Mining Spatial-Temporal *geoMobile* Data via Feature Distributional Similarity Graph

Arvind Narayanan, Saurabh Verma, Zhi-Li Zhang
Department of Computer Science & Engineering, University of Minnesota
Minneapolis, MN, USA
{arvind,verma,zhzhang}@cs.umn.edu

ABSTRACT

Mobile devices and networks produce abundant data that exhibit geo-spatial and temporal properties mainly driven by human behavior and activities. We refer to such data as *geoMobile* data. Mining such data to extract meaningful patterns that are reflective of *collective* user activities and behavior can benefit mobile network resource management as well as the design and operations of mobile applications and services. However, diverse feature distributions inherent in such data make such a task challenging. In this paper we advocate an approach based on advanced machine learning algorithms to transform original data matrices into a *feature distributional similarity graph* and extract "latent" patterns from complex structures of geoMobile data. Our analysis is further aided with a visualization technique. Using mobile access data records from an operational cellular carrier, we demonstrate the potentials of our proposed approach under multiple settings, and make some very interesting observations from the obtained results.

Keywords

call detail records; spatial-temporal; latent pattern extraction; feature distributions; similarity graph;

1. INTRODUCTION

Mobile technology has revolutionized the way how we live and interact with each other and the physical world. It was just back in mid-2014 when the number of active mobile devices surpassed the world population [1]. These mobile devices – together with the applications running on top and the underlying mobile infrastructures supporting them – have also enabled us to collect a whole gamut of (spatial-temporal) data that were possible otherwise. Combining such data arising from the cyber and/or physical worlds with *social data* arising from human actions and behavior allows us to gain a deeper understanding of the events and changes occurring in our environs. The real challenge today however is in designing methods and applying advanced analytics to extract *actionable* knowledge from such data and enable users to make sound and smart decisions. We coin the term *geoMobile* data to refer to datasets characterized by two salient features: i) they are associated with *geo*-locations (e.g. gathered by cell towers), and, ii) more often they capture human actions while *on-the-move*. For example, GPS-enabled smartphones produce feature-rich geoMobile datasets capturing human mobility and interactions with high spatial and/or temporal resolution. Evaluating human mobility can help unravel how populations move across regions. Human populations and their movements help understand carbon footprints and also aid in building smart grid energy networks by understanding the dynamically changing energy demands. With abundance of such geoMobile data, extracting meaningful patterns is an important yet non-trivial data analysis task that has wide applications, from network traffic engineering to urban transportation management, smart city planning, social behavior analysis, developing personalized services, geo-targeted mobile ads and cyber-physical world security.

In this paper, we aim to unravel underlying *latent* structures from a (spatial-temporal) geoMobile dataset driven by human mobility and behavior. Our dataset contains user-generated data (or Internet) access records registered by an operational cellular service provider covering approximately a 7000 sq. mile region in the San Francisco bay area. We try to ask questions such as: what is the impact of human mobility over cellular base-station (or tower) activities? If such an impact exists, is it diverse? If yes, is there a way to identify them from a large cellular network with millions of users and close to a thousand towers? Along with mobility, does time-of-the-day (temporal aspects) affect tower-level activities? If yes, is the nature of impact same for all towers? If not, how do we identify towers having similar impact due to human mobility and temporal factors? We believe understanding such underlying patterns driven by human mobility are vital to obtain insights about the events and changes that take place in the surrounding environment. For example, under future *"5G"* deployments, cellular service providers can get insights about tower-level activities to wheel out strategies for load balancing, dynamic provisioning of resources and transport layer networks, and other related services.

