 2018). ACM, New York, NY. 23 pages. <https://doi.org/10.1145/3274414>

## 1 INTRODUCTION

Public health, economic and other social outcomes depend on what we experience around us. Thus there is a need to identify the relevant environment around an individual in relation to a specific exposure. For example, in which areas of a city are we likely to feel happiness? Or conversely, where will we experience discrimination? In particular, many health outcomes depend

Authors’ addresses: Kunal Relia, New York University, Tandon School of Engineering, USA, [krelia@nyu.edu](mailto:krelia@nyu.edu); Mohammad Akbari, New York University, Tandon School of Engineering, Center for Data Science, USA, [akbari@nyu.edu](mailto:akbari@nyu.edu); Dustin Duncan, New York University School of Medicine, Department of Population Health, USA, [dustin.duncan@nyumc.org](mailto:dustin.duncan@nyumc.org); Rumi Chunara, New York University, Tandon School of Engineering, Center for Data Science, College of Global Public Health, USA, [rumi.chunara@nyu.edu](mailto:rumi.chunara@nyu.edu).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-0142/2018/11-ART145 \$15.00

<https://doi.org/10.1145/3274414>

on environmental exposures; it has been shown that there is a link between social exposures like racism and homophobia to disease and risky health behaviors [8, 30, 40]. There are many hypothesized mechanisms for this link. Discrimination against black workers in employment can lead to lower income and greater financial strain, which in turn have been linked to poor health outcomes [20]. Racism may also directly impact health by engaging psycho-biological mechanisms induced in the stress response [14]. A major challenge to this work has been obtaining measures of discrimination in the community. It has been shown that it is difficult to measure sensitive attitudes like racism or homophobia via surveys, due to biases like social desirability bias [54, 65]. Survey-based measures are costly and laborious to obtain widely or with high granularity, such as at within-city levels. Moreover, mapping results to Zip codes results in unnecessary and large spatial averaging [26]. Such attitudes can also be dynamic (e.g. socio-political events could cause increases in discrimination), and deploying surveys at an appropriate timescale to discern these changes would be labor and cost intensive. Measures of discrimination such as from hate crime reports also provide an incomplete picture of discrimination; they only capture discrimination that results in a crime being reported. As well, hate crime reports can also suffer from biases affecting the type of reports and/or who contributes them.

In social computing research, social media has been shown to provide a unique window into the social experience of people, and in particular Twitter has been used for assessing sensitive topics, such as discrimination [10, 21, 22, 60]. In order to use such measures of discrimination in downstream analyses, such as studying the link between online discrimination reports and health, we need a way to assess the latent sentiment in a location from the Tweets. This is challenging because a single Tweet may be noisy (e.g. its content regarding racism could be unclear). As well, when resolved to very high-resolution geographic areas, Twitter data can be sparse. Further, a single Tweet may not be representative of the social sentiment in an underlying area. Therefore, there is a need for methods to define the collective latent sentiment of an area based on inconsistently distributed point-observations from geo-located data such as Tweets.

Accordingly, we present a social computing pipeline for identifying relevant geographical areas of social processes, as defined by text from Twitter posts. Our approach, SS-SOMs divides the city into subareas, defining the optimal collective sentiment by location. Our work accounts for the observational nature of social media by accounting for varying Tweeting levels by location. We compare the method to appropriate baselines using prevalence of homophobia (expression of a homophobic sentiment, or of an experience of homophobia) and racism (conveys a racist message, or indicates someone experiencing racism) on Twitter, as significant and understudied forms of discrimination, and conducive to being measured via social media (though the pipeline can be applied to other social processes, especially those uniquely represented on social media). We focus on New York City (NYC) as discrimination within urban cities is highly varied due to proximity of neighborhoods with dramatically different norms and cultures. As well, health disparities for affected populations are significant in cities. Using data from a cohort of racially/ethnically and socioeconomically diverse young gay, bisexual and other men who have sex with men (MSM), who are disparately affected by racism and have a unique footprint in NYC [16, 27], we compute how the relevant online social media measures of racism and homophobia for them in NYC measured via SS-SOMs would be different than via Zip codes. Therefore, this work addresses the need for new measures of racism and homophobia especially within urban cities. Specific contributions are:

- A pipeline to partition an area into regions of collective, consistent sentiments from social media data (SS-SOMs).
- Evaluation of the quality of SS-SOM clusters compared to relevant baselines.
- Demonstration of impact by comparing online homophobia/racism exposure measures via SS-SOMs v.s. Zip codes for an MSM cohort in an area (NYC) where disparities are significant.

## 2 RELATED WORK

### 2.1 Characterizing Environments by Social Media

Concentration of online activity has been used to define locations from Flickr photos [24], and spatial statistics have been combined with such data to identify specific regions via particular photo tags [59] (but only identifying those locations captured in photos, and not partition an entire geographic area). Foursquare check-ins have been used to identify “neighborhoods” [18, 19, 49, 56]. This work has identified clusters of similar types of check-ins to characterize the physical environment based on activity type (e.g. shopping area versus restaurants, etc.). This line of work only focused on specific types of areas, is focused on identifying those specific “neighborhoods” and not partitioning an area comprehensively. Other work has assessed location data and the reliability of social media tags [37]. Finally, sound types have been identified via social media. Sound is a localized phenomena which can be mapped to precise locations (and can be compared to other data at that level), instead of by region [1]. Notably, the above approaches are built on observational, numerator-only data – thus resulting clusters can be confounded by areas where or types of locations in which individuals tend to check-in more frequently [13, 56]. This matters less when only specific locations/neighborhoods are being identified, compared to if partitions of an entire area are desired.

As the relevant environment around us goes beyond the physical and includes social processes around us, our work is novel because it considers the text of Tweets to characterize social environments (which hasn’t been done from a geographical point of view). As the social environment depends on anywhere there are people, our goal is also different from the above; we aim to partition (e.g. create contiguous clusters) to allow for understanding of exposures at any given location. Moreover, the task of describing the *social* environment helps us define constraints for the partitioning method. We require a set of subareas that are collectively exhaustive for the area they divide, contiguous and mutually exclusive, and each subarea should represent an exposure level that best exemplifies all of the individual social media posts they are defined by.

### 2.2 Methods for Identifying Spatial Structure and Generating Boundaries

Generating appropriate boundaries is an active research area, given the increasing amount of geo-located data. However the specific challenge of defining areas of consistent social attitude is different from previous work. For such social attitudes, individual Tweets can be noisy (the text of an individual Tweet may not be clear regarding the attitude), are not consistently generated everywhere, and to be useful in assessing health outcomes, must be linked to a unique (non-overlapping) area representing the underlying sentiment. Given these constraints, we are specifically charged with developing homogeneous, contiguous partitions of the specified geography (a continuous field, for example using kernel methods would not be appropriate for the application).

A variety of methods can be applied to uncover hidden spatial structure in geographic data, including clustering [24], density estimation [34, 43] and neural networks [31]. To define location representations/boundaries from Flickr and Foursquare check-in data some approaches have harnessed burst-analysis techniques which model the distribution probabilistically, highly peaked over a small number of more nearby values [59] or common clustering methods such as DBSCAN (Density Based Spatial Clustering of Applications with Noise) which is an algorithm for noisy data [39, 61], K-means clustering [49], and DBSC (Density-Based Spatial Clustering) which focuses on content similarity and spatial proximity equally but doesn’t guarantee to partition a region [53].

Other work identified irregularities in amount of Tweeting by location over time [50] using K-Means clustering and Voronoi polygons [29]. In epidemiology, environmental exposures are traditionally quantified via Zip codes and census tracts [12]. While these approaches (Voronoi polygons, Zip codes) fulfill the criteria for our social process area partitions: they define a set of

subareas that are collectively exhaustive for the area they divide, and are contiguous and mutually exclusive, the resulting regions are defined administratively or based on amount of data and not in a manner relevant to the exposure. Therefore computing the average social attitude over these areas will incur unnecessary spatial averaging.

A sophisticated approach for defining geographic areas uses artificial neural networks (ANNs); an unsupervised learning approach [47]. The input signal (vector containing information about the attributes of data to be mapped) is linked to a spatial location and the “self-organized map” (SOM), is organized based on the amplitude of these signals. Many different adaptations of SOMs have been proposed spanning organization of the input vector [69], algorithm [23, 45], or layout of the output space [51]. There has been a specific focus on preserving information about the topological distance between input nodes. Many such modifications can be grouped into 3 approaches: (1) including geo-coordinates as a part of the input vector, (2) calculating topological distance between output nodes instead of the distance between the weight of the nodes to localize the SOMs [23] and, (3) changing the SOM “neighborhood” function to cover a wider width [45]. Modifications on inclusion of geo-coordinates have included: (a) using a combination of the weight vectors and neuron spatial positions to measure topological distance between points and cluster them together [44, 69] and (b) searching for the best matching unit (nearest node) only within a predefined topological vicinity (called Geo-SOMs) [3]. Both of these modifications lead to well-defined but overlapping clusters. Consequently, a major shortcoming of this is the difficulty in developing contiguous positions of resulting areas [35]. Despite these varied approaches, to our best knowledge, there is no method that guarantees the resulting clusters to be at once topologically constrained, contiguous, non-overlapping, and allowing for control over the number of clusters formed.

### 2.3 Data for Tracking Racism/Homophobia

Geographers and social scientists have examined sentiments like racism for decades. Causal mechanisms between racism and health outcomes have been clearly established [40]. These conclusions have been reached in multiple settings, like in the workplace, racial discrimination has been shown to relate to adverse health outcomes [20] – the link being slightly larger for men versus women, but significant for both. More broadly are community-level influences, which are why here we focus on social media data across and within a city. Research has shown that neighborhood influences like racism and homophobia can play an important role in influencing health behaviors including substance use and condomless sex among MSM [8, 30]. Other recent work has shown links between racism in communities and mental health outcomes specifically in black populations [5].

