Variations in the temporal structure of sociability across American cities

Abstract

Though sociologists have been interested in how temporal patterns of sociability vary in urban contexts, the study of city-level dynamics at short timescales has been challenging historically. Social media and new computational methods provide a solution. Our study clusters cities using sociality as a metric. We collected three months of social media data to investigate variation in the temporal structure of sociability across American cities. We find that cities cluster into three distinct types ('Coastal', 'Transitional' and 'Heartland') and that geographic proximity together with race, education, and language associate with this clustering. Specifically, we found that clusters of Blacker cities tend to tweet more per capita, but also that more highly educated cities tend to tweet less per capita. These findings provide evidence that social media may be facilitating new opportunities to empower traditionally marginalized urban groups, a conclusion relevant to #BlackLivesMatter, the George Floyd protests, and other social movements.

Key words: city dynamics, network analysis, temporality, social media, daily life, urban studies

Dhiraj Murthy* (School of Journalism and Media and Department of Sociology, University of Texas at Austin)
John D. O'Brien (Department of Mathematics, Bowdoin College)
Alexander Gross (Intermedia Arts, University of Maine)
Nathan L. Meyers (Department of Mathematics, Brown University)

*Corresponding Author: e-mail: Dhiraj.Murthy@austin.utexas.edu, Twitter: @dhirajmurthy, Address: Moody College of Communication, University of Texas at Austin 300 W. Dean Keeton, A1000, Austin, TX 78712-1067, USA

Introduction

Urban sociology has historically been interested in studying city-level similarities in terms of clustering and categorization of cities. Traditional comparison has focused on geography as well as socioeconomic and demographic attributes, including poverty, race, and education. (Blau & Blau, 1982; Lichter, 1988; Schmid, MacCannell, & Van Arsdol, 1958). Classically, Wirth (1938) emphasized the value of comparing cities by population, density, and 'heterogeneity.' Molotch (1976) explored unemployment rates across clusters of US cities to understand the role of urban growth. Moreover, studies of segregation patterns in cities continue to be a focus for understanding city-level dynamics (Grigoryeva & Ruef, 2015) and

Lefebvre (2004) classically studied rhythms and the impact of time on the inhabitants of urban spaces. However, while sociologists argue that "time matters" (Griffin 1992, p. 405), the temporal impacts on sociality in an urban area remains understudied (Hughes, 1999).

As mammals, humans have an inborn diurnal pattern (Küller, 2002, Koski, 2011, p. 2161) and, correspondingly, how time governs human social dynamics is a fundamental concern of many social sciences. Sociology has traditionally argued that cities with similar socioeconomic/demographic attributes tend to have similar temporal patterns of human dynamics. For example, cities with high unemployment rates and poverty may engage in distinct 'routine activities' during daytime hours relative to affluent cities with low unemployment rates (Stahura & Sloan, 1988). Additionally, attributes such as race, religion and ethnicity have been used to explain some of the cultural practices that could be dominant in a city and dictated by time. For example, maritime and finance-oriented cities may have their sociability shaped by the tides and the opening/closing of stock exchanges, respectively.

In this study, we examine classic questions of urban sociability previously not answerable to Lefebvre, Wirth or Molotch, but now examinable using social media data. We collected nearly 26 million social media posts from Twitter in fifty United States cities over a fourteen-week period, with each observation stamped with the time of collection and origin city. This approach builds from Griffin's (1992, p. 406) argument, drawing from Giddens and Abrams, that "[a]ll actions and events take place in real/spatial/temporal contexts (Giddens 1979; Abrams 1982), and therefore all sociological data are time- and place-specific." We argue that the temporal patterns of these social actions can be aggregated to reveal consistent patterns across cities.

The fundamental assumption of this study is that social engagement online is a relatively consistent fraction of total social engagement in the urban areas under study. While this fraction may change over longer time scales (years or decades), we assume that this fraction is constant across cities on the time scale of study. This does not require that the fraction be constant across times within the week provided this variance is shared across cities. This assumption may be confounded if the fraction of sociality is consistent within demographic groups but varies across groups. In this case, the observed pattern of Twitter sociality within a city will then convolve the composition of those demographic categories with the overall sociality. How the overall fraction might be measured and assessing the consistency across demographic categories are crucial questions with implications for the interpretation of our results that we take up in the Discussion section.

We analyze these data by inferring the temporal intensity for each city, developing an appropriate metric for comparing the distance of cities in this space, and then using established statistical methods for inferring cluster and network structures among the cities. In order to facilitate exploration by other researchers of these patterns, we present our findings in a fully reproducible fashion in an online supplement[1], including original data, computer scripts, and additional visualizations. Our analysis indicates that major US cities tend to cluster into three distinct groups that we label as 'Coastal', 'Heartland', and 'Transitional'. The inferred clusters are stable across the weeks within the study and consistent across days of the week. While clusters obviously have a relationship with the distance between cities (with closer cities likely to inhabit the same cluster), this relationship only explains part of the phenomenon. We employ a network-based regression technique to explore other factors and find that language patterns, education, and racial composition all play some role.

The data show that these finely-grained sociality measures can be used to cluster cities. Traditional sociological approaches have been limited to use geographic, demographic, and economic variables aggregated at the scale of the city. We show that refined temporal data can also accomplish clustering of American cities that can be partially associated with traditional variables. Unexpectedly, instead of higher education and whiteness enabling sociality, social media-measured sociality is actually higher for Blacker[2] cities and lower for educated, whiter cities. This has tremendous implications for understanding some of the ways traditionally marginalized urban groups may be using social media to challenge historical inequalities and indeed digital divides.

As a final point, we highlight the logical importance that finely grained social media may have for the discipline. Sociology has developed in parallel with its statistical resolution. The ecological fallacy – a classical problem since Quetelet's (2013) social statistics and, for

[1] The online supplement is accessible at: https://github.com/cascobayesian/Twitter
[2] We follow the convention from American sociology of using Blacker adjectivally in the context of comparing neighborhoods and cities (e.g. Massey & Denton (1988: 592) and Tolnay et al. (2000: 1001))

example, Durkheim's sociology (van Poppel & Day, 1996) – rears its head again in the digital age: how do we reason about individual behavior from data aggregated across a large, possibly heterogeneous group and how should we account for how messy, unstructured data may affect these inferences? Our study shows that social media data are capable of addressing aspects of aggregation that are not accessible using standard sociological techniques, as well as emphasizing the power of statistical modeling in the development of sociology.

Clustering Cities

Drawing from Hoselitz's (1953) grouping of cities into 'political intellectual centers' and 'economic centers', Redfield and Singer (1954, pp. 56-57) categorize by historical epoch to develop post-industrial comparisons of 'Metropolis-cities' - London, New York, Osaka, Yokohama, Shanghai, Singapore and Bombay - against '[c]ities of modern administration' such as Washington DC, New Delhi, and Canberra. Crain's (1966, p. 476) seminal study on why cities adopted fluoridation reveals convergence between cities that adopted technologies based on 'peer-group influence' driven by 'a mass communication system', across peer cities that share certain cultural norms, beliefs, patterns, and outlooks. More recently, Graham (2002) provides evidence of city personalities within the United States describing 'global metropolitan areas of the nation' such as New York, Chicago, Los Angeles, San Francisco/San Jose, and Atlanta that cluster in their concentrations of transnational corporations, telephony traffic, and the quality of their physical infrastructure (Graham, 2002, p. 76, Graham, 2002, p. 82, Table 1).

In an international context, Sassen (2011) studied 'global cities' such as London and New York whose power is built from intercity networks that enable a level of independence that 'bypass[es …] national states.' This framework analyzes shared features of disparate urban areas, particularly connectedness, rather than proximity. For Sassen (2002, p.3), these cities become particularly connected through "specialized global circuits for economic activities". As such, American cities such as New York, Chicago, Boston, Los Angeles, and San Francisco might cluster together based on interconnected economic activities. Sassen (2002, p. 6) also found that other clusters can be derived by studying the presence of top advertising agencies (i.e., New York, San Francisco, Los Angeles, Miami, Atlanta, Dallas, Minneapolis, and

Washington DC).

Temporalilty and City Clustering

Time is capable of rendering similarities and differences among cities. Much work has been done comparing the pace and diurnal patterns of urban life (Bettencourt et al. 2007). Sociology has traditionally been interested in time and the rhythms of urban life. Indeed, Lefebvre's (2004) *Rhythmanalysis* applies dialectical methods to understand the intersections of time, space, and rhythm. The more analytic empirical work of Cohen and Felson (1979, p. 589) applies an ecological approach to understanding that certain acts (in their work, illegal) "occur at specific locations in space and time", arguing that 'routine activities' are interdependent. This approach has been empirically extended to highlight how such acts "feed on the spatial and temporal structure of routine legal activities (e.g. transportation, work, and shopping)" (Sampson & Raudenbush, 1999, p. 610). The contemporary generation of digital 'trace data' – online data that reflects everything from our mobile-phone derived movement to social, political, and economic actions – potentially renders visible rich city-level spatial-temporal structures across a wide variety of acts or aggregations of acts (Jiang, Joseph Ferreira, & Gonzalez, 2012). These data may also provide methods to extend Lefebvre's qualitative approach to quantitative methods that reveal some of the rhythmic patterns he saw as important.

Sociology and Sociability

Amirou (1989, p. 115) notes that one of the first definitions of 'sociality' by Max Scheler was used to 'establish what is called reality.' Indeed, Amirou (1989, p. 116) sees Max Weber's invocation of sociality as fundamentally anchored in 'everyday life.' Simmel also pivots the term towards more 'playful' iterations of social interactions in his 'The Sociology of Sociability' (Simmel & Hughes 1949), paving the way for quite contemporary notions of the social. Simmel and Hughes (1949, p. 255) ascribe to sociability "a feeling [... of] satisfaction in, the very fact that one is associated with others and that the solitariness of the individual is resolved into togetherness, a union with others."

Simmel also maintained that sociability inherently had a democratic streak, arguing "[r]iches and social position, learning and fame, exceptional capacities and merits of the individual have no role in sociability" (Simmel & Hughes 1949, p. 256). Notably, Simmel argues that sociability is rooted in 'banal experience' with conversation at its core, where "[t]alking is an end in itself." Moreover, for Simmel, the content of our 'talking' is not as important as the action of communication itself, arguing that "content is merely the indispensable carrier of the stimulation" (Simmel & Hughes 1949, p. 259). He argues that this is in stark contrast to a business association where speech is a means to an end; for sociability, speech is "the whole meaning and content of the social process (Simmel & Hughes 1949, p. 256)." Sociability has particular relevance to contemporary sociology in the realm of social media, where platforms are both used as a means of communication and performance (Murthy & Gross 2017). Twitter has the potential to foster online public spheres, allowing individuals to interact (Murthy 2018) and, following Simmel, engage in 'talking', whether serious or playful.