In the next section, we will briefly describe the dataset and illustrate the diversity inherent in such geoMobile data under multiple settings. We argue that classical approaches are ill-suited for handling high dimensionality and non-linearity apparent from the diverse nature of geoMobile datasets. Judicious feature engineering is therefore imperative. In § 3,

MobiData'16, June 30 2016, Singapore, Singapore

DOI: http://dx.doi.org/10.1145/2935755.2935760

we represent geoMobile data in the form of a matrix, where every row represents a data point and every column corresponds to a feature. We outline our proposed approach to obtain clusters from such (high-dimensional) data matrix through first converting the original data matrix to generate a feature distributional similarity graph and then applying spectral clustering to project the feature space into lower dimensional manifolds for latent significant feature extraction. To help visualize the resulting clusters, we leverage a density preserving mapping tool to embed the *"now"* labeled data points in a two (or three) dimensional space. In an attempt to answer the questions posed earlier, in § 4 we construct several data matrices from our geoMobile dataset under different settings, and show some initial yet promising results we have obtained applying our proposed approaches to these data matrices. We highlight some of the related work in § 5. The paper is concluded in § 6, where we highlight some ongoing work and future research directions.

2. DATASET & FEATURE DIVERSITY

In this section, we describe our dataset, present some preliminary analysis to show the diversity inherent in the data and highlight some challenges in analyzing the complex structures in geoMobile data.

2.1 Dataset

Our dataset consists of *anonymized cellular data access records* (2G/3G) collected from an operational cellular network. Also referred to as *call detail records (CDR)*, such data access records are collected by the cellular service providers for billing and other purposes. When a customer initiates a 3G-data session (e.g. by opening a website or could also be a background process) from a mobile phone, a communication link is established between the user's device and one of the radio base-stations that is within user's close geographical proximity. As soon as this session gets terminated, a record is generated at the radio-base station consisting of over 100 fields with information related to the customer, customer's device, and the session – such as customer identifiers, session start and end timestamps, bytes transmitted/received, and round trip time. Our dataset contains (anonymized) *per-session* records collected from over 900 base-stations (henceforth called towers) covering the San Francisco bay area for a period of 30 days, representing over a million users. Tower locations are known *a priori*, which we use as a proxy for user location. We use CDRs to study *collective* mobility patterns and *aggregate* activities of users at the tower level and their impact on the cellular infrastructure.

2.2 Diversity in Tower-Level Activities

A tower's volume of activity in a cellular network is often driven by the geo-characteristics of the tower's footprint (or coverage area). For example, a tower located in a highly populated downtown area may attract more activity than some other tower located in the suburbs. To further investigate, Figure 1 shows the number of unique users connected to every tower. It also shows the number of data access records that were generated by the towers. From this figure, we make two observations: 1) there is high diversity in the number of users connecting to towers, and, 2) to a lesser extent, there is also diversity in the number of records (or tower-level activity) among different towers. Obviously, human population plays a crucial factor. However, if geo-

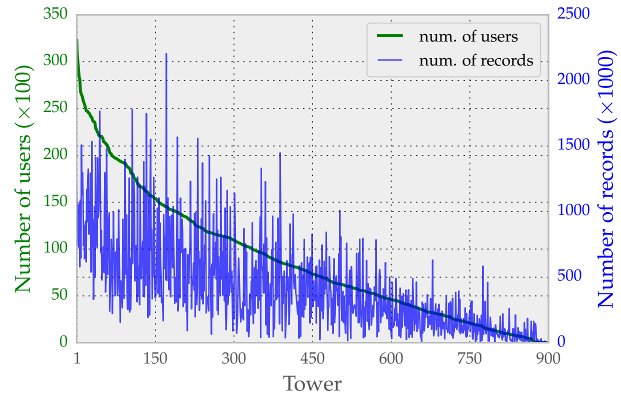


Figure 1. Volume of activity across all towers

characteristics indeed drives the tower's activity, it is intuitive to consider that, as populations move around a highly active tower for reasons such as grocery shopping, watching a movie, etc., it may cause its neighboring towers to also generate reasonable activity, and may further have a ripple effect across to other neighboring towers in the region. However, the question arises on how to identify such moving populations and its effect on the towers. We will get back to this discussion later in § 4.