One approach to assess racism has been based on geographic clustering or concentration of different racial groups, to extract measures of segregation [67]. While segregation is an important factor in ascertaining racism, it doesn’t directly speak to the experience of individuals. Approaches to measure experience with racism have largely been through survey mechanisms [2]. Concerns about how well surveys capture all aspects of the prejudice have been highlighted (e.g. based on terms being stigmatized, social censoring, different population groups having different self-report biases, and the validity and reliability of measures) [17]. A new proxy for an area’s racial animus from a non-survey source is the percent of Google search queries that include racially charged language [62]. Here it was noted that Internet-reported data is not likely to suffer from major social censoring, and is a medium through which it is easier to express socially taboo thoughts [48]. More recently, Twitter has been used to identify and spatially map racism [10]. This work has parsed Twitter data via keyword filtering, summarized and compared example Tweets expressing racism.

More recently, Twitter has also been used to systematically assess the main targets of on-line hate speech (defined as bias against an aspect of a group of people, thus broader than just

racism/homophobia in online social media) [60]. Notably from this work, race and sexual orientation were amongst the top hate speech targets on Twitter (as well as Whisper). Overall, the prevalence of hate speech on social media is considered a serious problem. Finally, because we aim to understand racism and homophobia as comprehensively as possible spatially, we include both reports of exposures and racist/homophobic statements, both of which are possible on Twitter.

3 DATA

3.1 Twitter Sample

The Twitter Application Programming Interface (API) was used to source geo-located Tweets having point coordinates within the boundaries of New York City (NYC). We initially used a generous bounding box enclosing all five NYC boroughs: 40.915256 N to 40.496044 N and 73.700272 W to 74.255735 W [55]. Tweets that were outside the precise NYC boundaries were later filtered out while mapping the Tweets to grid cells, as discussed in section 4.2.1. The time period for Tweets was selected to exactly overlap the time period of the MSM cohort mobility data (January 25, 2017 to November 3, 2017). This resulted in 6,234,765 Tweets. Data from NYC was used as it provides an ideal environment based on data volume, population density, and also addresses the increasing imperative to understand social and health disparities specifically within urban areas [36].

*3.1.1 Training Data for Racism Classification.* In line with the typical size of training data sets for Twitter classification efforts, we used 9785 Tweets for the training set [66]. As the topic is very nuanced, to capture a comprehensive set of training examples we used a combination of keyword filtering, labelling, and iterative learning to generate this training data.

*Keywords for Racism Classification.* We used findings from the existing literature on social media and various forms of discrimination [9, 10, 62], in combination with Urban Dictionary and Wikipedia to generate a dictionary of keywords (Appendix) to source “racism” Tweets. As discussed above, in order to ensure we were capturing Tweets that represented exposure to racism comprehensively, our definition for a “racism” Tweet is one that conveys a racist message, or indicates someone experiencing racism. We used this dictionary to source 100 recent Tweets for each keyword from the Twitter API. This initial keyword filter resulted in 3868 Tweets. Notably, some keywords were derived from a previously developed Internet-based measure of “area racism” that did not rely on the provision of responses to survey questions and thus is less susceptible to social desirability bias [62]. Further, we highlight that deciding what is a “racism” Tweet is a nuanced issue and we wanted to make sure not only that the keywords selected were consistent with other recent papers using Internet-data to measure racism, but that the data did not over-represent colloquial language. For example, not all text that contains the “n-word” are motivated by racist attitudes. Specifically, more colloquial forms of the “n-word” (i.e., ending in “-a” or “-as” vs. “-er” or “-ers”) were not included given that prior work found that these versions were used in different contexts [62]. Stephens et al. found that after this exclusion, prevalence of the resulting terms did relate to racial attitudes in a

Table 1. Top Twitter features after classification.

Classifier	Racism top features	Homophobia top features
Neural Model	useless_nigger, white_trash, coon, stop_chinaman, racist_bigot	hate_same_sex, gay_marriage, son_is_gay, hooker, lads_kissing
SVM	me_a_nigger, ching_chong_chinaman, white_trash, suck_it_niggers, useless_beaner	gay, gay_people, son, I_hate, being_gay_is



geographic area measured through disparate voting practices for a black candidate [62]. To further ensure the keywords selected do not result in colloquial language being included, we also examined the top features post-classification (using the final training data) (Table 1) which confirmed the resulting labels were indicative of our criteria and not colloquial language.

*Labelling of Tweets by Human Annotators.* We clearly defined the criteria for labelling a Tweet as indicating “racism” (versus “no racism”) as described above, to Amazon Mechanical Turk (AMT) workers for annotation, and through initial trial experiments confirmed the clarity of our instructions (included in the Appendix). As this task involved exposure of humans to potentially sensitive content, we clearly indicated the task is about racist Tweets, and created each Tweet as an individual Human Intelligence Task (HIT) giving workers a chance to discontinue at any point without losing payment if they felt uncomfortable. The workers were paid the minimum rate suggested by AMT when allocating the task (resulting in the amount paid being equivalent to \$14.40/hr [28]). Each Tweet (only the text of the Tweet was provided) was labelled by two workers on AMT and Cohen’s kappa statistic was 0.62 [15]. Our team member manually examined Tweets with a disagreement to resolve the label. In the end, 2040 of the 3868 Tweets sourced by keyword filtering were labelled as expressing “racism”. To add data from the negative class to our training set, in a ratio similar to other work, we added 4188 Tweets without any racist keywords in them [7, 66].

*Iterative learning.* Finally, to improve our training data set even further we aimed to find and include labels for Tweets that may be difficult to classify. To do this, we used an iterative learning approach as previously implemented with Twitter data [38, 52] to ensure our classifier was clearly capturing the definitions of racism and homophobia we stated. In this step we classified approximately 10000 unlabelled Twitter posts (from the NYC set used for this project), using the 8056 labelled Tweets from the previous section. We used the SVM classifier as described in section 5.1.1. We then manually labelled the edge cases (approximately 5% Tweets with an output label from classification that was near the threshold). We added these labelled edge cases to our training data set with the aim of increasing the number of edge labels in the training data which may improve the classifier’s performance [38, 52]. We did this successively until the SVM classifier performance plateaued (after three iterations). In the end, this increased the training data size from the 8056 to 9785 Tweets, 2398 of which indicated “racism” and 7387 “no racism” [66].

**3.1.2 Training Data for Homophobia Classification.** The same procedure as above (keyword filtering, labelling by human annotators and iterative learning) was used to create the training data for Tweets that express or represent an experience of homophobia.

*Keywords for Homophobia Classification.* In contrast to the study of racism on social media, there is not as much precedence in the literature regarding social media expressions of homophobia. Therefore the set of training Tweets for homophobia derived using keywords from Urban Dictionary and Wikipedia alone was not as comprehensive as the resulting Tweets for racism. Thus we added Tweets from the “@homophobes” Twitter account. The explicit purpose of this account is to Retweet homophobic Tweets from anywhere, so this would increase the linguistic diversity of the training data set. All 1626 Tweets and Retweets as of December 2017 were obtained from the account “@homophobes” via the Twitter API, and added to 1177 recent Tweets containing the homophobia keywords listed in the Appendix.

*Labelling of Tweets by Human Annotators.* We clearly defined the criterion to annotators that a Tweet labelled as “homophobia” (versus “no homophobia”) should either express a homophobic sentiment, or express an experience of homophobia. Tweets (only the text of the Tweet was provided) were labelled by two workers on AMT and Cohen’s kappa statistic was 0.71. Disagreements were resolved by our team. At the end of the labelling stage, 1990 of the 2803 Tweets sourced in the previous section were labelled as “homophobic”.

*Iterative learning.* Following the same approach as in racism classification, SVM and iterative labeling were used to classify the homophobia Tweets. Finally, this resulted in a training data set of 2176 “homophobia” Tweets, and 7535 “no homophobia” Tweets.

### 3.2 MSM Cohort Data

To demonstrate the effect of measuring online social discrimination through SS-SOMs versus other spatial partitioning methods, we use mobility data from a cohort of young MSM who are part of the National Institutes of Health funded P18 Cohort Neighborhood Study [36]. It is important to calculate the potential exposure-space for this group specifically, as 1) they are susceptible to multiple racism and homophobia-linked health outcomes, and 2) the locations they frequent and thus their environmental risk-space is unique compared to other populations [16, 27, 46, 64]. P18 is an ongoing prospective cohort study of sexual behavior, substance use, and mental health burdens that includes racially/ethnically and socioeconomically diverse young MSM. We measure the exposure of racism and homophobia specifically on this cohort of MSM because it is widely known that experiences of social discrimination among the members of minority groups like MSM can lead to poor health outcomes [11].

Participants were eligible for the original study if they were 18 or 19 years old, biologically male, lived in the NYC area, reported having had sex with another male in the 6 months before screening, and self-reported a HIV-negative or unknown serostatus. Those in the neighborhood part of the study consent to carry a GPS tracker for two weeks. The tracker logs the precise geo-location every 10 seconds. In total, at time of this paper preparation, we had mobility data from 226 men. Data was collected between January 25 and November 3, 2017. The minimum amount of time an individual wore the tracker was 6 days, maximum was 82 days. The mean duration was 21.3 days.

## 4 SOCIO-SPATIAL SELF-ORGANIZING MAP PIPELINE

The SS-SOM pipeline will partition an area into sub-areas that are collectively exhaustive, contiguous and mutually exclusive, where each subarea represents homogeneous exposure that best exemplifies the individual social media posts they are defined by. Implementation of this method is a three staged process where (1) Tweets are classified using a neural network and the training data generated above, (2) classified Tweets are then mapped to grid cells used to divide the entire NYC to account for varying level of Tweeting patterns across the city and (3) an augmented version of the Self-organizing Map is used to partition the city into the regions that are most similar in the social attitude, and contiguous. Figure 1 illustrates these steps.