Cities as an Aggregation of Sociability

Deleuze's and Guattari's (1987) arboreal thinking sees cities as rhizomatic, wherein a root top dies and there is a new rejuvenation from the top. They see the social as a de-centered process, wherein, as Savage (2011, p. 517), commenting on Deleuze and Guattari, remarks, 'fixed location can be seen as the sedimented product of intensive flows.' The aggregation of sociability can partially be represented as this sedimented product. In the context of social media, tweets, regardless of content are part of the flows of sociability, whose sedimented product represent some aggregation of sociability.

Though the city is a 'site of mobility and mobilities' (Bridge & Watson, 2011, p. 157), Simmel (1950) remarked how strangers do not approach others in the city but often seek to avoid them, creating an equilibrium of sociality (Sennett, 2011, pp. 390-391). However, in contemporary cities, this socially inhibitory quality may encourage virtual community with their fellow city peers leading to a new form of Simmelian equilibrium borne from engagement rather than avoidance (Soukup, 2006). Indeed, Simmel's (1949, p. 255) classic work on sociability emphasizes the notion of Gesellschaft ('togetherness') to understanding sociability. Digital forms of sociability mean that we can be 'together' in new ways in cities.

Social Media and Urban Sociology

Urban life is increasingly associated with social media and these interactions play a key role in the sociability of cities (Goodspeed 2017). As urban individuals are 'talking' via social media, rates of usage of these forums can be an effective proxy to observe city-level sociability and so points a way out of the limitations of traditional data collection methods. Unlike previous approaches, these data have the potential to directly observe social interactions at the atomic-level (individual behavior) that can be aggregated at whatever spatial scale we set. This may permit both the systematic examination of the structure and quality of aggregation as well as provide a means for direct testing of the statistical. Such uses of digital trace data have been encouraged through Burrows' and Savage's (2014) critique of sociology's historical lack of meaningful engagement with new media data.

The set of all social interactions of an individual person across a closed set of individuals constitutes the full scope of sociality (though not everything in the social field, as, for example, Bourdieu's habitus). In the case of Twitter, a social media platform widely used in social research, tweets form a subset of individuals' total social interactions. Understanding the structure of how tweets - and social media data generally - are distributed within this broader set of social interactions is critical to appropriate interpretation of these data in urban sociological analysis. Specifically, within a city, there are two broad sampling processes required by aggregation that are not directly observable: how an individual's Twitter use relates to their patterns of social interactions; and how that use varies across demographic categories. For instance, an individual might use social media more in the late evening while most of their face-to-face interactions are during school hours. This pattern may in turn be consistent across individuals sharing demographic features, say affluent Black teenagers. The observed city-level temporal social media distribution is the aggregation of these two processes across individuals.

Ultimately, traditional quantitative urban sociology has been often constrained to interrogate the structure of urban social environments with regression and spatial methods (Peterson & Krivo, 1993; Schnore, 1963; White, 1983). Though undeniably valuable, because these methods are not generative – i.e., they do not establish models of behavior, only determine patterns of association among variables – they are always threatened by the ecological fallacy. Computational sociology in the form of agent-based models points a possible way out of this by allowing researchers to simulate how individual behavior becomes group-level phenomena, but has yet to be connected to data (Bonabeau, 2002; Epstein, 1999; Macy & Willer, 2002).

Social Media and City-Level Dynamics

A diverse range of social media platforms and their data has been used to empirically study city-level dynamics and clustering patterns. Cranshaw et al.'s (2012) Livehoods project used 18 million location-based check-ins from Foursquare to study the social dynamics of a single American city, Pittsburgh, Pennsylvania. They argue traditional categories like neighborhoods do not capture the full social dynamics of cities and go on to develop an algorithm that uses their data to identify new types of social clustering. Their in-depth analysis of the opening of a branch of an upscale, organic-based market on the border between affluent and economically depressed areas shows rapid changes in the structure of location-based check-ins. The speed at which they were able to examine social changes through social media data highlights the utility of these approaches. Similarly, by treating users as social sensors, Pan et al. (2013) employed data from the Chinese social media site Weibo alongside GPS data from taxis in Beijing to identify traffic anomalies within six seconds.

Explorations of visual social media data have also provided comparisons of behavior across cities. Hochman and Manovich (2013) examined 2.3 million Instagram photos sampled from 13 global cities through a visualization procedure to decipher similarities and differences. Visual similarities across their montages potentially indicate patterns in social media image posting behaviors. Silva et al.'s (2013) use of Foursquare and Instagram data compares social dynamics in New York, São Paulo, and Tokyo by looking at sequences of social media behavior. Even after a year, social media sharing patterns on these platforms remain relatively constant, suggesting the development of certain entrenched routines and patterns and stable patterns of observation. Diurnal analysis found that the sharing pattern by day and time for American cities and Japanese cities usually peaked at lunch and dinner and in the Brazilian context, peaks did not necessarily coincide with mealtimes. Their work indicates particular sharing behaviors, what they label as a "cultural signature", a unique city-level signal in these data. These diverse, descriptive case studies all point to possibilities of discerning patterns and/or distinctive signatures for city-level dynamics, a prerequisite for studying temporal clustering. However, none undertakes the methodological challenge of measuring city-level dynamics at a big data scale.

Research Questions

Our research questions (RQs) draw from this theoretically-informed context. Specifically, we seek to understand the intersections of time and urban sociability using social media data in order to identify whether there exist signatures that enable us to cluster American cities. Tweets are highly varied in their content, but their occurrence patterns signal some form of sociability, connectedness, humor, social awareness, etc. We therefore ask the following RQs:

RQ1: What is the temporal structure of urban sociability seen through social media data? As Wajcman (2008, p. 62) argues, “changes in the temporal structure of modern societies transform the very essence of our culture, social structure, and personal identity”. Our ability to measure the temporal structure of urban life empirically has been restricted by limited data availability. How sociability varies through the day is a natural avenue for sociological research, as it connects structural/infrastructural, cultural/anthropological, and biological aspects of behavior into a single nexus that can be measured quantitatively.

RQ2: What is the structure of variation in this pattern of sociability across cities? How much variation exists is largely unknown and in our consideration of American cities, we explored whether variation from the baseline pattern is structured (e.g. organized along gradients, clusters, or other patterns).

RQ3: Can the metric of social media data be used to study urban sociability? Twitter data is increasingly used as a social sensor, what Mejova et al. (2015) term a ‘digital socioscope’. However, Twitter data, like most social media data, is a biased subset and its bias needs elaboration to evaluate whether these data can be meaningfully used to study the structure of sociability across American cities.

Data and Methods

Data collection

The data comprise 25,963,512 tweets from 48 cities in the continental United States, collected via Twitter’s Streaming API from April 1 to July 1, 2012. This API endpoint provides a ~1% sample of all tweets. Though these data are no longer contemporary, the early years of Twitter represent a unique time in the platform’s history where bots had not become a dominant force and users had high levels of geo-location enabled. Other recently published work has also used older datasets such as a particularly rich sample from 2010 for this reason (e.g, Rocco et al. 2019). Initially, 50 cities were selected, but two were removed due to data quality issues. The largest city possessed 1.8 million observations, while the smallest had slightly less than 50,000. While unique individual user identification was collected, this information was discarded and each recorded message is treated as equivalent. Due to slightly uneven stopping and starting times for each city, we trimmed the data to include only the fourteen weeks beginning at midnight April 8 and ending midnight June 30, adjusted for time zones. We then binned the data by city, day, week, into 15-minute intervals within each day. Since cities were in different time zones, we adjusted time stamps by the time zone of their observation rather than universal standard time as recorded by Twitter.

Using the US Census Administration's DataFerret online tool (U.S. Census Bureau, N/A), we collected 2010 Census demographic data on the 48 cities, including population, median age, fraction of English speakers, Spanish speakers, fraction of individuals self-identifying as 'White', 'Black', 'Hispanic', 'Asian', and 'American Indian', fraction of individuals with less than high school, high school, undergraduate, and graduate education levels, poverty rate, percentage of US citizens, and median income. Together with the total number of tweets, we present a summary of these in Table 1.

[Table 1 goes here]

As cities vary significantly in population and this strongly affects the number of tweets observed (see Figure 1), we first ensure that comparison across cities was possible by renormalizing the data. We accomplish this by arranging the data for each city into 672 15-minute bins for each week, and dividing by the total number of observations. This effectively controls for differences in counts across each city, rendering all as bins within cities in terms of fraction of total observations per week.

[Figure 1 goes here]

The number of tweets in total (y-axis) by the city population (x-axis), with both axes in log scale.

Removing non-human users

In order to ensure that the data reflect sociability rather than artifacts of artificially generated posts created by computer-assisted posting agents (i.e., bots) used both commercially and non-commercially, we performed a prior analysis on the city tweet data to assess these anomalous data. To do this, we processed each user into a time series, examining the frequency with which different lag times appeared in a user's data. Those users that exhibit near exact regularity in their lag times we deemed likely bots. In order to make an automated algorithmic cut-off for bots, we applied a fast-Fourier transform (FFT) to the lag data. Examination of highly probable bots indicated that a 3.5 ratio of convex hull area to the area of drawn by the FFT identified these users with a minimal false positive rate. We applied this cut off to all of the data, removing 2.3% of users.

Calibrating outages

The Twitter Application Programming Interface (API), the interface used to collect large volumes of public data from Twitter, is a commercial system that has outages, updates, and repairs. We took steps to ameliorate the impacts of these events on the analysis by plotting the fraction of cities that have no observations within a bin. Appendix 1 shows the resulting structure of these observations, revealing both regular troughs of usage in the early morning and large-scale outages. A regular outage occurs starting 12 am Monday mornings, presumably for standard maintenance: Twitter is not offline during these times but the API is not responding to queries. Other unanticipated outages occur sporadically: Tuesday in week 14; Tuesday evening in week 7; Wednesday evening in week 3; Thursday evening in week 7; and Saturday evening/Sunday morning in weeks 1 and 8. While some of these outages are partial, we elect to treat all observations during these periods as missing values that are removed for subsequent analysis. We distinguish the regular troughs in usage from outages by using a cut-off of 7 cities exhibiting zeros simultaneously. This leads to the identified outages in Appendix 2.

Visualizing patterns of usage

For each city, we plot the density vector for each day across the study as well as the median value for each bin and inspected each for systematic patterns and corresponding anomalies. With near-perfect consistency, we observe a diurnal cycle within each city and each day, with the lowest density values appearing between 3 and 6 am. The time of the highest value varied significantly by day and by city but usually occurred between 5 and 10 pm. Saturday and Sunday display distinct temporal patterns from the weekdays. The results for all cities are shown in Figure 2. We note significantly more variation between succeeding intervals in smaller cities than those observed in larger cities owing to differences in sample size. This variation motivates the smoothing procedure described next.