Figure 2. Nine geographically dispersed towers selected from the San Francisco bay area region to study change in towel-level activities over time

Tower activities may also change depending upon the time of the day. To get temporal insights on tower activities, we randomly select 9 towers $(A, B, C, ..., I)$ covering fairly different regions across the San Francisco bay area (see Figure 2). We pick one week's data access records, and aggregate them to find the number of *unique users* seen at every hour for each of the nine towers. We show the results in Figure 3 and make three interesting observations. First, the temporal-pattern distributions vary significantly between the towers. While certain towers (such A, B, C) peak during the evening hours, others do not. Second, the volume of users connected to the towers also vary between towers. For example, throughout the day, we observe that tower A attracts more users than towers D, E, F, G, H, I. Similarly, to a lesser degree, tower F attracts more users than towers G, H, I. Third, even though

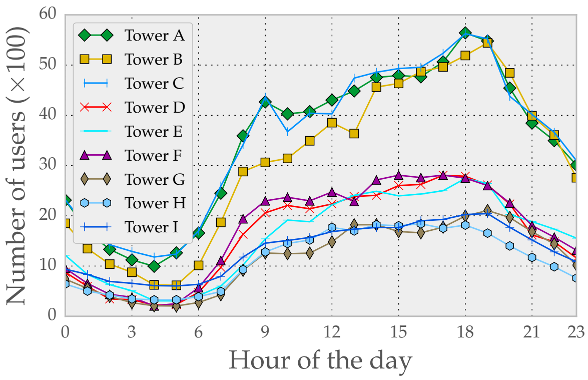


Figure 3. Distribution showing unique number of users seen by the nine selected towers in every hour

towers A and B are fairly located at different regions of the bay area, we see similarities in their temporal patterns. Similar observations can be made for other geographically dispersed towers as well. Given that there are more than 900 towers, the question arises on how to identify towers that possess such similar temporal behavior. This may become even more difficult if we consider a finer temporal granularity such as a combination of both the day (e.g. Mon/Tue, etc.) and the hour. In other words, we can formulate the problem such that towers can represent data points whereas time-of-the-day represents the feature space. Therefore, if we consider both day and hour, every data point will have a feature vector of length 168 (7 days $\times$ 24 hours), thus making it *high-dimensional*. Our goal is to cluster the data points (or towers) such that towers in the same cluster exhibit similar temporal structures. Next, we list certain properties of this problem that makes it challenging.

Depending upon the number of towers and temporal granularity, data can quickly grow large both in terms of data points and feature space. High dimensional feature space pose a challenging task for analysis due to the curse of dimensionality [4]. However, it is quite likely that there exist a lower dimensional latent feature space (or sub-manifolds of low intrinsic dimensions) corresponding to the observable feature space. Therefore, we can consider to extract latent feature space and then perform clustering to address the high-dimensionality issue. But in the process of extracting latent space, it may not be desired to preserve each characteristic of observable feature space and one can focus on a particular property (eg., geodesic distance, probability density) of the data based on the application.

Latent features can be a linear or non-linear function of raw features depending upon how the data is generated in observable space. Due to the high diversity observed in our data, it is reasonable to assume that the data is embedded non-linearly in observable space. In other words, our latent features would be a non-linear function of temporal or spatial observed feature-space distributions. This makes principal component analysis (PCA), a very popular technique used for analyzing high dimensional data, unsuitable for extracting latent features. From a geometrical point of view, PCA only captures orthogonal directions of maximum variance of the data without taking the curvature into account and thus loses many interesting properties of the data.

Even before dealing with extraction of latent features, it is important to discuss the feature construction process. Construction of features depends upon the pattern and activity of interest. For example, to get meaningful temporal patterns at each location, we can have feature space as time series while for spatial patterns, we can consider spatial features at each location. And to find spatio-temporal patterns, we can jointly consider time series and spatial features as a single feature space. We however emphasize that the *resolution* of such features and the *measure* quantifying the relation between the feature and data point should be decided *judiciously*. Finally to extract latent features, we favor feature extraction over feature selection process because the latter does not exploit the power of feature combination that could result in non-linear patterns in the data.

