

Manifold Learning of the Dominant Modes of Human Mobility

James R. Watson^{a,2}, Zach Gelbaum^{b,2}, Mathew Titus^b, Grant Zoch^a, and David Wrathall^a

^aCollege of Earth, Ocean and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA, 97330; ^bThe Prediction Lab LLC, Corvallis, Oregon, USA, 97330

This manuscript was compiled on January 8, 2019

When, where and how people move is a fundamental part of how human societies organize around every-day needs as well as how people adapt to risks, such as economic scarcity or instability, and natural disasters. Our ability to characterize and predict the diversity of short- and long-term human mobility patterns has been greatly expanded by the availability of Call Detail Records (CDR) from mobile phone cellular networks. These data can be used to quantify mobility patterns across a range of spatial (from local to global) and temporal scales (from minutes to years). The size and richness of these datasets is at the same time a blessing and a curse: while there is great opportunity to extract useful information from these datasets, it remains a challenge to do so in a meaningful way. In particular, human mobility is multi-scale, meaning a diversity of patterns of mobility occur simultaneously, which vary according to timing, magnitude and spatial extent. To identify and characterize the dominant modes of human mobility that emerge over time and space we examined CDR data from the Orange mobile network in Senegal using spectral graph wavelets, a form of Manifold Learning. This unsupervised data mining approach reduces the dimensionality of the mobile data to reveal key seasonal changes in human mobility, as well as mobility associated with large-scale punctuated religious and political events. This is achieved through a localized, multiscale decomposition of a dynamically evolving mobility network, identifying the dominant scales of importance at a given location, and tracking the changing features of the associated wavelet function at those scales. The novel insight on the range of human mobility patterns afforded by Manifold Learning methods like spectral graph wavelets have clear applications for urban planning, infrastructure design as well as hazard risk management, especially as climate change alters the biophysical landscape on which people work and live driving new patterns of migration around the world.

Manifold Learning | Human Mobility | Complex System | Dimension Reduction | Prediction | Geography

Introduction

Human mobility is a fundamental part of how individuals, households and communities organize to meet every-day needs, and to respond to infrequent risks and shocks like economic instability and environmental hazards. Human mobility is multi-scale in nature, that is for any given type of mobility, such as commuting, seasonal migration or holiday travels, individuals move as part of social collectives of varying size and interconnectivity, which span different magnitudes of spatial and temporal scale. Human mobility also has multiple modes of variability: people go to work each day,

8 they go on holiday during specific programmed periods within the year, they may migrate before and
9 after key agricultural seasons, or they may evacuate during floods or other environmental hazards.
10 For these reasons and others, it is a continuing challenge to identify, categorize and anticipate the
11 various simultaneous patterns of human mobility. Anticipating and planning for human mobility is
12 a non-trivial task for organizations whose core functions provide critical services to and address the
13 needs of moving people, such as urban planning and transport agencies, disaster first-responders
14 and international aid organizations.

15 To overcome these challenges and generate fundamental insight on human mobility, research is
16 leveraging novel data generated by users of the digital infrastructure, e.g. mobile phone subscribers.
17 So-called Big Data, routinely collected from a range of sources, most notably the explosion of mobile
18 phone usage throughout the world, provides rich information on users' locations through time.
19 Mobile network operators collect records of their users' calling patterns, a type of data called Call
20 Detail Records (CDR), which include the location of the receiving tower where each voice call or
21 text message is made, as well as the location of the recipient. Over time, each user's calling patterns
22 can be used to reconstruct a detailed record of their location history. The collective mobility history

Significance Statement

Mobile phone data can provide fundamental insight on the diversity of patterns of human mobility that unfold simultaneously over different spatial and temporal scales as people go about their every-day lives and as they respond to economic and environmental risk. The challenge is that these data are large and difficult to use and interpret, and any signal of a fundamentally important mobility pattern can be lost in the noise. To overcome this challenge, we have developed a new approach to human mobility data analysis using tools from applied mathematics. Employing spectral graph wavelets, we were able to extract the signal of different modes of human mobility, based on changes in overall mobility network scale. We were able to identify and characterize various contemporaneous and spatially overlapping dominant modes of human mobility. This novel approach to characterizing and predicting changes in human mobility has practical applications for urban planning, environmental risk management, and humanitarian intervention.

