9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Exploring the Use of Twitter Volume, Sentiment, and Content as a Proxy Measure of US Homelessness

Keywords: homelessness, social media, complex systems, signal detection, sentiment analysis

Extended Abstract

Unlike other public health measures, homelessness in America lacks a robust, uniformly-administered method for measuring local and nationwide prevalence [1]. Building on prior work that demonstrated common-sense links between Twitter data with state-level measures of public health [2], and motivated by a desire for low-cost, real-time, reliable tools to estimate local homelessness, we set out to test the ability of US-geotagged homelessness-related tweets to signal comparative measurements of homelessness and their fluctuations over time. Using the 2010-2019 and 2022 US Point in Time homelessness counts, a collection of over 960,000 US-geolocated tweets containing the word 'homeless,' and US Census land area and annual state-level population estimates, we examined how tweet volume, sentiment, and content correlated with per capita homelessness rates and per square land-mile density of homelessness. Across all years of the data, we found a statistically significant correlation between ranked state-level homelessness density and per capita tweet volume. Additionally, annualized within-state measures of homelessness across time exhibited statistically significant correlation with corresponding measures of tweet volume for many states experiencing widely variable or consistently high homelessness. Sentiment and content, however, signalled changes to homelessness trends only at the national scale, with composite nationwide sentiment scores anti-correlated to raw nationwide total homelessness counts and semantic tweet content becoming more politicized as national trends shifted.

We began with an investigation of Spearman ρ correlation measures between ranked per capita homelessness-related tweet rates and both per capita rates and per square land-mile densities of homelessness. Hypothesizing that high, increasing, or unstable measures of homelessness would increase per capita tweet rates, we tested correlations at the state level within each year of data, as well as across all years of data for each individual state. While overall state general-population density was not found to correlate at statistically significant levels with homelessness-related Twitter activity rates in any year, the correlation between a state's ranked *homelessness* density (number of persons experiencing homelessness per square land-mile) and its per capita homelessness-related Twitter activity was found to be statistically significant across all years of data. Likewise, ranked per capita homelessness rates correlated with per capita tweet rates among states from 2013-2019 and again in 2022.

We next sought to determine whether fluctuations within a single state across time might correlate to changes in the per capita homelessness-related tweet rate within that state. While in-year changes to per capita homelessness rates did not correlate across states to per capita tweet rates, sixteen states exhibited statistically significant nonparametric correlations across all eleven available years of data between some measure of localized homelessness (rate, density, or rate change since the prior year) and Twitter volume (rate or rate change). Among these states, six were both at or above the 75th percentile of per capita homelessness densities across all ten years and ranked in the top ten states with respect to variance in homelessness rate or

density, indicating both high point-in-time measures of homelessness and longitudinal instability. An additional two states ranked within the top ten states with respect to variance only; another was consistently at or above the 75th percentile only.

Further, we hypothesized that an increase in homelessness density within a given jurisdiction would result in more negatively-sentimented Twitter content. To the contrary, we found that, while ambient sentiment for tweets about homelessness was lower than that of English-language Twitter generally, it increased steadily over the 2010-2018 period, accelerating from 2016-2018 as changes in nationwide homelessness counts changed from decreasing to increasing for the first time in nearly a decade. We found a Pearson r correlation coefficient of -0.8421 (p-value = 0.0011) and a Spearman correlation coefficient of -0.6818 (p-value = 0.0208) between nationwide homelessness counts and the composite tweet sentiment score for all homelessness-related, US-geotagged tweets. Variance among state-level raw sentiment scores in any given year was extremely low, typically 0.01 or lower. Consequently, correlations were not considered between ambient sentiment and state-level homelessness, which was characterized by higher variance, rank-turbulence, and heterogeneity.

We then investigated the relationship of homelessness to the semantic content of tweets geoidentified with a given state. Using allotaxonometric measures of rank-turbulence divergence [3] between tweet corpora generated nationwide from 2010-2015 and from 2016-2019, we found that homelessness-specific and politicized language distinguished the later period of increasing homelessness, while words distinguishing the earlier period were more generic and less recognizably related to the phenomenon under study. Testing to determine whether this pattern of language use was also present at the state level, we found that it was present among states experiencing a consecutive multi-year period of increasing homelessness preceded by a period of decreasing homelessness, but not among states experiencing the reverse, suggesting that the language pattern correlated with time and not with changes to homelessness rates or densities. To generalize this finding, we constructed corpora from all state-years in which homelessness densities were increasing and those in which they were decreasing without respect to broader nationwide counts and found no meaningful semantic difference correlating with density or rate changes alone. Thus, these changes in language over time that are replicated at the state scale likely reflect changes to national discourse being diffused at the local level and not a mechanism that translates changes in local rates to sector-specific language.

In sum, our research suggests a mechanism by which the distributed probabilities of real-life, localized encounters with homelessness, as experienced by individuals both within a given year and across time, correlate significantly with state-level distributions of homelessness measures. At the same time, nation-scale shifts in homelessness counts may be signaled by more political language and positively-sentimented tweets.