Once latent features are extracted, we can apply traditional clustering algorithms to obtain clusters in the data. But if the reduced latent features are greater than three, it is not easy to visualize the clusters limiting the interpretation of data from the human observer's perspective. Visualization is important in the sense that it provides the most convenient way to justify the quality of clustering results and gives insight about distribution of the data. In the next section, we employ machine learning techniques to overcome the challenges listed here and to extract meaningful patterns.

3. PROPOSED APPROACH

Generally, geoMobile datasets (e.g. human mobility data) have complex structures driven by a large set of *latent features* giving rise to various observable features such as peak traffic hours, different density of traffic volume across the regions etc. Such pattern diversities give rise to varying observable feature distributions. For example, as discussed earlier in § 2 we see certain towers peak at different times of the day than other towers. However, it is quite likely that these large number of latent features *(driven by varying feature distributions)* lie in *low-dimensional sub-manifolds* forming various kind of clusters in the data. This means that each cluster is formed by few latent factors which are a (non-linear) function of the observable features. Based on this intuition, we are interested in an approach that find clusters while accounting for possible latent features in the data. For this purpose, we consider *spectral clustering* due to its sound theoretical foundation and ability to handle high dimensional as well as to some extent non-linearity in the data. Other reasons of not directly applying standard clustering algorithm such as k-means is due its dependency on having clusters to form convex regions which may not be true for the geoMobile data containing multiple linear sub-manifolds. Also, these methods do not attempt to find latent features instead directly perform clustering on observable features. Next, we provide details about our approach of applying spectral clustering on a similarity graph obtained from the observable feature distributions.

3.1 Clustering

Spectral clustering works on a notion of *similarity graph*. For data given in the form of a similarity graph, spectral clustering partitions the graph such that the edges between different groups have a very low weight. It works by computing a graph Laplacian $\mathbf{L}$ and performing the eigenvalue decomposition of $\mathbf{L}$ matrix so as to extract d eigenvectors which forms low dimensional representation of the data. Geometrically, spectral clustering force nearby data points (or clusters) to map around d mutually orthogonal points on the surface of unit d-dimensional sphere. In this low dimensional space, points in the same cluster have similar (spatial or temporal) representation – data points that are close in observable feature space distributions are also close to each other (governed by their similarity value) in the latent space. Traditional clustering algorithm can then be applied.

Constructing Similarity Graph: Effectiveness of spectral clustering heavily depends upon computing the appropriate *adjacency* or *similarity matrix* $\mathbf{A}$ obtained for some data matrix, say $\mathbf{X}$, where every row in $\mathbf{X}$ indicates the feature vector associated for every data point. In other words, every column of $\mathbf{X}$ represents a feature – which can be a spatial, temporal, or an activity-based measure of the data point. For example, data matrix can have rows representing towers and columns as users where each entry could indicate the frequency of data access records as seen by the tower (row index) for some user (column index). However, coming up with a similarity graph is not an easy task and it depends upon the application and nature of the dataset. For obtaining temporally driven mobility patterns from data matrices, each row or column can represent some time series. In cases with spatial patterns, it can comprise of distribution of some abstract values across different regions. In both cases, we adopt a similarity metric which takes feature distributions into account. We believe *Gaussian (or heat) kernel* is suitable (as shown below) for this purpose, for which more theoretical motivation can be found in [3].

$$\mathbf{A}_{ij} = e^{-\frac{\|\mathbf{x_i}-\mathbf{x_j}\|^2}{2\sigma_i^2}} \tag{1}$$