Z.G. and M.T. performed the mathematical analysis. Z.G. and J.W. performed the data analysis and wrote the paper. G.Z. and D.W. provided the data and contributed to the design of the analysis and the writing.

The authors declare no conflict of interest.

²To whom correspondence should be addressed. E-mail: zach@thepredictionlab.com; jrwatson@coas.oregonstate.edu

23 of all users' movements through time provides insight on total population flows between all cellular
24 network locations during any specified period of time. This enables the study of users' behaviors at
25 very high spatiotemporal resolution over both local and system-level spatial scales at time scales
26 of minutes to months to years. As each phone is embedded within an existing social fabric, CDR
27 allow the analysis of the changing structure of social organization as people (i.e. individuals, social
28 networks, communities, religious and ethnic groups, etc.) respond to a diversity of stimuli. CDR data
29 has been used to describe and evaluate commuting (REFS), vacationing (REFS), and environmental
30 displacement resulting from earthquakes and hurricanes (Bengtsson et al. 2012; Lu et al. 2016).
31 They have been used to evaluate mobility as a vector for the spread of infectious diseases such as
32 malaria (REF).

33 CDR data offer a vastly clearer and more detailed description of the communication and mobility
34 behaviors of people as they go about their daily lives, compared to traditional data on human
35 mobility, such as surveys and censuses, which are collected at relatively coarse time-scales (i.e.
36 months and years; REF). CDRs are compiled by network operators principally for the purposes
37 of billing customers for their use of the network, not for scientific analysis; therefore, the main
38 challenge with analyzing CDR is the size, complexity and richness of datasets. CDR are inherently
39 high dimensional and noisy, and quantitative analyses must incorporate dimensional reduction and
40 denoising approaches. Once done, the remaining challenge is retrieving a relevant signal from the
41 data at the appropriate spatial and temporal scale for each specific mobility pattern.

42 Here, we describe a data-driven approach to CDR analysis that explicitly addresses the multi-scale
43 nature of the mobility patterns embedded in the data and reflected in the system under study. We
44 analyzed CDR from users of the Orange/Sonatel cellular network, collected in Senegal between 1
45 January and 31 December 2013 (see Fig. 1 for an illustration of mobility in Senegal for a given day).
46 De-identified data entries included information on the time of the call, the mobile phone tower used
47 and the duration of call. To identify and characterize different modes of human mobility captured in
48 CDR, we developed a novel computational approach based on Spectral Graph Wavelets (SGWs),

49 an extension of classical wavelet analysis to the setting of networks. New mathematical analysis
50 involved the development of a specific wavelet kernel (REF) which was then applied to the cellular
51 data, decomposing these raw data into a set of functions describing the multiscale organization of
52 mobility. With this new approach we focused on dynamic multiscale mobility patterns to/from
53 a specific city - Touba, a market town and religious center in Senegal's agricultural breadbasket.
54 Our goal was to identify and characterize the different modes of human mobility that contrast in
55 overall spatial scale and temporal duration, focusing on inter-city (i.e. regional) forms of mobility
56 including seasonal migration relating to changes in agriculture and a mobile labor force, as well
57 as punctuated patterns relating to calendar festivals, local elections, or religious holidays. The
58 identification and interpretation of these dominant modes, and their effective scale, is novel and is of
59 value in extracting key dynamics from complex data sets not limited to those given by CDR.

60 Human Mobility Data

61 To analyze the multiscale nature of human mobility, we developed and implemented the SGW
62 method for de-identified, pre-processed extracts of CDR from the Orange/Sonatel mobile network
63 in Senegal generated between 1 January and 31 December 2013. These data include a number of
64 variables on individual mobility behaviors/locations for 300,000 randomly sampled users on a rolling
65 2-week basis. These data were arranged into a pairwise origin-destination network whose vertices, or
66 nodes, were cell towers and whose edges were weighted by the volume of phones flowing between
67 each tower pair for each 24-hour period over the entire year of 2013. To highlight inter-city mobility,
68 we first aggregated the towers of key cities into single nodes. Then, for each 24 hour period, a
69 one was added to a given edge i, j if a phone moved from node i to node j . To further filter out
70 intra-city mobility and patterns of daily commuting, repeated trips between the same towers for the
71 same user within a 24-period were not counted. Thus, mobility was defined and quantified as an
72 asymmetric affinity matrix A varying through time, where $A_{ij}(t)$ is the total number of unique trips
73 made by users between towers i and j on day t of 2013 (see Fig. 1A for a geographic representation