### 4.1 Tweet Classification using Neural Models

To classify Tweets we use neural learning models, which have been shown to have improved performance over traditional classifiers [25, 63], especially for short texts such as Twitter messages which contain limited contextual information. As the models combine the small text content with prior knowledge, they therefore perform better than models using just a bag-of-words approach. To implement the model, we create an embedding for words in our Twitter data set and use cosine distance to measure similarity between those Tweets in the embedded space. This approach is similar to the popular Word2vec model [33], but is trained on our actively learned racism and homophobia training data sets (section 3.1). The model is a shallow, two-layer neural network that is trained to reconstruct linguistic context of words from its training text corpus input. The model produces a high-dimensional vector space, with each unique word in the corpus being represented by a vector in the space. Thus words that share common contexts in the training corpus are located in close proximity to one another in the space. We validate this model by comparing its performance to a baseline classifier (Support Vector Machines (SVM)) [41].

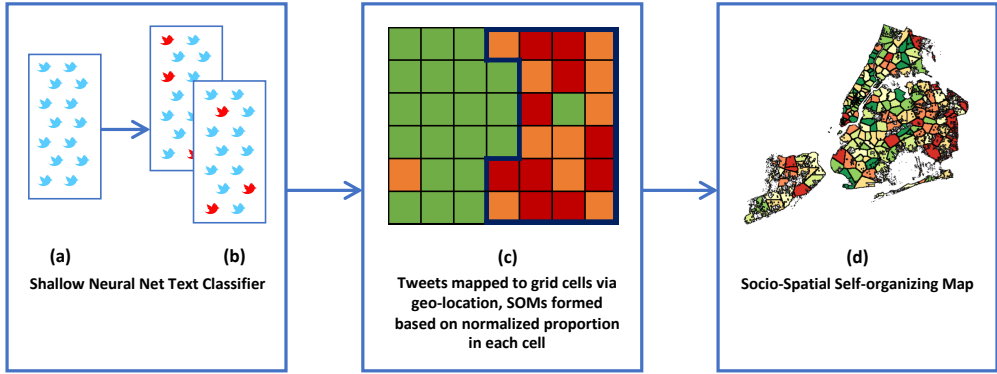


Fig. 1. The SS-SOM pipeline. (a) Tweets made within the NYC bounding box are classified as (b) racism vs. no racism and homophobia vs. no homophobia using a shallow neural network, where the resulting probabilities of Tweet is used to classify the Tweet. (c) NYC is divided into grid cells, geo-location of each Tweet is used to map Tweets to the cells, and the normalized count is represented by different colors of the grid cells - from red (high racism/homophobia) to green (no racism/homophobia). SS-SOM clustering is performed (example boundary represented by the dark blue border). (d) Resulting SS-SOM map for prevalence of racism in NYC.

**4.1.1 Shallow Neural Network.** In our model we maintain an  $n$ -gram approach coherent with the baseline SVM classifier (section 5.1.1); the features used for the embedding ranged from one to six words, half the number of average words per Tweet [34] (character  $n$ -grams did not improve performance as in other work with similar size data sets [38]). Given a word vocabulary  $W$  of size  $N$  derived from the labeled training data,  $w_i : i \in \{1, \dots, N\}$ , where the goal of neural embedding is to learn a distributed representation for each word  $w_i$  accounting for the semantic similarity of words while simultaneously computing the probability of distribution over the predefined classes. More specifically, the objective of embedding is to minimize the negative log-likelihood:  $-\frac{1}{N} \sum_{t=1}^N y_i \log(f(x_i))$  where  $x_i$  is the  $n$ -th document,  $y_i$  is the label, and  $f$  is an embedding function which maps the text document into a vector representation. Following research efforts in word embedding, we first embed each word in the text with an embedding vector and then employ the average of word embeddings to form the tweet embedding. This is in turn fed to the final layer, a softmax function, to define the probability of a class given the embedding of the input text. We will then have two outputs for each Tweet: the probability of being racist, and non-racist (and similar for homophobia), Tweets classified using a probability threshold of 0.5.

## 4.2 Adjusting for Varying Spatio-Temporal Tweeting Levels

**4.2.1 Tweeting Levels over Space.** In order to account for varying levels of Tweeting by location, we normalized the number of racism/homophobia Tweets by the total number of Tweets by dividing NYC into a series of grids (spatial bins), and using these grids with a normalized proportion in subsequent classification analyses. Work using Foursquare data has shown that the geographic size of the bins is important to consider as it can alter the resulting spatial distribution [57]. Here we discuss how we chose an appropriate spatial bin size that would account for the spatially inconsistent nature of Twitter data, but also minimize averaging and allow us to harness the precise geo-location information associated with Tweets.

To divide the city into grids we used available latitude/longitude coordinates of the sourced Tweets (which have 8 decimal places). By systematically rounding off each latitude/longitude pair to a lower



Table 2. Result of dividing NYC bounding box into grids and rounding coordinates to selected precision. The area covered by each grid cell for 3 decimal places is 9364 m<sup>2</sup>, which is similar to the area of 1 block in NYC (1 block is the area enclosed between 2 streets and 2 avenues  $\approx$  11,100 m<sup>2</sup>).

Decimal Places	Number of Grids	Area of Each Grid Cell (in m <sup>2</sup> )	Avg. Tweets per Grid		
			Total	Racism	Homo.
1	35	9.367x10 <sup>7</sup>	178,136.143	1738.629	3135.200
2	2,206	9.365x10 <sup>5</sup>	2,826.276	27.585	49.743
<b>3</b>	<b>150,385</b>	<b>9.364x10<sup>3</sup></b>	<b>41.459</b>	<b>0.405</b>	<b>0.730</b>
4	2,796,484	9.364x10 <sup>1</sup>	2.230	0.022	0.039
5	10,620,968	9.364x10 <sup>-1</sup>	0.587	0.006	0.010
6	14,466,241	9.364x10 <sup>-3</sup>	0.431	0.004	0.008
7	14,768,312	9.364x10 <sup>-5</sup>	0.422	0.004	0.007

precision (7 to 1 decimal place), we divided the city into grid cells at each resolution. For example, a Tweet with co-ordinates 40.83470082, -73.92287411 is rounded off from (40.8347008, -73.9228741) to (40.8, -73.9). Based on these coordinates, the example Tweet is mapped to the corresponding grid cell. Table 2 shows the results by precision level.

From the grid count and average Tweet count per grid (Table 2), we selected 3 decimal point precision, as this offered a balance between a substantial number of grids (using more grid points significantly increased computation time for analyses), while still representing a reasonably local geographic area and high enough proportion of racism/homophobic Tweets per grid cell. Later (in section 4.3), we show that this also resulted in a close relation between the number of SS-SOMs and Zip codes, which was important so that any findings related to cluster consistency are related to the actual clusters and not increased spatial averaging due to differences in the average size/number of clusters. We then removed grid cells that were not within the strict NYC boundaries via borough boundary shape files<sup>1</sup>, leaving 72,484 grids.

We highlight that while the normalized values by grid account for an unequal distribution of Tweets by location, it would be insufficient to use these grids alone to assess the prevalence of racism or homophobia by location due to an unequal distribution of Tweets per grid cell. Further, we note that we also tested a user-centric versus Tweet-centric approach to both grids and clusters, to assess if “loud” users may be skewing results. The results showed that there could be some loud users, but they are not changing results based on content. Pearson’s correlation was 0.87 for racism and 0.91 for homophobia, between user-centric and Tweet-centric formation of grids. Cluster similarity using user-centric and Tweet-centric binning was 0.93 and 0.95 for racism and homophobia respectively. Thus all further analyses use a Tweet-centric approach in order to capture the impact of all of the *content*.

**4.2.2 Tweeting Levels over Time.** As we controlled for the Tweeting levels over space by dividing the city into grid, we also assessed if Tweeting levels change over time and if that may influence results. We assessed the proportion of homophobia or racism Tweets each month using student’s t-test to assess if there was any significant difference between each two adjoining months’ normalized homophobia and racism values. Results of the tests showed no significant difference (p-values ranged from 0.54-0.91) between any adjacent months, indicating no major underlying differences

<sup>1</sup> <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm/data>

by month. Hence, classification results from each of the included months are similar and results would not differ by time.

### 4.3 Socio-Spatial Self-Organizing Map (SS-SOM)

Self-organizing Maps (SOMs) are frequently used in geographical clustering, the main difference being that the constraint of contiguous clusters is not imposed. The broad approach for implementing any version of SOMs remains fairly similar. Details of the SOM algorithm can be found elsewhere [6, 47], and here we focus on the specifics and unique aspects of SS-SOMs. Implementation can broadly be divided into 2 stages: (1) initialization, and (2) organization.

- **Initialization:** Similar to all SOMs, SS-SOM require an input vector (of weights, which in our case is the proportion of Tweets), output vector (initially consists of just the geographic locations of the nodes, and eventually to have linked information about the weights), and multiple operational parameters. Operational parameters are: learning rate (rate at which winning nodes, the final centers of each cluster, approach their destination cluster), number of learning cycles, and neighborhood radius which describes how “far away” in terms of weight, adjoining grid cells are to be considered in each learning cycle and in our case is adapted to include geographic distance also. The modification of constraining the neighborhood radius is operationalized by specifying the distance at which nodes are considered in re-weighting of the output nodes, to ensure geographical adjacency.
- **Organization:** Organization of the SS-SOM is similar to regular SOM except the neighborhood radius is augmented to include surrounding nodes that have similar weights in terms of both Tweet proportions (semantic similarity) and topological distance (geographic similarity).