There are many interpretations of Gaussian kernel, from *kernel density estimation* (KDE) to representing conditional probability $p_{j|i}$ of picking x_j as the neighbor of x_i data point. Choosing σ in Gaussian kernel is not trivial and greatly affects the similarity between two points in the data. Instead of setting a constant σ for all data points, we choose to compute σ_i at each data point x_i (as density can be different in different regions) such that the entropy of distribution is given as:

$$-\sum_j p_{j|i} \log p_{j|i} = \log k \tag{2}$$

where k is a user defined perplexity parameter and can be interpreted as a smooth measure of effective number of neighbors. For calculating σ_i, we performed a binary search over its value so that the entropy of distribution becomes equal to $\log k$ for each data point. It turns out that similarity matrix is robust for different values of k and typical values lie in the range 5 – 50.

Constructing Graph Laplacian: We employ a *symmetric normalized graph Laplacian* version proposed in [8] as it is less susceptible to bad clustering when different clusters are connected with varied degree.

$$\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{A}\,\mathbf{D}^{-1/2} \tag{3}$$

Here $\mathbf{D}$ is the diagonal degree matrix whose elements are sum of the rows of similarity matrix. From eigen decomposition of $\mathbf{L}$, d largest eigenvectors are stacked as columns in a $\mathbf{Y}$ matrix which is renormalized to yield a low-dimensional representation of data in $\mathbb{R}^d$ space. There are several ways to estimate the intrinsic dimension d of the data (eg., kernel PCA) but graph Laplacian implicitly provides a way to estimate d through examining drop in eigenvalues of $\mathbf{L}$. A better approach to approximate intrinsic dimension can be found in [15]. For our dataset, Laplacian eigenmaps approach was sufficient enough to yield faithful results. For several of our data matrices, we observe there is an eigenvalue drop, pointing to the existence of $15 - 30$ intrinsic dimensions. After obtaining latent features $\mathbf{Y}$, we apply traditional clustering algorithms to obtain clusters. In this paper, we use *DBSCAN* due to its robustness against outliers.

3.2 Cluster Visualization

Visualization plays an important role in assessing the quality of clusters formed, getting insights to interpret the results and enabling human reasoning. One of the corollaries of Gauss Theorema Egregium [10] shows that manifolds (data) with *intrinsic curvature* cannot be mapped to the $\mathbb{R}^2$ plane (as it has zero Gaussian curvature) without distorting distances. However, no such obstruction exists for density preserving maps. For details we refer readers to Moser Theorem [9]. Hence, we seek a method that *preserves* (probability) density maps rather than distances for visualization purpose. *t-stochastic neighbor embedding algorithm* (t-SNE) [14] is such a technique that attempts to preserve joint probabilities of data points in lower dimensional space by minimizing the KL-divergence between the distributions. We apply t-SNE over the raw (observable) feature space and apply cluster labels to them that were obtained through spectral clustering as discussed in the earlier subsection.

From the observable space, t-SNE computes the symmetric conditional probability p_{ij} (also referred to as pairwise similarity) for each pair of data points y_i and y_j to obtain a joint probability distribution P given by:

$$p_{ij} = \frac{exp(-\frac{\|\mathbf{y_i}-\mathbf{y_j}\|^2}{2\sigma_i^2})}{\sum_{k\neq l} exp(-\frac{\|\mathbf{y_k}-\mathbf{y_l}\|^2}{2\sigma_k^2})} \tag{4}$$