[Figure 2 goes here]

Averaged means taken across all cities for each 15-minute interval grouped by day. Dotted black line indicates overall mean across all days.

The study varies significantly in sample size by city: Boulder contains 49,920 observations, New York has 1,833,849. As this creates additional heteroscedasticity within days and weeks on top of genuine temporal variation, scaling each city by a single value will not address the issue. Instead, we smooth the data by averaging over a window of neighboring intervals. Defining t_{cwdq} to be the observed fraction withn city c, week w, day d, and bin q, we calculate the lag-smoothed value for $\tilde{t}_{cwdi}$ with a lag *l*, we average over the values from *i-l* and *i+l*:

$$\tilde{t}_{cwdi} = \frac{1}{2l+1} \sum_{q=i-l}^{i+l} t_{cwdq}$$

with the boundaries of the data set averaging only over available values. All values are then normalized to one across for each city within a week. We choose the value of *l* to balance the trade-off between reduced variation and a reduction in the overall temporal signal, finding that *l=2* is a reasonable compromise (see Figure 3). We note that this minimizes the effect, but does not entirely eliminate it (see, for instance, Boulder versus Denver in the online supplement). To emphasize the variational aspect, we further subtract out the overall diurnal average across all cities to get the smoothed variation from the mean for each city, as shown in Figure 4. We refer to these curves as mean-subtracted variations (MSVs). While noticeable variation persists within the smallest cities (see Boulder and Albuquerque, for instance) most variation between successive intervals is less than variation across days or cities.

[Figure 3 goes here]

Lag 2 smoothed tweet intensity profiles for each city. Days are colored as in Figure 2.

[Figure 4 goes here]

MSV lag 2 profiles for each city. Days are colored as in Figure 2.

Examining Figure 4, we observe a number of distinct patterns and several notable anomalies in the MSVs. The most salient is that some cities exhibit nearly ‘flat’ MSVs, indicating that for all days these cities exhibit usage patterns very close to the overall average. Dallas and Houston are particularly strong examples. Another apparent structure is the regularity of deviations from a flat pattern: many cities exhibit either a ‘bow,’ with fewer observations in the late evening and early morning (San Diego, San Francisco, New York, and Albuquerque) and an evening spike (Fresno, Richmond, and Pittsburgh). In the next section, we attempt to quantify these patterns. We also note city-specific anomalies against these patterns: New Orleans shows a strong spike in activity on Saturday evening, while Miami and Orlando show similar spikes early on Monday morning. These can be plausibly attributed to city-specific tourist traffic but should be considered when interpreting the network regression results presented in the Network construction section.

Network construction

To quantify the variation in MSV across cities, we employ the Jensen-Shannon (JS) metric applied to the original city-specific probability densities described previously. The JS is a function that measures the distance between two probability densities, with a value of zero when two distributions are identical, with increasing positive values as the differences between the distributions increase. We apply this metric to each pair of cities, creating a matrix of 1128 distinct entries.

To further understand the matrix, we construct a set of undirected networks where cities are vertices and edges form above a minimum JS distance. This level of similarity is a threshold where we separate cities into those that appear close from those that appear distant in the space of MSVs. Setting this threshold presents an element of choice; so, in all analyses, we perform several runs over a set of these values to ensure robust conclusions. A reasonable lower bound on these values is 0.0084, the smallest value such that all cities form a single, connected network. A reasonable upper bound is 0.0102, at which point more than 50% of all possible edges are included in the network.

Regression

To analyze the effect that the metadata might have on the similarities among cities, we employ a cluster-based latent-space approach to network regression. This model accounts for cities as being members of distinct clusters, whose membership permits the construction of a latent space that situates each within an unobserved social space, while accounting for the effects of additional covariates on the probability of observing a link between cities. This is accomplished via a logistic formulation for the likelihood (given in Handcock, Hunter, Butts, Goodreau, & Morris, 2008; Hoff, Raftery, & Handcock, 2002).

We began with a full set of US Census regressors (described previously) and computed the pairwise correlation between each (illustrated in Figure 5). Accounting for sample size (N=48), this indicates strong correlation amongst five distinct regressor blocks, with little cross-correlation across the blocks. We select a set of representative regressors from each block to limit the amount of correlation, with the exception of income, which associates with two blocks. From this set, we select the significant regressors, the dimension of the latent space, and the number of underlying clusters using the Bayesian information criterion embedded in the regression algorithm. This procedure selects age, rate of English spoken, frequency of African-Americans and distance between cities as significant regressors, the dimension of the latent space to be two-dimensional and three clusters partitioning the network.

[Figure 5 goes here]
Pairwise correlations between full set of regressors. Note the five distinct blocks of highly correlated regressors.

[Figure 6 goes here]
Projection of the network embedded in latent space into two dimensions, with cities marked by pie charts denoting proportion of membership in the three clusters: Heartland, Coastal, and Transitional.

[Figure 7 goes here]

Pairwise connections between cities plotted from white to black, with black representing the strongest connections (JS distance close to one).

[Table 2 results go here]

Regression results from latent network analysis with standardized effects. The inferred coefficients provide the log-odds ratio of two compared cities having a network connection beyond the contribution from the latent space structure. See Hoff, Raftery, and Handcock (2002) for a more complete interpretation. Regressors taken from 2010 US Census data. '% English' refers to the self-reported fraction of households that speak English at home; 'Age' is the median age for the city; 'Black' refers to the percentage of respondents who self-identified as Black; and, 'Distance' is a pairwise-variable giving the distance between any two cities in kilometers.

Results

Twitter-based city clustering

Figure 1 provides a first step in answering RQ2, showing a strong linear trend (ρ=0.88) between the population size and the observed number of counts over the study for each of the cities included, with San Jose, Boulder, Los Angeles, and Washington as outliers. This pattern provides context for our subsequent analysis, showing that larger cities can be treated as 'scaled-up' versions of smaller cities. The strong linear trend supports the later analytic assumption that cities' temporal count distributions are comparable. Figure 4 shows the smoothed and normalized diurnal profiles for each city with the mean profile subtracted out. We observe strong similarities among certain cities (for instance, Seattle/San Francisco and Detroit/Cleveland) taken across all seven days, arguing that RQ1 can be answered through a clustering of cities. Certain cities (Minneapolis, Dallas, and Nashville) exhibit patterns very close to the overall average, yielding a nearly flat profile, while most show appreciable divergence. Visual inspection indicates that most cities display similar patterns with other cities (for instance, Denver and Boulder) while retaining recognizable distinctions from the overall average, implying a broad pattern of clustering.

Figure 4 underlines this pattern by adjusting the temporal distribution for each city by subtracting the overall mean for each day. Contributing to both RQ1 and RQ2, this portrait provides a striking presentation of the differences and similarities among cities, showing that some cities exhibit patterns close to the overall average denoted by the grey dashed line (e.g. Houston and Chicago) while most cities differ in noticeable respects, most commonly by having significantly lower or higher observed counts in the late evening and middle day. This also highlights some subtle day-specific effects, such as the spike in early Monday counts seen in Miami, Orlando, Raleigh, Oklahoma City and Minneapolis (perhaps due to air travel). Many cities exhibit strong similarities (e.g., San Francisco/Seattle or Cleveland/Cincinnati) while being distinct across these clusters (i.e. San Francisco appears distinct from Cincinnati). These patterns, with clustering observed across cities of vastly different sizes, imply the RQ3 can be resolved in the affirmative, even if researchers must be mindful of how issues around data aggregation could influence the interpretation of these clusters.

The regression analysis and network structure in Figure 7 both indicate a distinct geographical aspect to the social distances among cities, with cities located closer to each other more likely to exhibit connections than more distant cities, speaking to RQ2. We observe strong inter-city connections among the cities on the West coast, the East coast, and land-locked cities. A number of inter-coastal ties exist, largely between Northeastern cities, (Boston, New York, and Philadelphia) and West Coast cities, with few ties to Southern cities. A number of ties connect cities located in the Rocky Mountains with Southern coastal cities.

Application of the latent space regression model to the inferred network indicates that three clusters situated in two dimensions provide the best explanation for the structure of the network. We label these clusters 'Coastal', 'Transitional' and 'Heartland.' These labels reflect that the most distinct cluster comprises cities solely on the Northeastern coast and the West coast. Cities in this cluster lie nearly wholly within their assigned cluster (see Figure 6). Transitional cities, including Minneapolis, Austin, Nashville, Los Angeles, Rochester and Raleigh, often exhibit some membership in the Coastal cluster and the Heartland cluster. Cities located in the Heartland cluster often exhibit fractional membership in the Transitional cluster, though not always (e.g., Atlanta). These clusters correspond almost identically to a k-medoids analysis based on the JS-matrix, providing statistical support to these assignments. The latent space approach simultaneously performs a regression analysis to find variables that account for additional linkages amongst cities not accounted for by the clustering structure. These results (see Table 2) indicate the strongest relationship between the distance between cities and their patterning ($p = 0.00004$), with weaker but still noticeable dependency on age ($p = 0.02917$) and the frequency of non-English speaking homes ($p = 0.04581$). Cities in the Coastal cluster exhibit younger ages and fewer English speakers, with cities in the Heartland showing the opposite.

In summary, for RQ1, our examination of city-specific patterns of sociality indicates broad similarity in temporality across all cities. We find strong diurnal patterns observed on each day with variation between the weekdays and the weekend. Across all cities, weekdays exhibit relatively less sociality during the day, with a noticeable peak in the late evening. Weekends exhibit an inversion: more sociality during the day and relatively less at night. For RQ2, we measured how close the sociality patterns for each city fall from each other. Our findings indicate that there is relative consistency in temporality and this can be further decomposed into our three broad categories of 'Coastal', 'Transitional', and 'Heartland'. We grouped all cities into their nearest cluster and then plotted the average mean-subtracted variations (MSV) by city. Lastly, in regards to RQ3, while we were not able to definitively answer this question, our response highlights important aspects of this question – in particular, how biases in aggregation can affect interpretation – that are highly relevant for future research. Moreover, our response brings a measure of clarity to the reliability of the 'social signal' by considering the structure of outages as well as attempting to rescale the data to account for varying sample size.