*Detailed description of SS-SOM.* We define the following input variables for the SS-SOM algorithm:  $g$ , the number of grid cells, and  $\mathbf{I}$ , an input vector space, where length of  $\mathbf{I} = g$ . The operational parameters help us specify how the SS-SOM algorithm functions:  $t_{max}$  is the number of learning cycles (50),  $\eta_0$  is the learning rate at time 0,  $\eta_0 \approx 0.1$ ,  $\eta_0 \in Q^+$  and  $\eta(t)$  is the learning rate at time  $t$ :  $\eta(t) = \eta_0 \cdot e^{-\frac{t}{t_{max}}}$ ,  $\eta(t) \in Q^+$ . The operational parameters are set in line with previous work [6]. Additionally, we performed a sensitivity analysis for the number of learning cycles, and found that the learning rate plateaus as we near 50 cycles (the chosen value), after which the number of resulting clusters also plateaus. Later, we found that this resulted in a number of clusters that was most comparable to the number of Zip codes at the chosen threshold value  $\tau$  (Figure 3(b)). And, since the average resulting area of SS-SOM’s with this threshold and number of learning cycles were  $8.39 \times 10^6 \text{ m}^2$  for racism and  $7.73 \times 10^6 \text{ m}^2$  for homophobia, which are very similar compared to  $3.69 \times 10^6$  for Zip codes, we deemed these settings appropriate. Moreover, the average number of Tweets in each SS-SOM for racism was 647.36 (0.98%) and for homophobia was 1075.8 (1.76%), which was very similar to Zip codes (284.36 (0.98%) and 512.77 (1.76%) respectively). Hence, resulting comparisons would only be due to homogeneity within the clusters and not because of a changing number or changing area of clusters.

Each output node has two attributes, fixed position and weight. While the fixed position is a constant attribute that helps in determining the initial position of the node in the grid system (NYC map), the weight of the node is a varying attribute and determines how the clustering takes place based on the total and racism/homophobia Tweet counts.  $S$  is a set of output nodes to be created:  $|S| = g$ ,  $n_i$  is a node from the set of output nodes:  $n_i \in S$ . Then,  $\forall n_i$ ,  $p_i$  is the position of node  $n_i$  (the coordinates of the node (grid cell’s) location on the map),  $p_i \in (\text{longitude, latitude})$ . And,  $w_i$  = the weight of node  $n_i$ , is initialized using random values of any range,  $\dim(w_i) = \dim(\mathbf{I})$ .

We modify the traditional SOM (discussed in section 5.2.2) to ensure we create topologically-constrained and non-overlapping clusters by defining the input vector space  $\mathbf{I}$  with two vectors:

number of Tweets in a grid cell, and number of racism/homophobia Tweets. Hence,  $\dim(\mathbf{I}) = 2$ , and correspondingly,  $\dim(\mathbf{w}_i) = 2$ . We also define the following operational parameters specifically for the modified version of the SOM:  $\phi$  is a cluster ID assigned to each grid cell that uniquely identifies the cluster that a grid cell belongs to,  $\phi \in N$ . The threshold  $\tau$  is a fixed neighborhood radius,  $\tau \in N$ . This parameter is similar to  $\sigma(t)$  (neighborhood radius) used in the traditional version of SOM. The difference is that  $\tau$  is modified to be a constant to ensure that the grid cells that are part of a cluster are adjacent to each other geographically.

**4.3.1 Determining the winning node.** “Winning nodes” are nodes assigned to a specific cluster in each learning cycle based on spatial and social attitude proximity. Therefore, a winning node,  $n_{win} \in S$ , is defined as a node that: (a) is socially (based on prevalence of the social process) closest to the input vector  $v$ . In other words, among all the output nodes within the threshold  $\tau$  of  $n_{win}$ , the difference between the weights of the winning node and the input vector  $v$  is the least, and (b) for nodes that are equally close socially, the cluster  $\phi$  that surrounds  $n_{win}$  the most topologically (cluster  $\phi$  with the maximum number of grid cells in radius  $\tau$  of  $n_{win}$ ), will influence  $n_{win}$  most.

To achieve this, at each iteration  $t \in (1, t_{max})$  we create a subset of grid cells,  $\mu$ , that are at a distance  $\tau$  from  $n_{win}$ . Then, each grid cell  $\mu_j$  that belongs to a cluster  $\phi$  is denoted by  $\mu_{j\phi}$ . Accordingly, the weight of  $\mu_j$ , denoted by  $\omega_j$ , is the same as the weight of all the grid cells in the same cluster,  $\mu_{j\phi}$ , and the number of grid cells in cluster  $\phi$  is  $|\mu_{j\phi}|$ . Hence, as  $\mu \subseteq S$ ,  $\forall \mu_j \in \mu$ ,  $\mu_{win}$  is the same as  $n_{win}$ . Therefore,  $\mu_{win}$  must satisfy the condition:

$$\forall \mu_j \in \mu, ||\omega_{win} - v|| |\mu_{j\phi}| \leq ||\omega_j - v|| |\mu_{j\phi}| \quad (1)$$

The difference is weighted by the size of the cluster in the radius  $\tau$  ( $|\mu_{j\phi}|$ ) to maintain topological similarity. Then for SS-SOM, the best matching unit,  $\text{BMU}(v)$ , is calculated using the same formula as in traditional SOM, but by replacing the distance between a node  $w_i$  and the input vector  $v$  with the distance weighted by the size of the cluster in radius  $\tau$ .

**4.3.2 Weight Adaptation.** In SS-SOMs, the weight adaptation step in traditional SOMs has been modified in accordance with the aim of giving equal importance to geographical as well as social distance. While the threshold  $\tau$  and the BMU ensure we select adjacent grid cells,  $\omega_i(t+1)$  is adjusted to also take the proportion of a social process in cluster  $\mu_{j\phi}(t)$  into consideration. Additionally, the neighborhood function,  $h(d, t)$  is modified [3, 6, 45] to ensure that the weight of each grid cell in the cluster, and the potential new grid cell are all considered:  $h(d, t) = e^{-\frac{d}{|\mu_{j\phi}(t)|+1}}$ .

Finally, the weight of each output node is adapted as in the original SOM case, for  $t \in (1, t_{max})$ :

$$\omega_j(t+1) = \omega_j(t) + (\eta(t) h(|\mu_{win} - \mu_j|, t) [v_j - \omega_j(t)]) \quad (2)$$

## 5 COMPARISON OF PIPELINE TO BASELINES

### 5.1 Tweet Classification

First, we validate the neural net text classification in the SS-SOM pipeline by comparing its performance to a baseline linear classifier (Support Vector Machines) which has been demonstrated as the most robust linear classifier for text classification [41].

**5.1.1 Support Vector Machines.** Grid search was used to optimize hyper-parameters of the classifier selecting from a radial basis function kernel (gamma: 1e-3 to 1e-4) and linear kernels each with C values of {1, 10, 100, 1000} [41]. Features (n-gram) used for the grid search ranged from one to six words, half the number of average words per Tweet [34], coherent for comparison with the neural network approach.

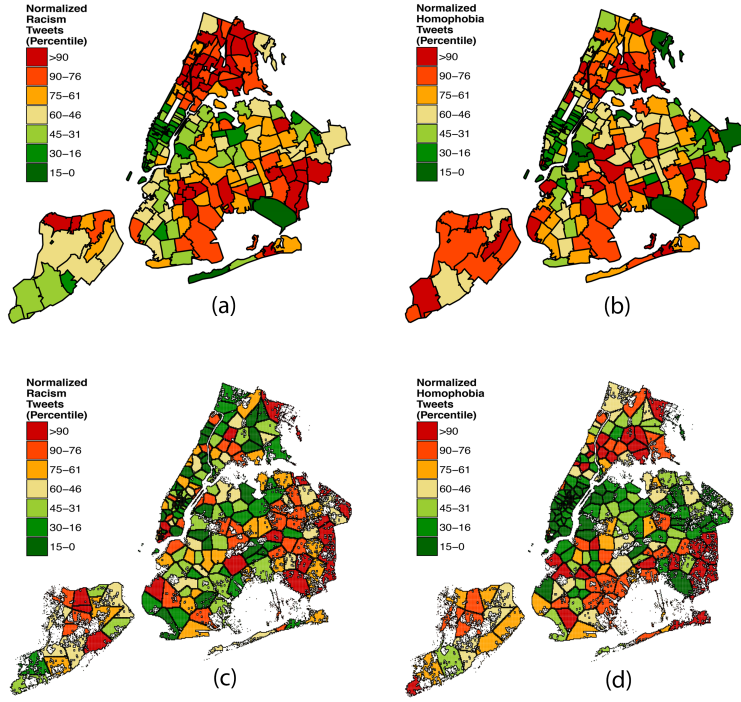


Fig. 2. (a) Map of NYC depicting prevalence of social media measured racism by Zip code; (b) homophobia by Zip code; (c) racism by SS-SOM with 94 clusters (threshold of 3); (d) homophobia by SS-SOM with 102 clusters (threshold of 3). It is clearly visible that areas of high/low exposure are resolved differently using SS-SOMs versus Zip codes. The white-areas in (c) and (d) within the city represent grid cells that had 0 Tweets, hence normalization for such grid cells was not possible.

## 5.2 Administrative Boundaries and Clustering Methods

We primarily compare the performance of resulting SS-SOM clusters with 2 baselines: (1) administratively defined Zip codes and (2) traditional version of SOM. We also discuss results of clustering by the Geo-SOM method [3].

**5.2.1 Zip codes.** Zip codes are generally used to quantify environmental exposures in health-related studies. To determine the prevalence of racism and homophobia in each Zip code, we mapped each grid cell to a Zip code using the NYC Zip code shapefile<sup>2</sup>. For each Zip code, we normalized the number of racism/homophobia Tweets with total number of Tweets in that Zip code. Figures 2(a) and 2(b) depict the prevalence of racism and homophobia Tweets by Zip codes.

**5.2.2 Traditional Self-Organizing Maps.** We define the following input variables (same as in SS-SOM):  $g$ , the number of grid cells; and  $I$ , an input vector space having four vectors: number of Tweets in a grid cell, number of racism/homophobia Tweets, and the latitude and the longitude of the grid [3]; length of  $I = g$ . The operational parameters used are in line with previous work [6]. Implementation led to 3622 clusters for racism, and 2899 clusters for homophobia. We used a maximum  $t_{max}$  of 17500 as used in [32] thus decreasing the number of clusters by only 20% to 2899

<sup>2</sup>[https://www.census.gov/geo/maps-data/data/cbf/cbf\\_zcta.html](https://www.census.gov/geo/maps-data/data/cbf/cbf_zcta.html)

for racism and 2328 for homophobia which was still well above the number from SS-SOM which represent a number of geographic areas of exposure that are more comparable to Zip codes.