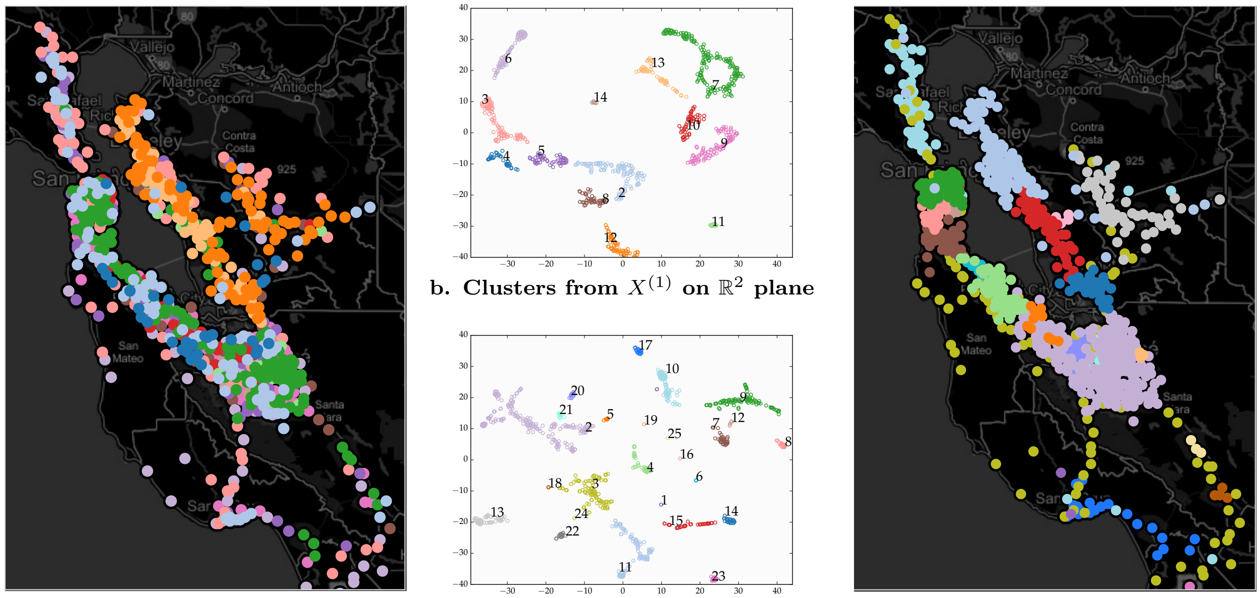
While in two (or three) dimensional space, t-SNE employs a *student t-distribution* with one degree of freedom to obtain the joint probability distribution Q, and pairwise similarity which is given by:

$$q_{ij} = \frac{(1+\|\mathbf{z_i}-\mathbf{z_j}\|^2)^{-1}}{\sum_{k\neq l}(1+\|\mathbf{z_k}-\mathbf{z_l}\|^2)^{-1}} \tag{5}$$

To obtain the final lower-dimensional coordinates $\mathbf{z}_*$ in $\mathbb{R}^2$, t-SNE minimizes the KL-divergence between P and Q probability distribution according to the objective function:

$$KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{6}$$

The above optimization problem is non-convex but gradient descent gives reasonable results for this problem. We refer readers to the original t-SNE paper [14] for further details. Compared to other visualization techniques [13, 12], t-SNE in general, works well for data with complex structures because it focuses on preserving probability densities rather than distances (though both quantities are inter-dependent). In the next section, we show the efficacy for our approach on multiple data matrices obtained in different settings.



b. Clusters from $X^{(1)}$ on $\mathbb{R}^2$ plane

a. Clusters from $X^{(1)}$ overlaid on a geographic map **c. Clusters from $X^{(2)}$ on $\mathbb{R}^2$ plane** **d. Clusters from $X^{(2)}$ overlaid on a geographic map**

Figure 4. Examples showing results obtained by applying the clustering approach described in § 3 on data matrices $X^{(1)}$ and $X^{(2)}$ *(best viewed in color)*