[Figure 9 goes here]
MSV for each cluster - solid lines indicate cluster mean and shaded region indicates the range of individual cities within the clusters

Discussion

In this study, we established that the variation in the temporal structure of digital sociability is stable across American cities and that these cities exhibit distinct yet structured patterns. This finding also opens important methodological questions about these data. We found that urban American cities can be situated into three clusters: 'Coastal', 'Heartland', and 'Transitional.' In conversation with the global cities literature, we asked whether the connectivity and shared features of disparate urban areas could be more responsible for the clustering of cities, rather than geographical proximity. Just like cities cluster in terms of sociocultural processes such as fitness patterns (e.g. Sallis et al., 1985), cities cluster separately from geographic proximity. Specifically, we found that the Coastal cluster includes the West Coast as well as Northeastern coastal cities, reflecting shared features and connectivity. Additionally, the outliers in the observed counts by city population (i.e., San Jose, Boulder, Washington DC, and Los Angeles) raise questions about what is happening in these cities to exhibit marked variance. For example, our work does not answer whether communities of white or Black users in Washington DC affect this pattern. Similarly, Boulder and San Jose are highly educated, tech cities and their low tweet counts could be attributable to this city-level signature. Los Angeles, on the other hand, is likely affected by its concentration of media and entertainment industries.

Though aggregation across cities indicated particular signatures (e.g., spikes at meal times), we found associations between American cities' cultural signature and a city's tweeting pattern. The value of this in terms of deciphering the social is fundamentally important. First, this approach - following Cranshaw et al. (2012) - indicates new clusters of cities that are not bound to traditional socioeconomic or geographical clustering (e.g. Los Angeles, Salt Lake City, and Raleigh). Second, following the work of Silva et al. (2013), the time of day in which a particular peak occurs could indicate particular signatures (e.g., based on shared features or connectedness). Notwithstanding the limitations outlined in the Limitations and Future Work section, we found that our social media-based clustering can be associated with traditional socioeconomic variables such as education level, citizenship rates, and racial composition. Uniquely, we found evidence that not only do clusters of Blacker cities tend to tweet more per capita, but also more highly educated cities tend to tweet less per capita.

As Deleuze & Guattari (1987) argue, sociality does have parallels to processes in an ecosystem, such as their invocation of sedimentation and the layering process of a de-centered formation of sociality. The sedimentary processes they reference are perhaps not exclusively bounded to just one city, but interconnections at the national or even international levels that we can measure using our novel method of social media data analysis. Our diurnal analysis has not only made an argument that clustering exhibits interconnections at a more macro-scale (i.e., nationally in our case) where ecological space is not constricted to the city. Rather, the beats and rhythms, following Lefebvre (2004), at the city-level also have resonance potentially at the cluster level, where they are manifested as signatures which echo the patterns seen with other cities, ultimately enabling us to cluster cities. This speaks to some of sociology's attempts to understand interconnectedness at larger scales as well through the notion of 'global cities' (Sassen 2011). So seeing cluster formation among cities we would traditionally not expect to see similarities between provides opportunities for future research to identify why such clustering may be happening.

Our study of urban American tweeting patterns also opens up new methodological/statistical approaches. Social media such as Twitter have biases and we believe that innovative approaches are needed to not only evaluate data biases, but to also interrogate how samples from Twitter are biased by demographics. Ultimately, we observe strong differences in the temporal tweet distribution across cities that may be attributed to broad changes in individual behavior (e.g., individuals in a city tend to tweet relatively more in the evenings) or the relative changes within a single demographic group (e.g., affluent, Black teenagers tweet more in a particular city). Our analysis here cannot distinguish between these two causes and for the purpose of analysis we have assumed that the proportion of sampled tweets occurs homogeneously across a city. This assumption, while strong, is a natural starting point for research: having established the existence of distinct temporal patterns across cities, one can disentangle them through the use of more involved study designs as we discuss in the following section.

Limitations and Future Work

The issue of incorrect inference owing to unseen aspects of aggregation lurks behind the interpretation of this and similar work. Subsequent research to understand the structure of aggregation and further examination of the effects of sampling will be invaluable in calibrating these data types for sociological analysis. While our analysis indicates broad clustering of temporal variability following geographic and cultural features, some of our observed clustering patterns are not easily explainable and require further analysis. Fortunately, Twitter and other social media data allow for closer examination of how granular data become larger patterns: by more closely examining how individual users' behavior combine to create these larger features, researchers can definitively ascertain the effects of this aggregation. The strong linear relationship between city size and tweet frequency gives some support to the unbiased collection procedure, though more complex sampling procedures will need to be deployed to assess the relative bias of social media observations with other data types. Future work can also build upon our study by examining how American cities compare with cities worldwide, which would provide more robustness in overall inference. As tweeting is a minority activity and only one metric of temporal intensity, future work can also extend our theoretical and methodological framework to other social media platforms. Recent advances in machine learning could also be applied to discern events and their role in shaping tweeting behaviors. Furthermore, our approach can be applied to phone data or non-computer mediated communication, such as speech interaction. As future work explores these avenues, we will also have a clearer understanding of how well social media subsets capture urban dynamics and better address the ecological issue of how the process of aggregation limits how strongly one can interpolate city-level structure into individual-level behavior.

Conclusions

The study of city-level social dynamics at large urban scales has traditionally been difficult. Trying to study the pace of life between cities or anomalies in behavior within cities manifests empirical challenges. Though classic sociological studies such as Lefebvre's (2004) *Rhythmanalysis* did so qualitatively, large-scale quantitative attempts have largely been absent. In the social sciences, census-derived data has been fundamental to studying city-level clustering. Despite being collected regularly, they take time to be published after they are gathered. Additionally, some of these data are reported by respondents (and subject to so-called 'Hawthorne Effects' (Stephen, 1992)). As Burrows and Savage (2014, p. 3) argue, 'Big Data digital tracing', the study of metadata such as the reporting of location via smartphones, can give us a better picture of social life based on actual actions rather than respondent-provided 'accounts of actions.' Users act as real-time social sensors (Sagl, Resch, Hawelka, & Beinat, 2012), producing data from many facets of their daily life (e.g., location check-ins). Here, we have extended computational methods using rich social media 'trace' data to address the type of classic sociological questions Lefebvre, Simmel, and others were only partially able to answer with the methods of their time.

In this study, we collected 25,963,512 tweets from April - July 2012 from 48 US cities and paired this with US Census data as regressors. Our analysis evaluated whether cities are stable in the temporal distribution of observations. Previous literature on social media and diurnal behavior patterns suggested strong stability in these patterns (Grinberg, et al., 2013), which we confirm. However, we also found that cities significantly vary in aspects of their temporal structure. Second, and perhaps more importantly, we found that cities do not have unique signatures per se, but group in broad, discernible clusters that associate with traditionally important social/cultural/demographic factors such as race and education level. Through a regression analysis, we find that the clustering is based on spatial distance as well as on language, age, and race. By simultaneously incorporating messy, unstructured data from Twitter with traditional, structured census data, our study is able to provide unique insights into similarities and differences between American cities. We found evidence that not only do clusters of Blacker cities tend to tweet more per capita, but also that more highly educated cities tend to tweet less per capita. We hope that future work can further evaluate these claims, particularly through studies of other social media platforms as well as newer data sets.

We find that studying more fine-grained temporality matters. Sociology has traditionally studied time and cities from a diurnal perspective, where city life may be affected by the temporal cycles of the day. Moreover, what happens in cities at particular times of day could be partially explained by, for example, primary industries or demographic attributes including race, education-level, employment status, age, income, etc. Furthermore, if cities have a similar geographical position and share similar sociological variables, they could be clustered by these factors. Pioneering work by sociologists studying 'global cities' (e.g., Sassen (2011)) argued that new networks fostered by intensive globalization enabled cities to transgress traditional geographic limitations and cities could cluster on connectedness enabled by various networks. Sociality remained challenging to study within this fairly economic-anchored model. However, unlike traditional sociological approaches that have been bound by geographic, demographic, or connectivity variables, our analysis of fine-grained temporality using large-scale social media data shows a distinct perspective on urban clustering that only can be partially explained through these traditional variables. Unexpectedly, instead of higher education level and race (particularly whiteness) enabling sociality, social media-measured sociability is actually higher for Blacker cities and lower for educated, more white cities. These findings provide evidence that social media may be facilitating new opportunities to empower and give voice to traditionally marginalized urban groups, a conclusion highly relevant to understanding social media's role in contemporary social movements such as the George Floyd protests.

Acknowledgments

The authors would like to thank Sarah Brayne for feedback on an earlier version of the manuscript and members of the Social Network Innovation Lab for their feedback on methods.

References

Amin, A & Thrift, N 2002, *Cities: reimagining the urban*, Polity Press.

Arthur S. Alderson & Jason Beckfield 2004, 'Power and Position in the World City System', *American Journal of Sociology*, vol. 109, no. 4, pp. 811-51.

Bettencourt, LMA, Lobo, J, Helbing, D, Kühnert, C & West, GB 2007, 'Growth, innovation, scaling, and the pace of life in cities', *Proceedings of the National Academy of Sciences*, vol. 104, no. 17, pp. 7301-6.

Blau, JR & Blau, PM 1982, 'The Cost of Inequality: Metropolitan Structure and Violent Crime', *American Sociological Review*, vol. 47, no. 1, pp. 114-29.

Bonabeau, E 2002, 'Agent-based modeling: Methods and techniques for simulating human systems', *Proceedings of the National Academy of Sciences*, vol. 99, no. suppl 3, pp. 7280-7.

Bridge, G & Watson, S 2011, *The new Blackwell companion to the city*, Wiley-Blackwell, Oxford.

Burrows, R & Savage, M 2014, *After the crisis? Big Data and the methodological challenges of empirical sociology*, vol. 1.

Cohen, LE & Felson, M 1979, 'Social change and crime rate trends: A routine activity approach', *American Sociological Review*, pp. 588-608.

Crain, RL 1966, 'Fluoridation: the diffusion of an innovation among cities', *Social Forces*, vol. 44, no. 4, pp. 467-76.

Cranshaw, J, Schwartz, R, Hong, JI & Sadeh, NM 2012, 'The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City', in *ICWSM*.

Deleuze, G & Guattari, F 1987, *A thousand plateaus (B. Massumi, Trans.)*, University of Minnesota Press, Minneapolis.

Edwards, Je & Booth, Ae 1973, *Social participation in urban society*, Schenkman Pub Co., Cambridge.

Epstein, JM 1999, 'Agent-based computational models and generative social science', *Complexity*, vol. 4, no. 5, pp. 41-60.

Goodspeed, R 2017, 'Community and Urban Places in a Digital World', *City & Community*, vol. 16, no. 1, pp. 9-15.

Graham, S 2002, 'Communication Grids: Cities and Infrastructure', in S Sassen (ed.), *Global networks, linked cities*, Routledge, New York ; London, pp. 71-92.

Griffin, LJ 1992, 'Temporality, Events, and Explanation in Historical Sociology: An Introduction', *Sociological Methods & Research*, vol. 20, no. 4, pp. 403-27.