References

1. “America’s first homelessness problem: Knowing who is actually homeless.” Swenson, Kyle. Washington Post, August 24, 2022. https://www.washingtonpost.com/dc-md-va/2022/08/ 24/homeless-seattle-hud-statistics/.

2. Alajajian, S. E., Williams, J. R., Reagan, A. J., Alajajian, S. C., Frank, M. R., Mitchell, L., Lahne, J., Danforth, C. M., & Dodds, P. S. (2017). The Lexicocalorimeter: Gauging public health through caloric input and output on social media.

3. Dodds, P. & Minot, J. & Arnold, M. & Alshaabi, T. & Adams, J. & Dewhurst, D. & Gray, T. & Frank, M. & Reagan, A. & Danforth, C.. (2020). Allotaxonometry and rank-turbulence divergence: A universal instrument for comparing complex systems.

State-level general-population densities, per capita tweet rates

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2022
Spearman ρ	-0.071	-0.019	0.103	0.066	0.054	0.083	-0.064	-0.097	-0.015	-0.064	-0.143
p-value	0.619	0.896	0.473	0.645	0.707	0.563	0.653	0.497	0.914	0.658	0.318

State-level homeless-population densities, per capita tweet rates

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2022
Spearman ρ	0.060	0.672	0.822	0.669	0.689	0.634	0.638	0.613	0.685	0.52	0.452
p-value	0.677	6.64e-8	1.46e-13	7.93e-8	2.13e-8	6.00e-7	4.66e-7	1.77e-6	2.87e-8	9.26e-5	8.59e-4

Table 1: Spearman ρ correlation between population density and homelessness-related tweet rates, statistically significant with respect to homelessness densities (bottom) but not general population densities (top)

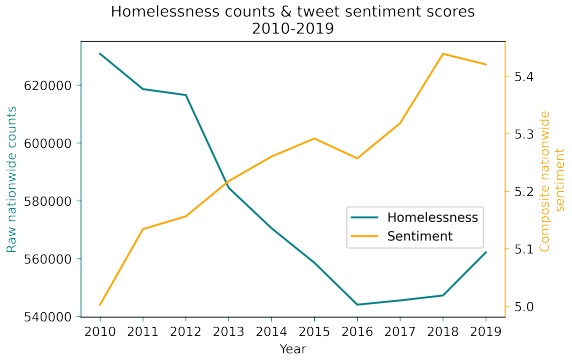

Figure 1: Annual measures of nationwide homelessness counts and composite tweet sentiment, positive change in sentiment scores accelerating as the nationwide trend changes from decreasing to increasing

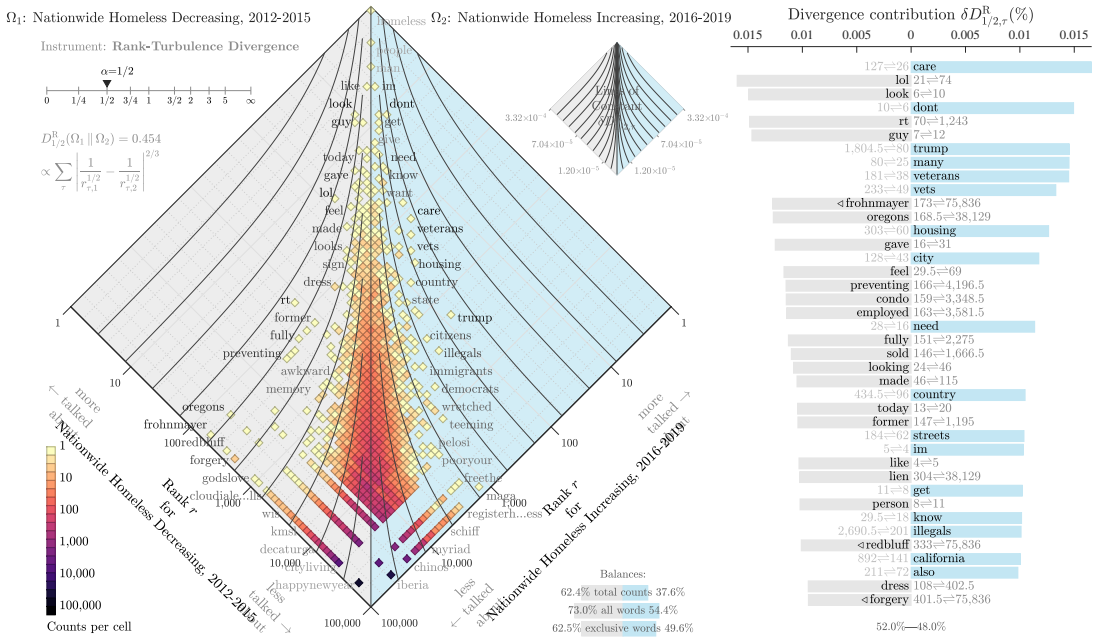

Figure 2: Allotaxonometric analysis of 2010-2015 versus 2016-2019 tweet content demonstrating higher-frequency political, homelessness-specific language in the period of increasing homelessness