4. EXAMPLES & TACTICAL RESULTS

Earlier in § 2.2, we empirically observed that certain geographically dispersed towers have similar activity rates (i.e. distribution of number of unique users connected to them) across different hours of the day (see Figure 3). With around 900 towers, the question was how to identify such cluster of towers exhibiting similar temporal patterns. We construct a data matrix $X^{(1)}$, *s.t.*, rows represent towers, and columns represent hourly intervals across an entire week, and the matrix cell indicates the number of unique users seen by some tower for an hour. For example, $x^{(1)}_{i=towerA, j=\{MON,08\}}$ cell in $X^{(1)}$ indicates the number the unique users seen by tower A on Monday between 08:00 and 09:00 hours. Therefore, for every data point (or tower), we have a feature vector of size 168 (7 days of the week × 24 hours), making $X^{(1)}$ a matrix of size 900 × 168. We now apply the approach discussed in the previous section. We obtained a total of 12 clusters, and further by applying t-SNE visualize the obtained clusters in a 2-dimensional map (see Figure 4b). We also map the results (i.e. clustered towers) over the geographical map of San Francisco bay area (see Figure 4a). All the figures are best viewed in *color*. The results suggest there to be 12 distinct activity patterns across the region. Certain patterns such as clusters ●12 and ●13 exhibit a *locality effect*, whereas most of the other patterns travel across the region. We also find that clusters ●7 and ●10 were mostly located at the central zones (i.e. highly populated) regions. Looking back at the nine towers in Figure 3, we observed towers A, B, C to be grouped together into cluster ●7, towers D, E, F into ●10, and towers G, H, I into cluster ●9. The results, thus obtained, make sense as the towers grouped together also evidently exhibit similar feature (or *temporal* in this case) distributions across time in Figure 3.

In our preliminary analysis, we observed certain towers tend to attract more activity (i.e. generate high number of data access records) compared to others. As suggested earlier, intuitively, a tower located in a highly populated region will generate high volume of activity. As people from such a location are on the move (e.g. for grocery shopping), neighboring towers will start generating activity, which may then travel to other neighboring towers later which in some way causes a directional-bound ripple effect. We ask the question how to identify regions that show such ripple effects of tower-level activity. In other words, we try to identify if activities of certain towers are mainly driven by some subset of user population, if so, how their movements affect neighboring towers. We perform exploratory analysis by constructing a data matrix $X^{(2)}$, *s.t.*, data points (or rows) represent all the 900 towers, and the features (or columns) are individual users. Cell value in the matrix indicates the number of data access records generated by a particular user on some tower. We apply our approach on this data matrix $X^{(2)}$, and obtain close to 25 *interesting* clusters. Figure 4c shows the low-dimensional map of the data points in $\mathbb{R}^2$ plane. We further map the labeled (i.e. colored) data points on a geographic map (see Figure 4d). We highlight three of the many observations captured under this setting. First, we clearly find pockets of *regional* towers (such as in clusters ●2, ●4, ●15, etc.) that exhibit similarities in serving a certain subset of user population. In other words, we find boundaries of the extent of ripple effects arising from human mobility across distinct regional communities. Second, cluster ●3 travels across the entire region suggesting towers in this cluster are mainly located at transit zones. Further analysis of cluster ●3 suggest the activity of its towers are fairly less than others. Third, we find the geographic footprint of the regional clusters vary significantly. For example, clusters such as ●6, ●19, ●20, ●25 have smaller regional footprints than say clusters ●2, ●4, ●15. This suggests the clusters exhibit certain *locality-specific* effects of varying size. There are many other interesting observations however due to lack of space, we omit them here.

For both the examples mentioned in this section, we leave a more detailed analysis and interpretation of the results for a longer version of this paper. Nonetheless, it is interesting to see meaningful patterns emerging from such geoMobile data, especially when no location information was explicitly fed as input to our approach.