74 for one of these daily affinity matrices). In order to normalize the data so that the signal from high
75 volume mobility in areas such as Dakar, the capital of Senegal, did not wash out all other signals,
76 we calculated relative mobility by dividing the entries of $A(t)$, the daily total mobility, by their
77 respective row sums. In order to perform the SGW analysis, we then converted this asymmetric
78 affinity matrix to a symmetric one by taking the average of two-way mobility patterns between
79 pairs of cell-towers. Last, so as to not include intra-city traffic, we subtracted phones making calls
80 within tower cells (i.e. we set the diagonal of A to zero). The result was a symmetric affinity
81 matrix/network with zeros on the diagonal for every day of 2013.

82 **Multiscale Analysis using Spectral Graph Wavelets**

83 Networks of human mobility and communication are naturally modeled by graphs, sets of nodes
84 connected by edges, each with an associated weight representing the strength of the connection.
85 In this work, higher edge weights indicate a greater flow of human migration between the edge's
86 two endpoints. Graphs can be viewed as discrete analogues of smooth objects such as geometric
87 manifolds and surfaces, and the geometric content of these graphs is often analyzed using an
88 associated Laplacian operator (REFS), as it encodes the geometric structure of the graph in a
89 natural way. We will use such an approach in the present study.

90 Given that human mobility networks are inherently multiscale as well as highly heterogeneous,
91 we require a method of analysis that is both multiscale and localized. In the classical settings
92 of time series analysis and image analysis, wavelets are the proper tool for conducting localized,
93 multiscale decompositions and analysis. Here we employed and advanced a generalization of wavelets
94 to graphs, Spectral Graph Wavelets. One caveat to using wavelets is that it is up to the user to
95 choose the analyzing *wavelet kernel*. The wavelet kernel essentially defines its shape, and this choice
96 can have profound effects on the analysis. In order to successfully use SGWs to extract meaningful
97 information, we have developed a wavelet analysis based specifically on the heat kernel yielding
98 what we call Hermitian Graph Wavelets.

99 By defining our wavelet kernel in terms of the heat kernel, we are able to produce an analysis
100 that faithfully and efficiently extracts key geometric information and presents it through time. The
101 outputs of the analysis are a set of wavelet functions associated to each vertex (i.e. cell tower in
102 Senegal) in the graph representing the local multiscale geometry at that vertex, as well as a single
103 dominant scale for each vertex representing the scale containing the most information for that vertex.
104 Because these multiscale wavelet functions are rich in geometric content, by performing clustering
105 and other analyses on them we were able to detect, characterize and distinguish between different
106 regimes of human mobility through time.

107 **Applying Spectral Graph Wavelets to Mobility Networks.** Given the set of daily affinity matrices,
108 $A(t)$, the application of Spectral Graph Wavelets is as follows. First, for each of the daily affinity
109 matrices, we constructed their associated Laplacian matrixes:

$$110 \quad \Delta(t) = D(t) - A(t), \quad [1]$$

111 where $D(t)_{ii} = \left(\sum_j (A(t))_{ij} \right)^{-1}$ and $D(t)_{ij} = 0$ for $i \neq j$. Once done, the eigenvectors and
112 eigenvalues of $\Delta(t)$: $\{\phi_{k,t}\}$ and $\{\lambda_{k,t}\}$ were then computed. Then, Hermitian graph wavelets were
113 formed as:

$$114 \quad \psi_{s,x}(y) = \sum_k s \lambda_k(t) e^{-s\lambda_k(t)} \phi_{k,t}(x) \phi_{k,t}(y), \quad [2]$$

115 for a chosen set of scales $s \in \{s_n\}$. The last step was to compute the dominant scale at each vertex,
116 defined as

$$117 \quad S(x, t) = \arg \max_{s_n} \sum_y |\psi_{s_n,x}(y)|^2. \quad [3]$$

118 It is important to note that it is up to the investigator to define the set of scales $\{s_n\}$ to be
119 input into the above algorithm. This determination is largely ad hoc and some trial and error is
120 required. However some general points can be made: The scales should always be positive and the
121 largest should not be bigger than the largest eigenvalue. The goal is to get a good partitioning of

122 the interval $[0, \max\{\lambda(t)\}]$, however as the eigenvalues are in general not uniformly spaced some
123 experimentation may be required to find at what scales resolution is most informative.