**5.2.3 Geo-SOM.** The other closest baseline algorithm we can compare to is the Geo-SOM algorithm [3]. The aim of this work, to develop homogeneous regions and perform geographic pattern detection by taking into account spatial dependency along with social dependency, was achieved by changing the neighborhood function to search for BMU only within a predefined topological vicinity. While such an approach guarantees homogeneous clusters, we implemented this algorithm using the software (Geo-SOM Suite)<sup>3</sup> to confirm that it does not fulfill our constraints regarding generation of contiguous and non-overlapping partitioning of the city and hence, we do not further evaluate this baseline with SS-SOM.

## 6 EVALUATION

First, we evaluate the text classification performance (for identifying racism or homophobia Tweets). Then, we evaluate clustering quality of the SS-SOMs. Finally, we evaluate how meaningful resulting SS-SOMs are by examining how online homophobia/racism exposure measured through SS-SOMs differs from what would be measured through Zip codes, and qualitatively examining the resulting prevalence of racism and homophobia.

### 6.1 Classification Performance

To evaluate performance of racism and homophobia classification, we compare the results of neural network classification with our baseline (SVM) based on the F1 score and Receiver operating characteristic (ROC) curve from each method.

### 6.2 Cluster Quality

We evaluate implementations that result in clusters that geographically segment NYC (SS-SOM, Zip codes and SOM) along the following 4 axes related to the cluster quality: (1) **Similarity** indices to quantify how grid cells land in clusters across methods, thus conveying how different the overall clustering in NYC is by SS-SOM and the baseline methods. (2) **Robustness of results to missing data** for each method. Specific to the application here, missing data may arise from an under-represented area on social media or inconsistent data from Twitter API. Accordingly, we evaluate sensitivity to missing data in two different ways: (i) missing Tweets and (ii) missing grids. (3) **Mean variance** of social processes across grids in resulting clusters. This directly assesses homogeneity of the resulting clusters; how internally similar the resulting clusters are. (4) **Mean Squared Prediction Error (MSPE)** is used to assesses reliability of the clusters by calculating the mean squared difference of the prevalence of racism/homophobia between clusters formed using complete versus held-out data.

**6.2.1 Similarity of SS-SOM to Baselines.** As using the SS-SOM method results in a different number of clusters compared to Zip codes and traditional SOM method (and also depends on the threshold selected for the SOMs), we compare the clustering between methods using the similarity index  $c2(Y, Y')$  [58]. This is the ratio of the number of similar assignments of (grid) points to the total number of grid point pairs. The higher the value of  $c2$ , the more similar the resulting clusters from the comparing methods are. More precisely, given  $n$  points:  $x_1, x_2, \dots, x_n$ , and clustering results:

$Y = Y_1, \dots, Y_{k1}$  and  $Y' = Y'_1, \dots, Y'_{k2}$ ,  $c2(Y, Y') = \frac{\sum_{i < j}^N \gamma_{ij}}{\binom{N}{2}}$  where  $\gamma_{ij} = 1$  if there exist  $k$  and  $k'$  such that

<sup>3</sup><http://www.novaims.unl.pt/labnt/geosom/GeoSOM.suite.htm>



both  $x_i$  and  $x_j$  are in both  $y_k$  and  $y_{k'}$ , or if there exist  $k$  and  $k'$  such that both  $x_i$  is in both  $y_k$  and  $y_{k'}$ , and  $x_j$  is in neither  $y_k$  or  $y_{k'}$ , or 0 otherwise.

**6.2.2 Robustness to Missing Data.** For missing Tweets, we used a holdout range of  $n_1=\{25\%, 50\%$  and  $75\%\}$  of the total number of Tweets and assessed the resulting clusters formed by traditional SOM and SS-SOM. For missing grids, we used an approach from clustering evaluation [58]. We compare traditional SOM and SS-SOM regarding missing grid sensitivity, as Zip code boundaries don't change if data is missing. We first perform the clustering using a subset of  $n_1$  grids. Then, we implement the same clustering methods on additional  $n_2 = n - n_1$  grids (for a total  $n = n_1 + n_2$  grids). We compare how the original  $n_1$  grids are clustered using the larger data set ( $n$ ) using a  $k$ -fold cross validation ( $k=10$ ) for  $n_1$  (and  $n = 72,484$  grid cells) and similarity index  $c2$ .

**6.2.3 Variance of Social Processes by Cluster.** Mean variance of grid proportions in each cluster is:  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  where  $n$  is the number of grid cells in a cluster,  $x_i$  is the normalized homophobia/racism count of the  $i$ th grid cell and  $\bar{x}$  the normalized homophobia/racism Tweet count in the entire cluster.

**6.2.4 Mean Squared Prediction Error (MSPE).** MSPE for SS-SOM is calculated by holding out 10%, 25%, 50% and 75% of data (we performed this for both missing grid cells, and missing Tweets separately). We re-learn the spatial model for each size of the data, disregarding the “weight” for the held out grid cells, and compute the prevalence of racism/homophobia for the newly formed clusters. We then compute the MSPE using the prevalence of racism/homophobia in the clusters each of those grid cells is assigned to in each case above. We do a  $k$ -fold cross-validation ( $k=10$ ) to prevent over-fitting. MSPE is computed by:  $E \left[ \sum_{i=1}^n (g(x_i) - \hat{g}(x_i))^2 \right]$  where,  $n$  are the number of grid cells in the held out data,  $g(x_i)$  is the mean prevalence of the cluster grid cell  $i$  is assigned to in the full model, and  $\hat{g}(x_i)$  when it is held out.

### 6.3 Cohort's Exposure to Racism and Homophobia via SS-SOM vs. Zip codes

We calculated exposure differences ( $E_i$ ), for each person  $i$  in the P18 cohort by assessing the difference in proportion of racism/homophobia Tweets for each cell visited by individual  $i$ , as a proportion of all their visits.

$$E_i = \left( \sum_{k=1}^n \left( \frac{|E_{zip_k} - E_{SOM_k}|}{E_{zip_k}} * V_{i_k} \right) \right) / V_i \quad (3)$$

where,  $E_{zip_k}$  is the proportion of racism or homophobia Tweets in Zip code containing grid cell  $k$  and  $E_{SOM_k}$  is the proportion of racism or homophobia Tweets in SS-SOM cluster containing grid cell  $k$ .  $V_{i_k}$  is the number of visits made to grid cell  $k$  by person  $i$  and  $V_i$  is the total number of visits made in entire NYC by person  $i$ .

### 6.4 Prevalence and Co-prevalence of Racism and Homophobia Measured Through Social Media

We know methodologically that Zip codes will suffer from spatially averaging of the racism and homophobia sentiments, but will this result in differing prevalence of racism and homophobia online sentiment in NYC? Here we demonstrate the impact of results by qualitatively examining the prevalence of racism and homophobia sentiment through the Zip code and SS-SOM measures.

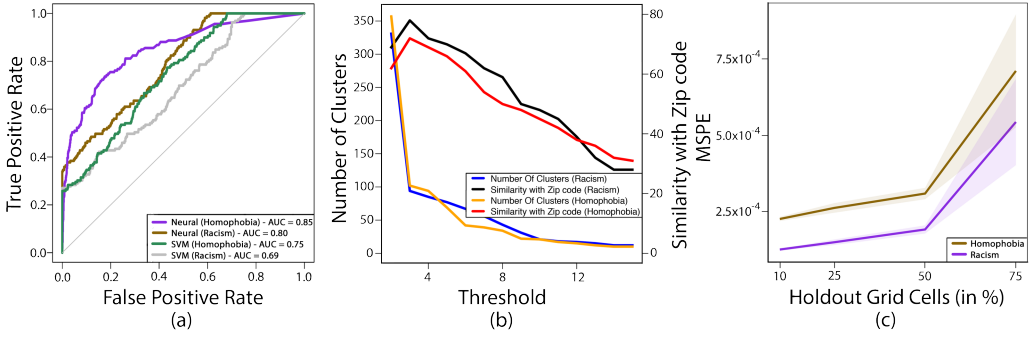


Fig. 3. (a) ROC curve for racism and homophobia classification using neural embedding and SVM. (b) Number of clusters and similarity between Zip codes and SS-SOM by threshold value. (c) Mean squared prediction error (MSPE) for SS-SOM (threshold of 3) across different proportions of held out grid cells (shaded region represents 95% confidence interval).

## 7 RESULTS

### 7.1 Classification Performance

Neural embedding for racism classification (F1 score: 0.89, Area Under Curve (AUC): 80) (Figure 3a) was slightly better than SVM (F1: 0.85, AUC: 0.69). This was also the case for homophobia classification; F1 scores of 0.89 and 0.86 respectively and AUC of 0.85 and 0.75 respectively [63]. We found that the neural embedding classifier yields a low false positive rate (precision for racism-0.89, homophobia-0.87) and low false negative rate (recall for racism-0.80, homophobia-0.84) on test data. Therefore, we can conclude that it is efficiently able to classify Tweets into racism versus no racism and homophobia versus no homophobia.

### 7.2 Cluster Quality

**7.2.1 Similarity of SS-SOM to Baselines.** For  $n=72,484$  grid cells, we find that Zip code and SS-SOM methods are most similar in terms of homophobia and racism by cluster (72% and 78% respectively). A comparison between homophobia and racism clusters can only be assessed via similarity metrics for the SS-SOM method (82%) showing a high similarity between clustering of the two processes. The traditional SOM is least similar to both Zip codes (25% and 28% for homophobia and racism respectively) and SS-SOM (34% and 36% for homophobia and racism respectively), due to the high number of clusters obtained (3622 clusters for racism, and 2899 clusters for homophobia).