5. RELATED WORK

Analysis of geoMobile datasets emerging from mobile phones, public transportation networks and crowd sourced mobile platforms have been a popular area of research for a long time. In the past, several studies have been conducted using such datasets for purposes including, but not limited to, modeling and understanding human mobility [6, 16], constructing social networks and community detection [7], urban analytics [11] and understanding human factors or behavior [5]. With a goal to extract meaningful and actionable knowledge from geoMobile datasets, majority of the previous work have one or more of the following lacunas: 1) to extract meaningful patterns, classical approaches such as PCA (e.g. in [2]) and *non-negative matrix factorization* (e.g. in [16]) have been used extensively; however, a fundamental premise of both approaches are that the relations/correlations between entities in the data are linear; as argued earlier, such an assumption may not always hold true for geoMobile datasets due to the high diversity inherent in them, 2) another challenge is that the features of geoMobile datasets can be diverse as well – amalgam of spatial, temporal, user-based, and/or activity-based features may exist; certain latent patterns in the data may be a non-linear function of a subset of such diverse features, identifying such patterns are often arduous, and, 3) need for a promising approach that can extract meaningful latent patterns from any kind of geoMobile dataset and avoid the hassle of feature selection. This paper tries to address all of these challenges.

6. CONCLUSION

Using *anonymized* data access records from an operational cellular network as a case study, we argued that classical PCA and related linear methods are ineffective in mining diverse geoMobile data, due to inherent high variability, nonlinearity and skewed feature distributions. To address these challenges, we have proposed a novel approach where we first transform original data matrices into a *feature distributional similarity graph*, and then apply advanced machine learning algorithms and visualization techniques to help extract meaningful "latent" patterns from complex structures of geoMobile data. We demonstrated the potential of this approach by applying it to our dataset under multiple settings and observed some insightful patterns. As ongoing research, we will take a deeper look to understand the underlying theoretical guarantees of our proposed approach, and develop mechanisms to characterize the extracted clusters in order to interpret the results better. We also plan to apply this promising approach to other forms of geoMobile datasets from different application-domains.

7. ACKNOWLEDGMENTS

This research was supported in part by NSF grants CNS-1117536, CRI-1305237, CNS-1411636 and DTRA grant HDTRA1-14-1-0040 and DoD ARO MURI Award W911NF-12-1-0385.

8. REFERENCES

[1] GSMA Intelligence. http://gsmaintelligence.com. April, 2016.

[2] R. A. Becker, R. Cáceres, K. Hanson, J. M. Loh, S. Urbanek, A. Varshavsky, and C. Volinsky. Clustering anonymized mobile call detail records to find usage groups.

[3] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[4] C. M. Bishop. Pattern recognition. *Machine Learning*, 2006.

[5] J. Candia, M. C. González, P. Wang, T. Schoenharl, G. Madey, and A.-L. Barabási. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224015, 2008.

[6] S. Isaacman, R. Becker, R. Cáceres, M. Martonosi, J. Rowland, A. Varshavsky, and W. Willinger. Human mobility modeling at metropolitan scales. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 239–252. ACM, 2012.

[7] K. Kianmehr and R. Alhajj. Calling communities analysis and identification using machine learning techniques. *Expert Syst. Appl.*, 36(3):6218–6226, Apr. 2009.

[8] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pages 849–856, 2002.

[9] A. Ozakin, N. V. II, and A. Gray. Manifold learning theory and applications - density preserving maps. pages 57–71, 2011.

[10] A. Pressley. *Elementary Differential Geometry.* Springer, 2010.

[11] A. Psyllidis, A. Bozzon, S. Bocconi, and C. T. Bolivar. A platform for urban analytics and semantic data integration in city planning. In *Computer-Aided Architectural Design Futures. The Next City-New Technologies and the Future of the Built Environment*, pages 21–36. Springer, 2015.

[12] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[13] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290, 2000.

[14] L. van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research (JMLR)*, 2008.

[15] L. Zelnik-manor and P. Perona. Self-tuning spectral clustering. In *Advances in Neural Information Processing Systems*, pages 1601–1608, 2005.

[16] D. Zhang, J. Huang, Y. Li, F. Zhang, C. Xu, and T. He. Exploring human mobility with multi-source data at extremely large metropolitan scales. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 201–212. ACM, 2014.