Grigoryeva, A & Ruef, M 2015, 'The Historical Demography of Racial Segregation', *American Sociological Review*, vol. 80, no. 4, pp. 814-42.

Grinberg, N, Naaman, M, Shaw, B & Lotan, G 2013, 'Extracting Diurnal Patterns of Real World Activity from Social Media', in.

Handcock, MS, Hunter, DR, Butts, CT, Goodreau, SM & Morris, M 2008, 'statnet: Software tools for the representation, visualization, analysis and simulation of network data', *Journal of statistical software*, vol. 24, no. 1, p. 1548.

Hochman, N & Manovich, L 2013, 'Zooming into an Instagram City: Reading the local through social media', *First Monday*, vol. 18, no. 7.

Hoff, PD, Raftery, AE & Handcock, MS 2002, 'Latent space approaches to social network analysis', *Journal of the american Statistical association*, vol. 97, no. 460, pp. 1090-8.

Hoselitz, BF 1953, 'The role of cities in the economic growth of underdeveloped countries', *The Journal of Political Economy*, pp. 195-208.

Hughes, G 1999, 'Urban revitalization: the use of festive time strategies', *Leisure Studies*, vol. 18, no. 2, pp. 119-35.

Jiang, S, Joseph Ferreira, J & Gonzalez, MC 2012, 'Discovering urban spatial-temporal structure from human activity patterns', paper presented to Proceedings of the ACM SIGKDD International Workshop on Urban Computing, Beijing, China.

Koski, SE 2011, 'Social personality traits in chimpanzees: temporal stability and structure of behaviourally assessed personality traits in three captive populations', *Behavioral Ecology and Sociobiology*, vol. 65, no. 11, pp. 2161-74.

Küller, R 2002, 'The Influence of Light on Circarhythms in Humans', *Journal of PHYSIOLOGICAL ANTHROPOLOGY and Applied Human Science*, vol. 21, no. 2, pp. 87-91.

Lefebvre, H 2004, *Rhythmanalysis: Space, time and everyday life*, A&C Black.

Lichter, DT 1988, 'Racial Differences in Underemployment in American Cities', *American Journal of Sociology*, vol. 93, no. 4, pp. 771-92.

Macy, MW & Willer, R 2002, 'From factors to actors: Computational sociology and agent-based modeling', *Annual review of sociology*, pp. 143-66.

Massey, DS, & Denton, NA 1988, 'Suburbanization and Segregation in U.S. Metropolitan Areas', *American Journal of Sociology, 94*(3), 592-626.

Mejova, Y, Weber, I & Macy, MW 2015, *Twitter: a digital socioscope*, Cambridge University Press.

Molotch, H 1976, 'The City as a Growth Machine: Toward a Political Economy of Place', *American Journal of Sociology*, vol. 82, no. 2, pp. 309-32.

Murthy, D. & Gross, A. J. 2017, 'Social media processes in disasters: Implications of emergent technology use', *Social Science Research*, vol. 63, pp. 356-370.

Murthy, D. 2018, Twitter: Social Communication in the Twitter Age. Cambridge, UK: Polity Press, 2018.

Pan, B, Zheng, Y, Wilkie, D & Shahabi, C 2013, 'Crowd sensing of traffic anomalies based on human mobility and social media', in *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 344-53.

Peterson, RD & Krivo, LJ 1993, 'Racial Segregation and Black Urban Homicide', *Social Forces*, vol. 71, no. 4, pp. 1001-26.

Podsakoff, PM, MacKenzie, SB & Podsakoff, NP 2012, 'Sources of method bias in social science research and recommendations on how to control it', *Annu Rev Psychol*, vol. 63, pp. 539-69.

Quetelet, LAJ 2013, *Treatise on man and the development of his faculties*, Cambridge: Cambridge University Pres.

Rawson, H & Miner, M 2006, *The Oxford dictionary of American quotations*, Oxford University Press.

Redfield, R & Singer, MB 1954, 'The Cultural Role of Cities', *Economic Development and Cultural Change*, vol. 3, no. 1, pp. 53-73.

Rocco, LD, Buscaldi, D, Bertolotto, M, Catania, B & Guerrini, G 2019, 'The role of geographic knowledge in sub-city level geolocation', paper presented to Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, Limassol, Cyprus.

Sagl, G, Resch, B, Hawelka, B & Beinat, E 2012, 'From social sensor data to collective human behaviour patterns: Analysing and visualising spatio-temporal dynamics in urban environments', in *Proceedings of the GI-Forum 2012: Geovisualization, Society and Learning*, pp. 54-63.

Sallis, JF, Haskell, WL, Wood, PD, Fortmann, SP, Rogers, T, Blair, SN & Paffenbarger, RS, Jr. 1985, 'Physical activity assessment methodology in the Five-City Project', *Am J Epidemiol*, vol. 121, no. 1, pp. 91-106.

Sampson, RJ & Raudenbush, SW 1999, 'Systematic social observation of public spaces: A new look at disorder in urban Neighborhoods 1', *American Journal of Sociology*, vol. 105, no. 3, pp. 603-51.

Sassen, S 2002, *Global networks, linked cities*, Routledge, New York ; London.

Sassen, SJ 2011, *Cities in a World Economy*, Sage Publications, Thousand Oaks; London.

Savage, M 2011, 'The Lost Urban Sociology of Pierre Bordieu', in G Bridge & S Watson (eds), *The new Blackwell companion to the city*, Wiley-Blackwell, Oxford.

Schmid, CF, MacCannell, EH & Van Arsdol, MD 1958, 'The Ecology of the American City: Further Comparison and Validation of Generalizations', *American Sociological Review*, vol. 23, no. 4, pp. 392-401.

Schnore, LF 1963, 'The Socio-Economic Status of Cities and Suburbs', *American Sociological Review*, vol. 28, no. 1, pp. 76-85.

Sennett, R 2011, 'Reflections on the Public Realm', in G Bridge & S Watson (eds), *The new Blackwell companion to the city*, Wiley-Blackwell, Oxford, pp. 380-7.

Silva, TH, Vaz de Melo, PO, Almeida, JM, Salles, J & Loureiro, AA 2013, 'A comparison of foursquare and instagram to the study of city dynamics and urban social behavior', in *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, p. 4.

Simmel, G & Hughes, EC 1949, 'The Sociology of Sociability', *American Journal of Sociology*, vol. 55, no. 3, pp. 254-61.

Simmel, G 1950, 'The Metropolis and Mental Life', *The Sociology of Georg Simmel. London: The Free Press of Glencoe Collier-MacMillian Ltd*.

Simon, RM 2011, 'Habitus and Utopia in Science: Bourdieu, Mannheim, and the Role of Specialties in the Scientific Field', *Studies in Sociology of Science*, vol. 2, no. 1, pp. 22-36.

Soukup, C 2006, 'Computer-mediated communication as a virtual third place: building Oldenburg's great good places on the world wide web', *New Media & Society*, vol. 8, no. 3, pp. 421-40.

Stahura, JM & Sloan, JJ 1988, 'Urban Stratification of Places, Routine Activities and Suburban Crime Rates', *Social Forces*, vol. 66, no. 4, pp. 1102-18.

Stephen, RGJ 1992, 'Was There a Hawthorne Effect?', *American Journal of Sociology*, vol. 98, no. 3, pp. 451-68.

Thompson, EP 1967, 'Time, work-discipline, and industrial capitalism', *Past and present*, pp. 56-97.

Thrift, N 2004, 'Intensities of Feeling: Towards a Spatial Politics of Affect', *Geografiska Annaler. Series B, Human Geography*, vol. 86, no. 1, pp. 57-78.

Tolnay, SE, Adelman, RM, & Crowder, KD 2000. '"Narrow and Filthy Alleys of the City"?: The Residential Settlement Patterns of Black Southern Migrants to the North', *Social Forces, 78*(3), 989-1015. doi:10.1093/sf/78.3.989

U.S. Census Bureau N/A, *DataFerrett*, <http://dataferrett.census.gov/>.

van Poppel, F & Day, LH 1996, 'A Test of Durkheim's Theory of Suicide--Without Committing the "Ecological Fallacy"', *American Sociological Review*, vol. 61, no. 3, pp. 500-7.

Wajcman, J 2008, 'Life in the fast lane? Towards a sociology of technology and time', *The British journal of sociology*, vol. 59, no. 1, pp. 59-77.

Wajcman, Ja 2015, *Pressed for time : the acceleration of life in digital capitalism*, University Of Chicago Press, Chicago.

White, MJ 1983, 'The Measurement of Spatial Segregation', *American Journal of Sociology*, vol. 88, no. 5, pp. 1008-18.

Wirth, L 1938, 'Urbanism as a Way of Life', *American Journal of Sociology*, vol. 44, no. 1, pp. 1-24.

Author Biographies

Dhiraj Murthy is an associate professor in the School of Journalism and Media and the Department of Sociology at the University of Texas at Austin, where he also directs the Computational Media Lab. Murthy's research explores social media, digital research methods, computational social science, qualitative/mixed methods, race and ethnicity, and virtual organizations.

John O'Brien is an associate professor of mathematics at Bowdoin College. O'Brien's research connects Bayesian statistics, computational approaches, genetics and ecology to construct new methods of biological analysis.

Alexander Jones Gross is a technologist and was an Associate and Programmer at the Social Network Innovation Lab and researcher in Intermedia Arts at University of Maine, Orono. His work specializes in natural language processing and computational social science.

Nathan Meyers is an applied mathematician and data scientist at Bevi.