Figure 3b shows  $c2$  versus threshold value. Similarity between SS-SOM and Zip codes for both homophobia and racism peaks at a threshold of 3 which also has a reasonable balance between amount of Tweets per cluster and computation time (as discussed in section 4.3). Thus all further analyses related to SS-SOM use a threshold of three.

**7.2.2 Sensitivity to Missing Data.** Missing Tweets resulted in no effect in both SOM and SS-SOM. This is because for each grid, we used a normalized proportion of Tweets. As racism/homophobia Tweets are just as likely to be missing as any other Tweets, normalization accounts for missing data. For missing grid cells, a similarity ( $c2$ ) of up to 95.3% and 99.3% (SOM and SS-SOM respectively) between the clustering from the original  $n$  and  $n_1$  grid cells. Overall, for a 10-fold cross validation for 25%, 50% and 75% of missing grid cells, the average similarity for traditional SOM was 95.3%, 89.2% and 82.8% respectively (standard deviation up to 1.2%), and the average similarity for SS-SOM

Table 3. Mean variance of social process by clustering technique. Traditional SOM shows the lowest variance for racism and homophobic-specific clusters.

Clustering	Racism Mean Variance (Min, Max)	Homophobia Mean Variance (Min, Max)
<b>Zip code</b>	$1.1 \times 10^{-3}$ (0,0.029)	$1.7 \times 10^{-3}$ (0,0.013)
<b>SOM</b>	$9.1 \times 10^{-7}$ ( $2.3 \times 10^{-8}$ , $7.2 \times 10^{-4}$ )	$6.2 \times 10^{-7}$ ( $8.3 \times 10^{-9}$ , $1.4 \times 10^{-5}$ )
<b>SS-SOM</b>	$9.4 \times 10^{-4}$ ( $1.6 \times 10^{-6}$ , $6.3 \times 10^{-3}$ )	$2.9 \times 10^{-3}$ ( $4.4 \times 10^{-6}$ , $1.4 \times 10^{-3}$ )

was 99.3%, 93.8% and 85.8% (standard deviation of up to 1.3%) which indicate that SS-SOMs are generally stable with respect to missing data.

**7.2.3 Variance of Social Processes by Cluster.** Traditional SOM and SS-SOM had lower mean variance for racism clusters than Zip codes (Table 3), and all methods were similar for homophobia clusters. Notably, minimums for both racism and homophobia by Zip codes are zero for Zip code 10470 (northern Bronx). This is because only 85 tweets from the entire data set were located in this Zip code, and none were racist or homophobic. Both SS-SOM and traditional SOM methods create a more consistent set of clusters than Zip codes (in terms of variance), thus improving on Zip codes for creating social-process segmented regions. Though the traditional SOM had the least variance leading to more homogeneous clusters, there are two reasons which make the use of traditional SOM not feasible: (a) the high number of clusters that the city gets divided into, and (b) the clusters are not topologically constrained and also overlap, which are basic requirements for the problem.

**7.2.4 Mean Squared Prediction Error (MSPE).** For grid cells, MSPE values increased from  $1.25 \times 10^{-4}$  to  $5.4 \times 10^{-4}$  (racism) and  $2.3 \times 10^{-4}$  to  $7.1 \times 10^{-4}$  (homophobia) from 10% to 75% of grid cells held out (Figure 3(c)). The low MSPE values suggest that the clusters created by SS-SOM are homogeneous. The MSPE for SS-SOM after holding out Tweets instead of grid cells is nearly zero (maximum MSPE is  $7.7 \times 10^{-5}$  for 75% holdout Tweets) across all hold-out sizes; this is because SS-SOMs are robust to missing Tweets due to normalization. Finally, MSPE of traditional SOM and Zip codes are not calculated as the former doesn't meet our clustering requirements as discussed at the end of section 7.2.3 and the boundaries of the latter don't change.

### 7.3 Cohort's Exposure to Online Racism and Homophobia

There was a clear difference in the cohort's measured exposure to online racism and homophobia for SS-SOM versus Zip codes. Mean racism exposure difference was 41.3% (SD: 17.8%). For 35% (78/226) of individuals, the difference was over 50%. Mean homophobia exposure difference was 29.6% (SD: 18.7%). Qualitatively we can identify places with large differences in exposure that would be important based on how many times they are visited; a grid cell which had 71,903 visits in upper Manhattan (near the Bronx border) had the highest difference in racism Tweet prevalence through Zip codes (2.2%) and SS-SOM (0.7%). A grid cell (which had 109,876 visits) near the City Hall in Lower Manhattan had the highest difference in homophobia Tweet prevalence through Zip codes (3.0%) and SS-SOM (1.0%).

### 7.4 Prevalence and Co-prevalence of Racism and Homophobia Measured Through Social Media

The maps in Figure 2 show prevalence of racism and homophobia in Zip codes and SS-SOM clusters. For Zip codes, the proportion of racism Tweets in each Zip code ranged from 0% to 3.60% and the proportion of homophobia Tweets ranged from 0% to 3.20%. For the SS-SOM clusters, the proportion

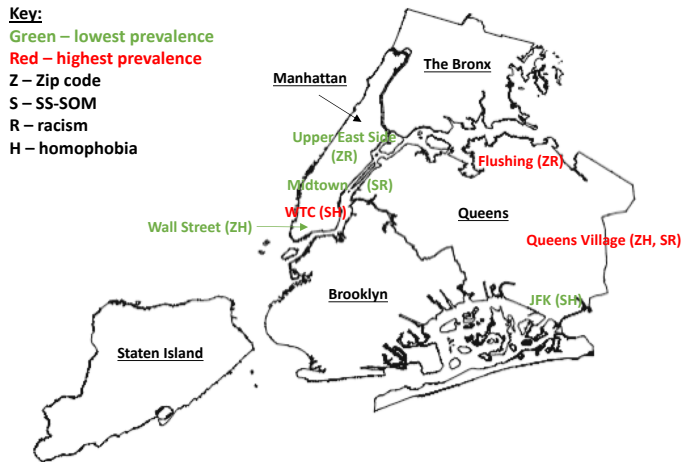


Fig. 4. NYC map displaying areas of online homophobia/racism prevalence for geographic reference.

of racism Tweets in each cluster ranged from 0.28% to 1.23% and the proportion of homophobia Tweets ranged from 0.09% to 2.92%. An important observation about the ranges for proportion by Zip code is that extreme values result for both racism and homophobia due to a very few number of Tweets being made in some Zip codes. For example, the Zip code with the lowest prevalence of racism just had 306 Tweets, and the Zip code with the lowest prevalence of homophobia just had 71 Tweets. The SS-SOM cluster with the lowest prevalence of racism had 100,774 Tweets and the SS-SOM cluster with the lowest prevalence of homophobia had 16,365 Tweets.

Qualitatively, we see that differences from these methods result in some overall similarities and differences. For racism, both Zip codes and SS-SOMs show high prevalence in areas of Staten Island and Queens. However, one striking difference is in the Bronx where most of the Zip codes show high prevalence, but many of the SS-SOMs show low prevalence. We examined the proportion of racist Tweets in this area and found that many Zip codes appear to have a high proportion of racism when it is mostly concentrated in one or two grid cells which can be adjacent to grid cells with much lower prevalence. These grids get split into different SS-SOMs, which internally will be more consistent, and overall the proportion in all of those SS-SOMs will be lower. Thus, the proportion of racism in Zip codes is being mis-represented due to a high proportion in one or two grid cells. Similarly for homophobia, the neighborhood with highest homophobia online via Zip codes is in Queens Village, however for SS-SOM it is near the World Trade Center (Figure 4), due to a similar spatial averaging effect. Both methods show a relatively low level of racism and homophobia in Manhattan, though SS-SOMs define some areas with higher racism in Manhattan.

Comparing the SS-SOM racism and homophobia clusters, we found a similarity of 82% (as discussed in 7.2.1). Some qualitative differences were in northern Staten Island (Mariners Harbor) (higher racism but below average homophobia) and near the world trade center (WTC) which had higher homophobia but below average racism. Overall we found that there were a lot of clusters that ranked low in racism but were very high in homophobia (like Rockaway Park Queens, Lower Manhattan near the World Trade Center). This indicates that the prevalence of online racism is generally accompanied by prevalence of homophobia but the opposite is not true; areas may have higher levels of homophobia independent of racism. This also holds true partially due to a higher prevalence of homophobia on Twitter in NYC as compared to racism, hence prevalence of

homophobia may not always be accompanied by racism. We also analyzed the co-prevalence of racism and homophobia at a user level and found the same result: only 2.5% of the users with a homophobia Tweet also had a racism one, but the reverse was 10%.

## 8 DISCUSSION AND CONCLUSION

### 8.1 Summary of Findings and Contributions

SS-SOMs are a simple modification of SOMs that address the need for partitioning a complete area into non-overlapping, contiguous and homogeneous areas that best represent the latent sentiment behind sparse and noisy Tweets, while accounting for varying social media levels in different places. This work adds to the social computing literature using novel data sources to augment our understanding of spatial regions [37, 39, 59]. Overall, we found that SS-SOMs show robustness to missing data, lower maximum variance in cluster content and less evidence of spatial averaging as compared to Zip codes. Mean variance of homophobia/racism in SS-SOM clusters was similar to that of traditional SOM (lower maximum variance) and lower than variance of Zip codes. Both SOM and SS-SOM had a narrower range of variances across all clusters compared to Zip codes, indicating more consistency. Thus, we see that the spatial averaging resulting by using Zip code measures of online racism and homophobia would be significant compared to SS-SOMs, for a real population for which racism and homophobia are relevant. We further found that SS-SOM had low MSPEs across a range of proportions of grid cells missing, indicating good reliability.