City	Tweet.counts	pop	age	english	spanish	white	black	amerindian	asian	grad	doctoral	bachelors	high	less	citizen	poverty	income
Albuquerque	115740	712937	35.3	0.57	0.34	0.73	0.03	0.07	0.02	0.08	0.01	0.11	0.2	0.3	0.95	292.7	47366.47
Atlanta	1200830	3987990	33.1	0.8	0.1	0.63	0.3	0.01	0.04	0.07	0.01	0.15	0.18	0.31	0.93	334.3	56921.91
Austin	416541	1167216	32.1	0.64	0.26	0.75	0.08	0.01	0.04	0.08	0.01	0.17	0.14	0.28	0.91	325.1	54341.31
Boston	913906	4809251	36.7	0.71	0.08	0.85	0.07	0.01	0.05	0.1	0.01	0.15	0.2	0.27	0.92	345	85829.37
Boulder	49920	2481349	34	0.74	0.17	0.83	0.05	0.02	0.03	0.08	0.01	0.16	0.17	0.28	0.92	339.1	70432.18
Charlotte	503307	1337705	34.7	0.84	0.09	0.74	0.21	0.01	0.02	0.05	0	0.14	0.19	0.32	0.95	325.5	51706.73
Chicago	1915317	8277152	34.5	0.66	0.19	0.68	0.19	0.01	0.05	0.07	0.01	0.13	0.18	0.32	0.9	332.3	68324.21
Cincinnati	311810	1552982	35.4	0.89	0.03	0.83	0.15	0.01	0.02	0.06	0.01	0.12	0.22	0.31	0.98	327.4	50480.64
Cleveland	783919	2948392	37.1	0.85	0.04	0.8	0.17	0.01	0.02	0.06	0.01	0.11	0.25	0.3	0.98	319	48729.66
Columbus	530005	1443293	34.3	0.86	0.04	0.82	0.14	0.01	0.03	0.06	0.01	0.14	0.21	0.28	0.97	326.3	50723.52
Dallas	1168239	5043876	32.5	0.67	0.24	0.71	0.14	0.01	0.04	0.06	0.01	0.13	0.17	0.33	0.89	318.5	44427.75
Denver	291034	2481349	34	0.74	0.17	0.83	0.05	0.02	0.03	0.08	0.01	0.16	0.17	0.28	0.92	339.1	70432.18
Detroit	931518	5150384	35.4	0.83	0.04	0.74	0.23	0.01	0.03	0.06	0.01	0.1	0.21	0.31	0.96	334.6	54671.98
ElPaso	229537	676220	32.2	0.16	0.79	0.77	0.03	0.01	0.01	0.03	0	0.07	0.16	0.42	0.84	224.7	26239.72
Fresno	145029	924612	32.3	0.47	0.4	0.59	0.06	0.03	0.08	0.03	0	0.07	0.16	0.41	0.85	238.5	43089.5
Houston	1206387	4663267	32.5	0.6	0.3	0.65	0.17	0.01	0.05	0.05	0.01	0.11	0.17	0.36	0.87	301.6	38976.97
Indianapolis	406276	1603021	35	0.89	0.05	0.83	0.15	0.01	0.01	0.06	0.01	0.12	0.24	0.31	0.98	329.9	46672.3
Jacksonville	333325	1101766	35.3	0.84	0.07	0.74	0.22	0.01	0.03	0.05	0	0.11	0.22	0.31	0.97	311.7	42598.17
KansasCity	243412	1073972	35.7	0.88	0.06	0.81	0.16	0.01	0.01	0.05	0	0.11	0.24	0.29	0.98	321.7	38551.77
LosAngeles	952608	1375174	34.9	0.65	0.23	0.75	0.1	0.02	0.06	0.04	0	0.08	0.22	0.32	0.89	307.7	51685.36
LasVegas	3874031	16375849	33.4	0.43	0.39	0.59	0.08	0.02	0.12	0.05	0.01	0.1	0.15	0.37	0.81	286.9	80907.62
Louisville	135841	699532	36.8	0.89	0.04	0.78	0.2	0.01	0.02	0.06	0	0.11	0.22	0.31	0.98	309.4	44832.53
Memphis	395079	893369	33.8	0.87	0.05	0.48	0.49	0.01	0.02	0.06	0.01	0.11	0.19	0.33	0.97	293.1	43380.67
Miami	1012929	3845904	37.3	0.41	0.45	0.73	0.22	0	0.02	0.06	0.01	0.1	0.19	0.35	0.78	285.2	50154.86
Milwaukee	488478	1684056	35.7	0.82	0.08	0.8	0.16	0.01	0.02	0.05	0	0.12	0.22	0.31	0.97	328.6	52134.82
Minneapolis	535979	2856295	34.3	0.83	0.05	0.88	0.06	0.01	0.04	0.07	0.01	0.16	0.18	0.26	0.95	357.8	57549.99
Nashville	442556	1234004	34.8	0.87	0.06	0.81	0.16	0.01	0.02	0.06	0.01	0.13	0.21	0.31	0.97	319.8	53097.65
NewOrleans	492037	1246651	35.1	0.83	0.07	0.56	0.4	0.01	0.03	0.05	0.01	0.1	0.21	0.36	0.98	268.4	40535.8
NewYork	1833849	12679672	36	0.53	0.22	0.6	0.22	0.01	0.08	0.09	0.01	0.12	0.19	0.33	0.85	307.9	89746.98
OklahomaCity	235431	892347	34.8	0.82	0.09	0.78	0.12	0.06	0.03	0.06	0.01	0.12	0.19	0.3	0.96	287	33180.47
Omaha	146847	584099	34	0.84	0.08	0.84	0.11	0.01	0.02	0.06	0.01	0.14	0.19	0.28	0.97	326.8	42835.08
Orlando	473560	1652742	35.8	0.7	0.19	0.78	0.15	0.01	0.03	0.05	0	0.12	0.21	0.3	0.93	306	43771.29
Philadelphia	1259983	3834217	36.8	0.79	0.07	0.72	0.22	0.01	0.04	0.08	0.01	0.13	0.23	0.31	0.96	321.9	52021.76
Phoenix	590275	3070331	34.5	0.66	0.25	0.8	0.04	0.02	0.03	0.06	0.01	0.11	0.17	0.31	0.89	310.3	52498.58
Pittsburgh	411737	2285064	39.6	0.88	0.02	0.9	0.09	0	0.01	0.06	0.01	0.12	0.28	0.27	0.99	306.8	37401.67
Portland	348765	1874608	35.4	0.77	0.11	0.86	0.03	0.02	0.05	0.06	0.01	0.13	0.18	0.28	0.92	318.9	67520.3
Raleigh	266144	1182869	33.9	0.81	0.09	0.71	0.23	0.01	0.03	0.09	0.02	0.17	0.16	0.28	0.93	328	57600.37

Richmond	255178	995112	35.8	0.86	0.05	0.66	0.31	0.01	0.02	0.07	0.01	0.14	0.2	0.3	0.97	328.1	49057.63
Rochester	113013	1030303	36.5	0.82	0.06	0.85	0.11	0.01	0.02	0.07	0.01	0.12	0.21	0.3	0.97	318.8	38804.54
Sacramento	246630	1802907	34.9	0.69	0.13	0.74	0.08	0.02	0.11	0.06	0.01	0.12	0.17	0.3	0.92	313.9	63515.71
SaltLakeCity	159408	1331833	31.3	0.77	0.13	0.9	0.01	0.01	0.03	0.05	0.01	0.11	0.18	0.31	0.94	319.3	61377.84
SanAntonio	551254	1551396	33.9	0.44	0.49	0.74	0.07	0.01	0.02	0.05	0	0.09	0.19	0.34	0.94	278.8	31772.09
SanDiego	608566	2807873	34.5	0.56	0.27	0.71	0.07	0.02	0.1	0.07	0.01	0.13	0.16	0.29	0.87	304	88002.51
SanFrancisco	1095772	7051730	35.9	0.54	0.19	0.63	0.08	0.02	0.2	0.1	0.01	0.17	0.14	0.28	0.85	353.7	132259.47
SanJose	339979	7051730	35.9	0.54	0.19	0.63	0.08	0.02	0.2	0.1	0.01	0.17	0.14	0.28	0.85	353.7	132259.47
Seattle	401765	3483448	35.3	0.77	0.06	0.83	0.06	0.02	0.1	0.07	0.01	0.15	0.18	0.26	0.94	340.4	78269.91
Tampa	213480	2386781	40.4	0.77	0.12	0.85	0.11	0.01	0.02	0.05	0	0.11	0.24	0.3	0.95	303.6	39451.36
Tucson	185083	843732	36.6	0.61	0.29	0.78	0.04	0.04	0.03	0.07	0.01	0.11	0.18	0.29	0.93	282	43814.87
Washington	548794	571753	36.4	0.72	0.12	0.32	0.61	0.01	0.03	0.14	0.02	0.14	0.17	0.29	0.91	280.7	80165.07

Table 1

Regressor	Effect	2.5%	97.5%	p-value
(Intercept)	4.720	2.87	5.73	< 2.16e-16
% English	-1.427	-5.63	0.03	0.0458
Age	0.413	-0.002	3.49	0.0292
Black	1.356	1.453	-1.123	0.1203
Distance	-0.0015	-0.0094	-0.00021	0.0004

Table 2

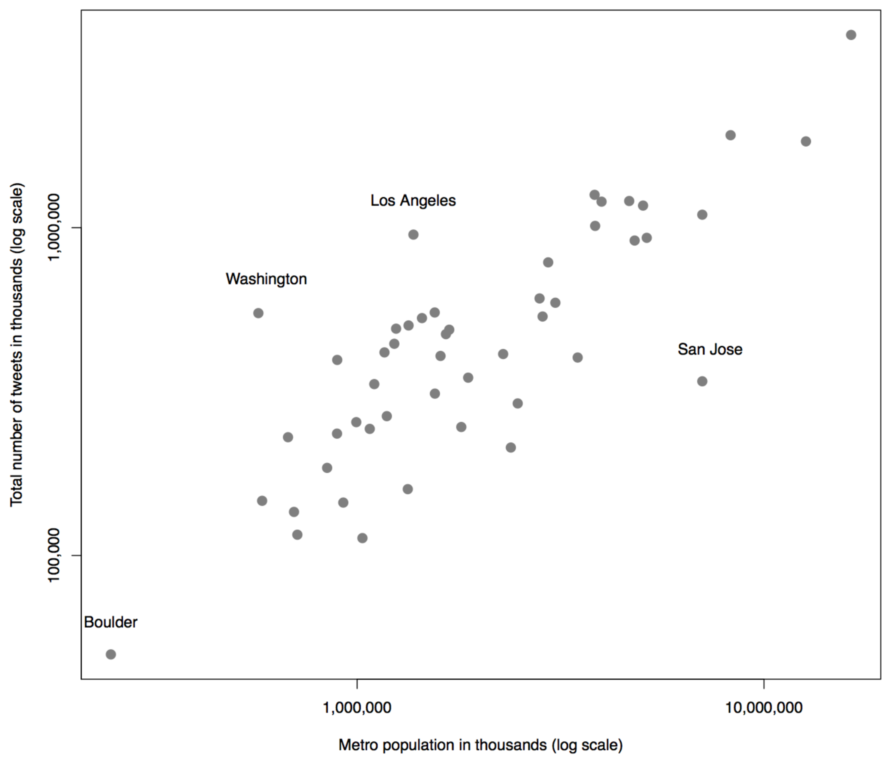


Figure 1

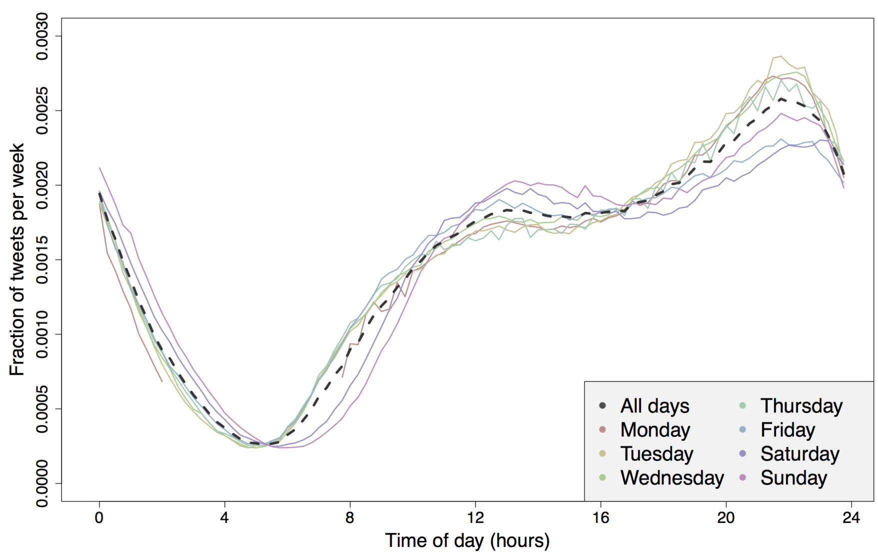


Figure 2

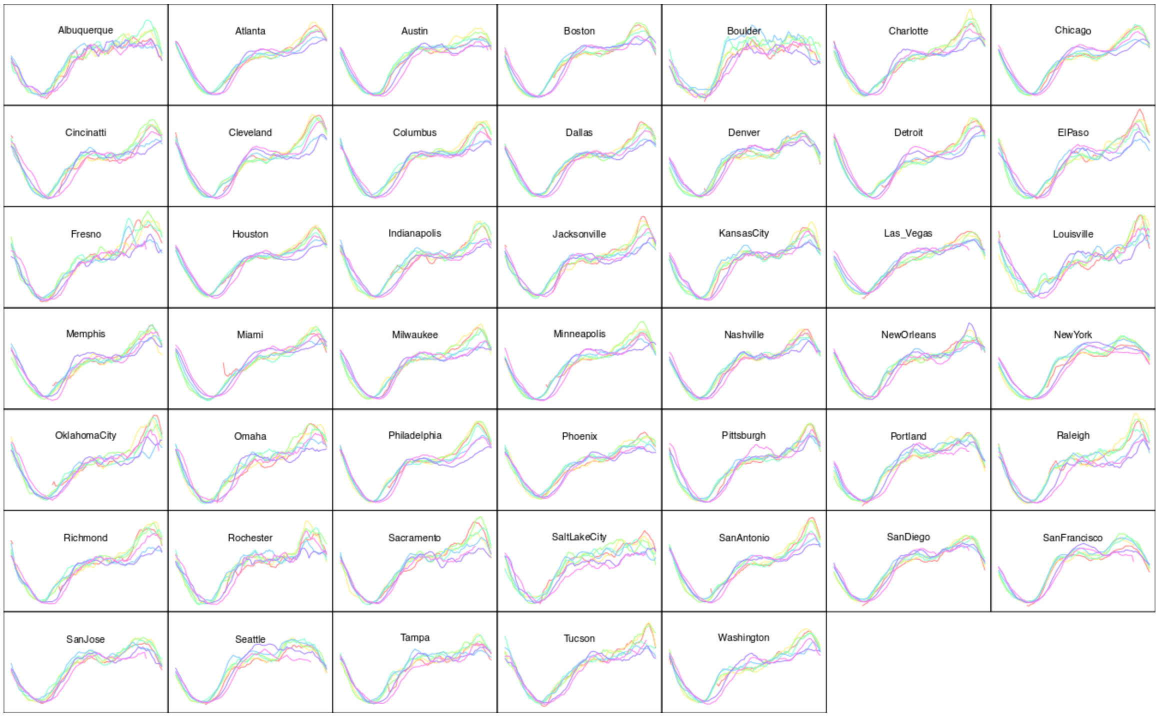

Albuquerque
Atlanta
Austin
Boston
Boulder
Charlotte
Chicago
Cincinatti
Cleveland
Columbus
Dallas
Denver
Detroit
ElPaso
Fresno
Houston
Indianapolis
Jacksonville
KansasCity
Las_Vegas
Louisville
Memphis
Miami
Milwaukee
Minneapolis
Nashville
NewOrleans
NewYork
OklahomaCity
Omaha
Philadelphia
Phoenix
Pittsburgh
Portland
Raleigh
Richmond
Rochester
Sacramento
SaltLakeCity
SanAntonio
SanDiego
SanFrancisco
SanJose
Seattle
Tampa
Tucson
Washington

Figure 3

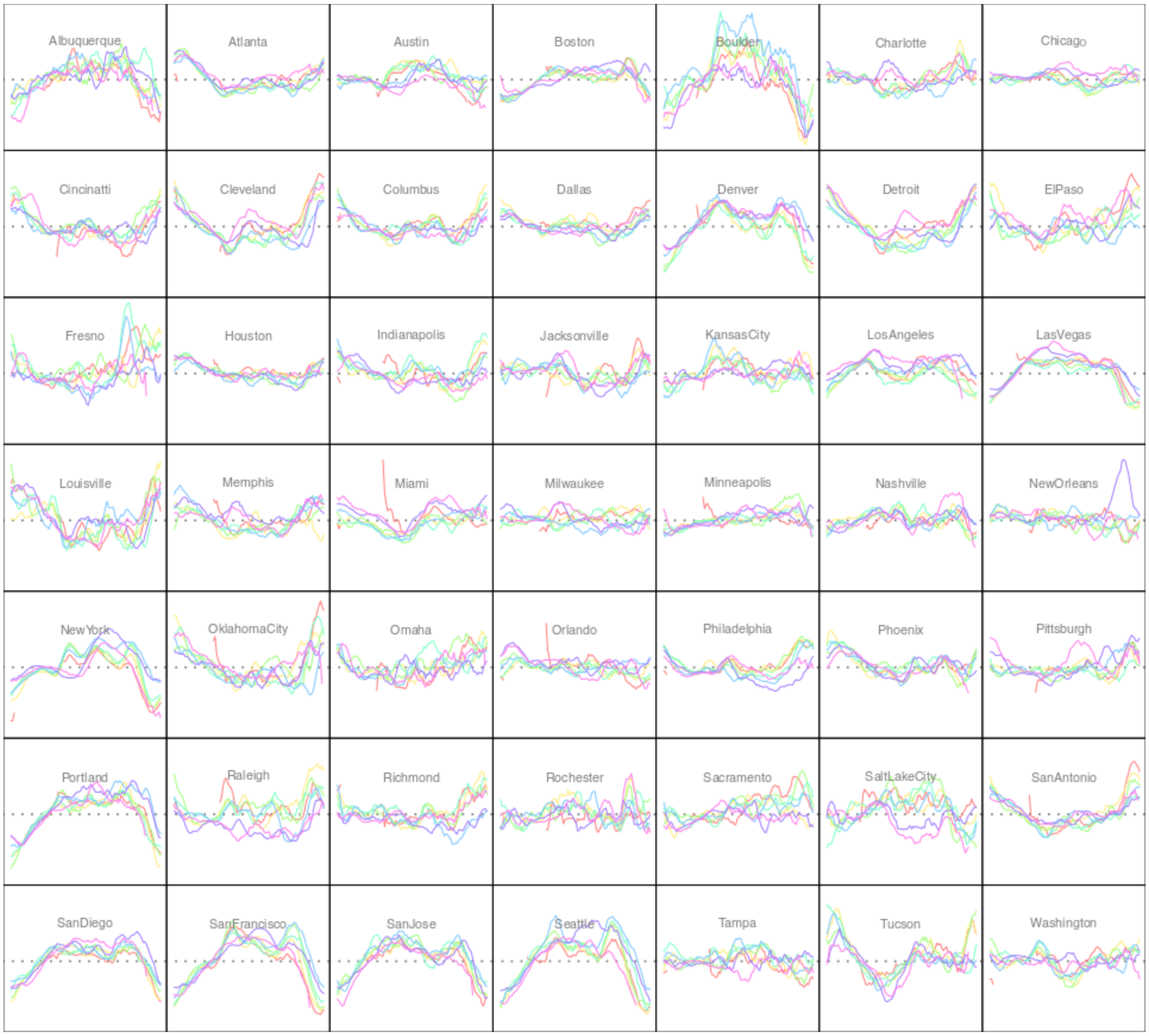

Albuquerque
Atlanta
Austin
Boston
Boulder
Charlotte
Chicago
Cincinatti
Cleveland
Columbus
Dallas
Denver
Detroit
ElPaso
Fresno
Houston
Indianapolis
Jacksonville
KansasCity
LosAngeles
LasVegas
Louisville
Memphis
Miami
Milwaukee
Minneapolis
Nashville
NewOrleans
NewYork
OklahomaCity
Omaha
Orlando
Philadelphia
Phoenix
Pittsburgh
Portland
Raleigh
Richmond
Rochester
Sacramento
SaltLakeCity
SanAntonio
SanDiego
SanFrancisco
SanJose
Seattle
Tampa
Tucson
Washington