This work creates a novel approach to understand underlying sentiment in areas based on individual, sparse and noisy social media data. This is useful for topics that are difficult to measure in other ways, like racism or homophobia. It has been shown that people are vocal about such sensitive subjects on social media where they perhaps feel more comfortable than on surveys or other face-to-face situations [68]. Therefore, there are no “gold standard” methods that capture the same types of prejudice as found on social media (e.g. hate crime reports capture prejudice that results in crimes), and can be exactly compared and used to validate the findings here. Instead, social media measures of prejudice should be seen as a complement to other measures. The thorough qualitative examination of the identified areas of racism and homophobia, as well as assessment of resulting exposure to a real cohort, confirms the importance of developing these measures and ways to operationalize the use of the social media data and compare and integrate it with other measures of prejudice.

In comparison of SS-SOM and traditional SOM to Zip code, SS-SOM clusters were more similar to Zip code for both racism and homophobia compared to similarity between traditional SOM and Zip codes. This is not surprising given that the traditional SOM clusters are high in number as compared to Zip code. However, both methods had lower similarity with Zip codes for racism (compared to homophobia) indicating racism may have determinants more unrelated to Zip code-defined areas. We also found that an increase in racism was generally associated with homophobia, but not the opposite. Thus this work can also be used to develop studies to assess the reason behind levels or co-prevalence of these social processes.

Finally, we demonstrated via mobility of a cohort of MSM (for whom exposures to homophobia and racism are important health risk factors), that exposures measured via Zip codes and SS-SOM can be largely different (over 50% different for racism exposure in a third of the cohort). This is an important finding; exposure measures linked to Zip codes suffer from averaging whereas SS-SOM clusters may represent the underlying exposure more consistently; SS-SOM clusters that were high or low for one social process were in regions that had average or opposite levels based on Zip code or traditional SOM definitions.



## 8.2 Limitations and Future Work

An inherent limitation in this work is the use of the Twitter API. Though results are robust to missing data so amount of data should not be an issue, any inherent biases in the API sample (e.g. by content, location, language) would affect the results. As well, augmented text processing methods such as including sentiment could also be explored in order to more optimally encompass the diversity of human language; we used a fairly simple approach to text classification via n-grams.

An important next step should assess how the sentiment measured through the online environmental measures connects to the offline world (it should be noted this is not the same as linking a Tweet to the *home* location of a user – doesn't exactly correspond to the "localness assumption" [42]). As discussed, the literature indicates there is a need for better and new racism and homophobia measures, which motivates this work. In some way, relationships between online data and outcomes have been demonstrated by work examining links between area-based measures of racial animus from Google searches and black mortality outcomes [9] or differential votes for a black candidate [62]. As well, this assumption is used in many studies which harness geo-location of social media and offline attributes such as in assessing place semantics or sound maps [1, 59]. However a systematic study for our measures and health outcomes, such as through using surveys (which are currently not available for such social processes) could be performed. As described above, racism/homophobia measures from social media may be different from self-reported racism sentiment or experience on surveys, so this study would have to be carefully designed, perhaps using focus groups with select social media users. Further analytic work can be done to tease apart the types of racism/homophobia online (e.g. expressions versus reports of encounters), which may have different spatial patterns. As well, since some people are more vocal in expressing than others, a study should examine if "louder" users on Twitter correspond to an increase in offline sentiment.

As social media has high temporal resolution and is available in real-time, future efforts could also explore how clusters could be monitored over time to assess changes in exposures. Although we specifically selected NYC for this study based on the P18 data, and the understanding that racism and homophobia can be important social determinants for groups like MSM in urban areas like NYC, the generalizability and utility of this work in other cities and non-urban areas should be studied. Finally, our finding that the prevalence of online racism is generally accompanied by prevalence of homophobia but the opposite is not true should be investigated for its sociological underpinnings.

## 8.3 Implications and Design of Social Computing Systems

The SS-SOM pipeline can be of use for researchers and practitioners when exploring the geographical properties of social processes, especially those that may uniquely be communicated over social media. In order to facilitate appropriate interventions, many studies have been done to understand the prevalence of hate speech in the online world [4, 7, 10, 60]. From this work, areas defined by specific social topics can be used to assess relevant social influence or context for an individual or to control (normalize) for their social environment. System design, actively deploying interventions, or generation of feedback based on an individual's social environment can also be better focused based on a contextual measure of neighborhood. For example, approaches to mitigating divisive sentiment at the right place and time can be targeted based on knowledge of spatial patterns of these social processes.

## A LIST OF KEYWORDS

**Racism keywords:** nigger, niggers, racism, racist, coon, ape, abid, abeed, ali baba, alligator bait, beaner, beaney, bootlip, buffie, burrhead, ching chong, chinaman, abcd, coolie, groid, haji, hajji, sooty, spic, spook, tar-baby, toad, wigger, “white nigger”, white trash, “white trash”.

**Homophobia keywords:** “batty boy”, dyke, faggot, “fag bomb”, “gay mafia”, “gold star gay”, “gold star lesbian”, “no homo”, queer, twink, “kitty puncher”, “brownie king”, twink, gaysian, “gym bunny”, daffy.

## B AMT INSTRUCTIONS

The racism labelling task instructions are given as an example, homophobia instructions were the same with appropriate racism/homophobia changes: “We are interested in labeling Tweets that exhibit racism or exposure to it. So, please label a Tweet as positive only if: 1) it exhibits a prejudice against any race (e.g. black, hispanic, asian, etc.), 2) it has a negative/derogatory remark about any race (e.g. black, hispanic, asian, etc.), and 3) it indicates that the person Tweeting was exposed to racism or observed someone being exposed to racism.” These instructions were followed by examples of racism and non-racism Tweets and explicit discussion of special cases (e.g. song lyrics or colloquial terms); through iteration we found specific direction about these was needed.

## ACKNOWLEDGMENTS

The authors acknowledge support by the National Institutes of Health and National Institute of Mental Health (NIMH) under grant R21 MH110190, and by the National Science Foundation under award MRI-1229185.

## REFERENCES

- [1] Luca Maria Aiello, Rossano Schifanella, Daniele Quercia, and Francesco Aletta. 2016. Chatty maps: constructing sound maps of urban areas from social media data. *Open Science* 3, 3 (2016), 150690.
- [2] Allison C Aosved, Patricia J Long, and Emily K Voller. 2009. Measuring sexism, racism, sexual prejudice, ageism, classism, and religious intolerance: The intolerant schema measure. *Journal of Applied Social Psychology* 39, 10 (2009), 2321–2354.
- [3] Fernando Bação, Victor Lobo, and Marco Painho. 2004. Geo-self-organizing map (Geo-SOM) for building and exploring homogeneous regions. *Geographic Information Science* (2004), 22–37.
- [4] Jamie Bartlett, Jeremy Reffin, Noelle Rumball, and Sarah Williamson. 2014. Anti-social media. *Demos* (2014), 1–51.
- [5] Jacob Bor, Atheendar S Venkataramani, David R Williams, and Alexander C Tsai. 2018. Police killings and their spillover effects on the mental health of black Americans: a population-based, quasi-experimental study. *The Lancet* (2018).
- [6] Christoph Brauer. 2012. An Introduction to Self-Organizing Maps. (2012).
- [7] Pete Burnap and Matthew L Williams. 2015. Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy & Internet* 7, 2 (2015), 223–242.
- [8] Richard M Carpiano, Brian C Kelly, Adam Easterbrook, and Jeffrey T Parsons. 2011. Community and drug use among gay men: The role of neighborhoods and networks. *Journal of Health and Social Behavior* 52, 1 (2011), 74–90.
- [9] David H Chae, Sean Clouston, Mark L Hatzenbuehler, Michael R Kramer, Hannah LF Cooper, Saboby M Wilson, Seth I Stephens-Davidowitz, Robert S Gold, and Bruce G Link. 2015. Association between an internet-based measure of area racism and black mortality. *PloS one* 10, 4 (2015), e0122963.
- [10] Irfan Chaudhry. 2015. # Hashtagging hate: Using Twitter to track racism online. *First Monday* 20, 2 (2015).
- [11] Kyung-Hee Choi, Chong-suk Han, Jay Paul, and George Ayala. 2011. Strategies of managing racism and homophobia among US ethnic and racial minority men who have sex with men. *AIDS education and prevention: official publication of the International Society for AIDS Education* 23, 2 (2011), 145.
- [12] Rumi Chunara, Lindsay Bouton, John W Ayers, and John S Brownstein. 2013. Assessing the online social environment for surveillance of obesity prevalence. *PloS one* 8, 4 (2013), e61373.
- [13] Rumi Chunara, Lauren E Wisk, and Elissa R Weitzman. 2017. Denominator issues for personally generated data in population health monitoring. *American journal of preventive medicine* 52, 4 (2017), 549–553.