Figure 4

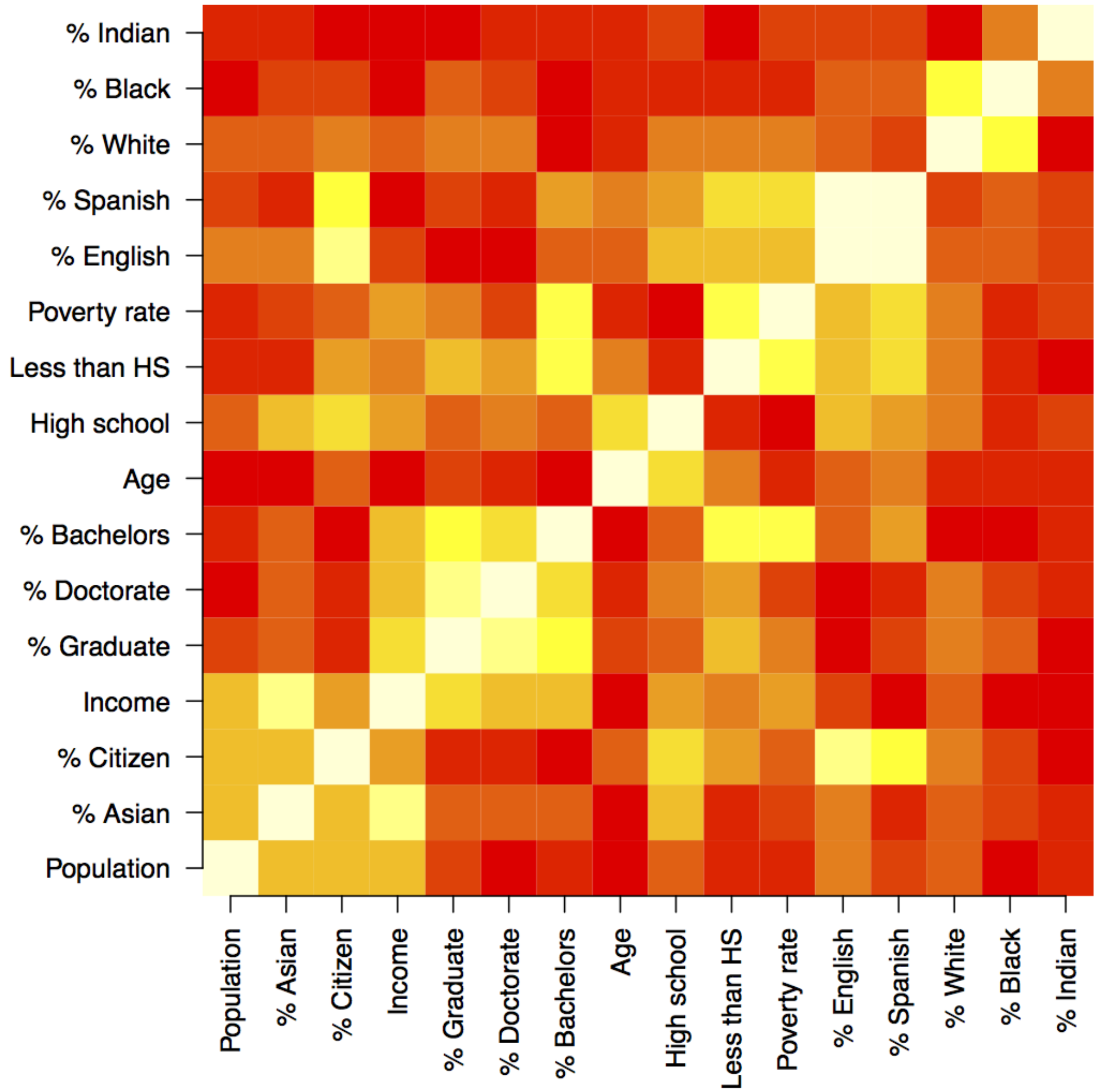
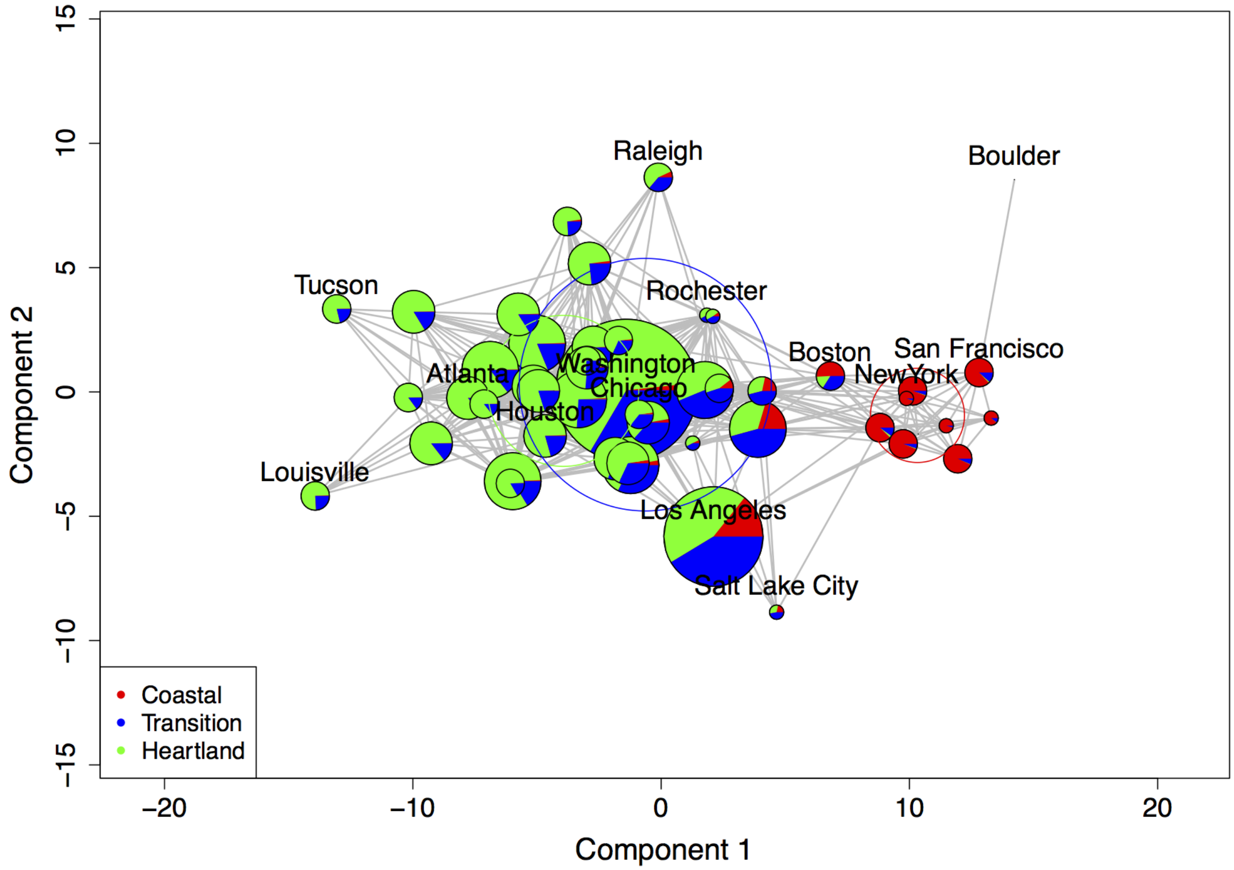

% Indian
% Black
% White
% Spanish
% English
Poverty rate
Less than HS
High school
Age
% Bachelors
% Doctorate
% Graduate
Income
% Citizen
% Asian
Population
Population
% Asian
% Citizen
Income
% Graduate
% Doctorate
% Bachelors
Age
High school
Less than HS
Poverty rate
% English
% Spanish
% White
% Black
% Indian

Figure 5

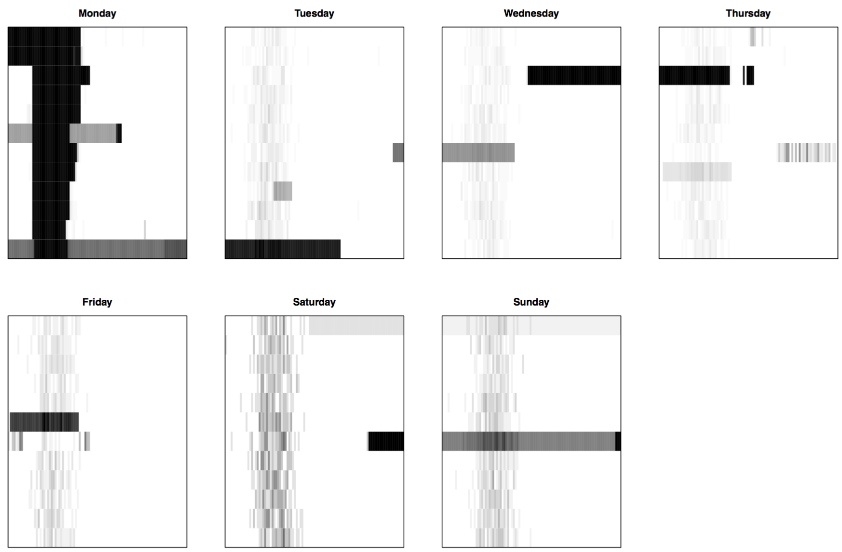


Raleigh
Boulder
Tucson
Rochester
Boston
San Francisco
New York
Atlanta
Washington
Chicago
Houston
Louisville
Los Angeles
Salt Lake City
Coastal
Transition
Heartland
Component 2
Component 1
15
10
5
0
−5
−10
−15
−20
−10
0
10
20

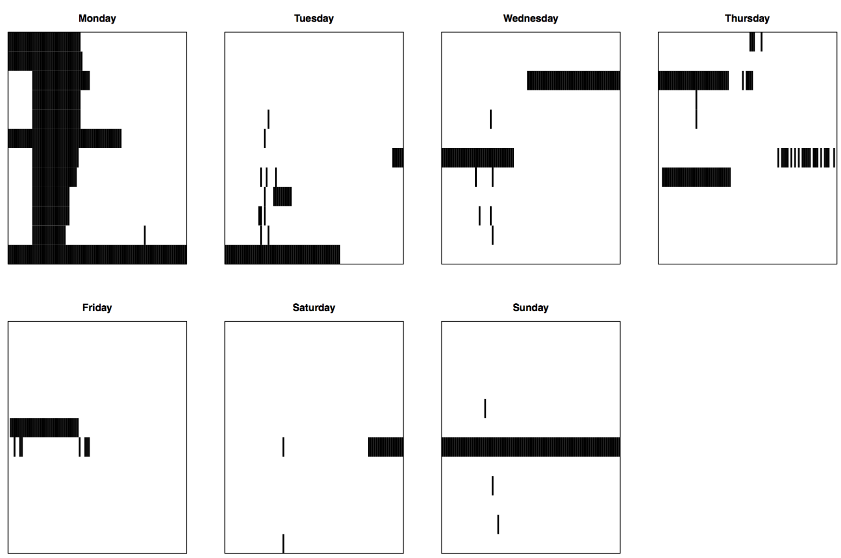
Figure 6


Seattle
Portland
Minneapolis
Milwaukee
Rochester
Detroit
Boston
Chicago
Cleveland
Omaha
SaltLakeCity
Boulder
Denver
Pittsburgh
NewYork
Philadelphia
Indianapolis
Columbus
Cincinnati
KansasCity
Washington
Sacramento
SanFrancisco
SanJose
Fresno
Richmond
LosAngeles
Nashville
Raleigh
Albuquerque
OklahomaCity
Memphis
Charlotte
LasVegas
Phoenix
SanDiego
Dallas
Tucson
ElPaso
Austin
Houston
NewOrleans
Jacksonville
SanAntonio
Orlando
Tampa
Miami

Figure 7



Appendix 1: Intensity of outages across the study period, organized by day (panels), weeks (rows within panels) and bins (columns within panels). Gray-scale shows frequency of zeros, with black indicating all cities exhibiting zeros and white indicate no cities exhibiting zeros.



Appendix 2: Identified outages using 7-city cut-off