- [14] Rodney Clark, Norman B Anderson, Vernessa R Clark, and David R Williams. 1999. Racism as a stressor for African Americans: A biopsychosocial model. *American psychologist* 54, 10 (1999), 805.
- [15] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 1 (1960), 37–46.
- [16] Dana Collins. 2009. “We’re There and Queer” Homonormative Mobility and Lived Experience among Gay Expatriates in Manila. *Gender & Society* 23, 4 (2009), 465–493.
- [17] Angelo Brandelli Costa, Denise Ruschel Bandeira, and Henrique Caetano Nardi. 2013. Systematic review of instruments measuring homophobia and related constructs. *Journal of Applied Social Psychology* 43, 6 (2013), 1324–1332.
- [18] Justin Cranshaw, Raz Schwartz, Jason I Hong, and Norman Sadeh. 2012. The livelihoods project: Utilizing social media to understand the dynamics of a city. In *ICWSM*. 58.
- [19] Justin Cranshaw and Tae Yano. 2010. Seeing a home away from the home: Distilling proto-neighborhoods from incidental data with latent topic modeling. In *CSSWC Workshop at NIPS*, Vol. 10.
- [20] William A Darity Jr. 2003. Employment discrimination, segregation, and health. *American Journal of Public Health* 93, 2 (2003), 226–231.
- [21] Thomas Davidson, Dana Warmley, Michael Macy, and Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. *arXiv preprint arXiv:1703.04009* (2017).
- [22] Munmun De Choudhury, Shagun Jhaver, Benjamin Sugar, and Ingmar Weber. 2016. Social Media Participation in an Activist Movement for Racial Equality.. In *ICWSM*. 92–101.
- [23] Da Deng and Nikola Kasabov. 2000. ESOM: An algorithm to evolve self-organizing maps from online data streams. In *Proc. IJCNN*, Vol. 6. IEEE, 3–8.
- [24] Dong-Po Deng, Tyng-Ruey Chuang, and Rob Lemmens. 2009. Conceptualization of place via spatial clustering and co-occurrence analysis. In *Proc. LBSN*. ACM, 49–56.
- [25] Cícero Nogueira Dos Santos and Maira Gatti. 2014. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts.. In *COLING*. 69–78.
- [26] Dustin T Duncan, Ichiro Kawachi, SV Subramanian, Jared Aldstadt, Steven J Melly, and David R Williams. 2013. Examination of how neighborhood definition influences measurements of youths’ access to tobacco retailers: a methodological note on spatial misclassification. *American journal of epidemiology* 179, 3 (2013), 373–381.
- [27] James E Egan, Victoria Frye, Steven P Kurtz, Carl Latkin, Minxing Chen, Karin Tobin, Cui Yang, and Beryl A Koblin. 2011. Migration, neighborhoods, and networks: approaches to understanding how urban environmental conditions affect syndemic adverse health outcomes among gay, bisexual and other men who have sex with men. *AIDS and Behavior* 15, 1 (2011), 35–50.
- [28] Tim Finin, Will Murnane, Anand Karandikar, Nicholas Keller, Justin Martineau, and Mark Dredze. 2010. Annotating named entities in Twitter data with crowdsourcing. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*. Association for Computational Linguistics, 80–88.
- [29] Vanessa Frias-Martinez, Victor Soto, Heath Hohwald, and Enrique Frias-Martinez. 2012. Characterizing urban landscapes using geolocated tweets. In *2012 PASSAT and SocialCom*. IEEE, 239–248.
- [30] Victoria Frye, Beryl Koblin, John Chin, John Beard, Shannon Blaney, Perry Halkitis, David Vlahov, and Sandro Galea. 2010. Neighborhood-level correlates of consistent condom use among men who have sex with men: a multi-level analysis. *AIDS and Behavior* 14, 4 (2010), 974–985.
- [31] George C Galster. 2008. Quantifying the effect of neighbourhood on individuals: Challenges, alternative approaches, and promising directions. *Schmollers jahrbuch* 128, 1 (2008), 7–48.
- [32] JL Giraudel and S Lek. 2001. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. *Ecological Modelling* 146, 1 (2001), 329–339.
- [33] Yoav Goldberg and Omer Levy. 2014. word2vec Explained: deriving Mikolov et al.’s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722* (2014).
- [34] Rachel Green and John Sheppard. 2013. Comparing Frequency-and Style-Based Features for Twitter Author Identification. In *FLAIRS Conference*.
- [35] Diansheng Guo, Mark Gahegan, Alan M MacEachren, and Biliang Zhou. 2005. Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach. *Cartography and Geographic Information Science* 32, 2 (2005), 113–132.
- [36] Perry N Halkitis and Rafael Perez Figueroa. 2013. Sociodemographic characteristics explain differences in unprotected sexual behavior among young HIV-negative gay, bisexual, and other YMSM in New York City. *AIDS patient care and STDs* 27, 3 (2013), 181–190.
- [37] Livia Hollenstein and Ross Purves. 2010. Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science* 2010, 1 (2010), 21–48.
- [38] Tom Huang, Anas Elghafari, Kunal Relia, and Rumi Chunara. 2017. High-resolution temporal representations of alcohol and tobacco behaviors from social media data. *Proceedings of the ACM on human-computer interaction* 1, CSCW

- (2017).
- [39] Suradej Intagorn and Kristina Lerman. 2011. Learning boundaries of vague places from noisy annotations. In *Proc. ACM SIGSPATIAL*. ACM, 425–428.
  - [40] Jennifer Jee-Lyn García and Mienah Zulfacar Sharif. 2015. Black lives matter: a commentary on racism and public health. *American journal of public health* 105, 8 (2015), e27–e30.
  - [41] Thorsten Joachims. 1998. Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*. Springer, 137–142.
  - [42] Isaac L Johnson, Subhasree Sengupta, Johannes Schöning, and Brent Hecht. 2016. The geography and importance of localness in geotagged social media. In *Proc. CHI*. ACM, 515–526.
  - [43] Malia Jones and Anne R Pebley. 2014. Redefining neighborhoods using common destinations: Social characteristics of activity spaces and home census tracts compared. *Demography* 51, 3 (2014), 727–752.
  - [44] Edson C Kitani, Emilio M Hernandez, Gilson A Giraldo, and Carlos E Thomaz. 2011. Exploring and Understanding the High Dimensional and Sparse Image Face Space: a Self-Organized Manifold Mapping. In *New Approaches to Characterization and Recognition of Faces*. InTech.
  - [45] Kimmo Kiviluoto. 1996. Topology preservation in self-organizing maps. In *IEEE Neural Networks*, Vol. 1. IEEE, 294–299.
  - [46] Beryl A Koblin, James E Egan, Andrew Rundle, James Quinn, Hong-Van Tieu, Magdalena Cerdá, Danielle C Ompad, Emily Greene, Donald R Hoover, and Victoria Frye. 2013. Methods to measure the impact of home, social, and sexual neighborhoods of urban gay, bisexual, and other men who have sex with men. *PloS one* 8, 10 (2013), e75878.
  - [47] Teuvo Kohonen. 1990. The self-organizing map. *Proc. IEEE* 78, 9 (1990), 1464–1480.
  - [48] Frauke Kreuter, Stanley Presser, and Roger Tourangeau. 2008. Social desirability bias in CATI, IVR, and Web surveys the effects of mode and question sensitivity. *Public Opinion Quarterly* 72, 5 (2008), 847–865.
  - [49] Géraud Le Falher, Aristides Gionis, and Michael Mathioudakis. 2015. Where is the Soho of Rome? Measures and algorithms for finding similar neighborhoods in cities. In *ICWSM*.
  - [50] Ryong Lee, Shoko Wakamiya, and Kazutoshi Sumiya. 2011. Discovery of unusual regional social activities using geo-tagged microblogs. *World Wide Web* 14, 4 (2011), 321–349.
  - [51] Cheng-Yuan Liou and Hsin-Chang Yang. 1996. Handprinted character recognition based on spatial topology distance measurement. *IEEE TPAMI* 18, 9 (1996), 941–945.
  - [52] Jason Liu, Elissa R Weitzman, and Rumi Chunara. 2017. Assessing Behavioral Stages From Social Media Data. In *Proc. CSCW*, Vol. 2017. NIH Public Access, 1320.
  - [53] Qiliang Liu, Min Deng, Yan Shi, and Jiaqiu Wang. 2012. A density-based spatial clustering algorithm considering both spatial proximity and attribute similarity. *Computers & Geosciences* 46 (2012), 296–309.
  - [54] Scott B MacKenzie and Philip M Podsakoff. 2012. Common method bias in marketing: causes, mechanisms, and procedural remedies. *Journal of Retailing* 88, 4 (2012), 542–555.
  - [55] Ruchit Nagar, Qingyu Yuan, Clark C Freifeld, Mauricio Santillana, Aaron Nojima, Rumi Chunara, and John S Brownstein. 2014. A case study of the New York City 2012–2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *Journal of medical internet research* 16, 10 (2014).
  - [56] Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. 2011. An empirical study of geographic user activity patterns in foursquare. *ICWSM* 11 (2011), 70–573.
  - [57] Daniel Preotiuc-Pietro, Justin Cranshaw, and Tae Yano. 2013. Exploring venue-based city-to-city similarity measures. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*. ACM, 16.
  - [58] William M Rand. 1971. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association* 66, 336 (1971), 846–850.
  - [59] Tye Rattenbury and Mor Naaman. 2009. Methods for extracting place semantics from Flickr tags. *ACM Transactions on the Web (TWEB)* 3, 1 (2009), 1.
  - [60] Leandro Araújo Silva, Mainack Mondal, Denzil Correa, Fabrício Benevenuto, and Ingmar Weber. 2016. Analyzing the Targets of Hate in Online Social Media.. In *ICWSM*. 687–690.
  - [61] Shivam Srivastava, Shiladitya Pande, and Sayan Ranu. 2015. Geo-social clustering of places from check-in data. In *Data Mining (ICDM), 2015 IEEE International Conference on*. IEEE, 985–990.
  - [62] Seth Stephens-Davidowitz. 2014. The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics* 118 (2014), 26–40.
  - [63] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. 2014. Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification.. In *ACL (1)*. 1555–1565.
  - [64] KE Tobin, M Cutchin, CA Latkin, and LM Takahashi. 2013. Social geographies of African American men who have sex with men (MSM): A qualitative exploration of the social, spatial and temporal context of HIV risk in Baltimore, Maryland. *Health & place* 22 (2013), 1–6.
  - [65] Roger Tourangeau, Kenneth A Rasinski, and Norman Bradburn. 1991. Measuring happiness in surveys: A test of the subtraction hypothesis. *Public Opinion Quarterly* 55, 2 (1991), 255–266.

- [66] Zeerak Waseem and Dirk Hovy. 2016. Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter.. In *SRW@ HLT-NAACL*. 88–93.
- [67] Michael J White. 1983. The measurement of spatial segregation. *American journal of sociology* (1983), 1008–1018.
- [68] Matthew L Williams and Pete Burnap. 2015. Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. *British Journal of Criminology* 56, 2 (2015), 211–238.
- [69] Hujun Yin. 2002. Data visualisation and manifold mapping using the ViSOM. *Neural Networks* 15, 8 (2002), 1005–1016.

Received April 2018; revised July 2018; accepted September 2018