124 The above SGW wavelet functions yield a complete multiscale analysis at each vertex in the
125 graph (i.e. for every cell tower in Senegal). To highlight the ability of Spectral Graph Wavelets to
126 identify multiscale patterns, we chose to build our wavelets with a fixed base point at the vertex
127 corresponding to the city of Touba (denoted x_T), and track the dominant scale $S(x_T, t)$ as above,
128 through time. We abbreviate this as $S(t)$. We thus obtain a sequence of dominant wavelet functions
129 on the network $\psi_{S(t),x_T}(t, y)$, where the first argument indicates the dependence on the changing
130 network structure encoded in $A(t)$ and the second argument y corresponds to vertices of the graph
131 on which the function takes its values.

132 Results

133 For any given day, human mobility to/from a given location can be extracted from the migration
134 matrix A and plotted as a map. The matrix row associated with Touba, a city in the center of
135 Senegal that is a destination for religious migration, describes the daily traffic going to and coming
136 from the city. One finds that large cities such as Dakar on the west coast (i.e. compare the location
137 of red nodes in Fig. 2A) account for most of this traffic. Once the daily dominant wavelet functions
138 were calculated, they too can be plotted geographically. In Fig. 2B we show the dominant wavelet
139 function for the same day as plotted in Fig. 2A. The function $\psi_{S(t),x_T}(t, y)$ is spatially very different.
140 One noticeable spatial pattern is the form of the wavelet function, which is centered on Touba: this
141 can be seen in Fig. 2B as a positive value on Touba, which then rapidly diminishes to negative values
142 in surrounding nodes, before becoming less negative (towards zero) at geographically far off nodes.
143 While the wavelet function was calculated from the to/from mobility data, and did not include
144 spatial distance explicitly, due to the spatial dependence of human mobility, which is reflected in
145 these data, spatial proximity to Touba has an imprint on the resulting wavelet function.

146 To characterize changes in human mobility through time, focusing on Touba, we clustered days

147 based upon origin-destination mobility. First, clustering was performed on the raw mobility data: for
148 every day, the vector of origin-destination mobility values corresponding to Touba were extracted from
149 the mobility matrices A , with each component of the vector corresponding to a node of the Senegal
150 cell tower graph. We performed the clustering using the Louvain Method for community detection
151 (REF), which involved calculating the Euclidean distance between days based on their mobility
152 values. We chose the Louvain method primarily because it provides an objective determination of
153 the number of clusters. This is in contrast to other common clustering approaches which require the
154 number of clusters to be chosen (e.g. k-means).

155 To further explore the raw-data we simply examined changes in the total traffic to/from Touba
156 over time (Fig. 3A; this is calculated by summing the daily mobility values to/from Touba for all
157 locations). \log_{10} total traffic over time reveals a variety of qualitative features: most striking is a
158 large peak in traffic corresponding to the commemoration of Touba's mosque's 50th anniversary,
159 which occurred May 30th, 2013 (day 150), as well as the end of Ramadan August 7th, 2013 (day
160 219) and Eid al-Adha on October 15 (day 287). The clustering of these raw mobility for Touba data
161 identified two dominant time periods, corresponding to seasonal changes in the Senegalese weather
162 and agriculture, as the rainy season begins around September. There is a third cluster corresponding
163 to a short intermediate period (see Fig. 3A, changes in the marker color: green, grey and blue
164 correspond to the dry, intermediate and wet seasons respectively). Interestingly, the clustering only
165 picked out these long-term changes in mobility, and not the short-punctuated events relating to the
166 religious festivals. This is because while these events are associated with a change in total traffic
167 to/from Touba, they do not reflect changes in the topology of the migration network through time.

168 In contrast to these results, clustering dominant wavelet functions over time revealed more nuanced
169 information. Clustering was done in the same way: calculating the Euclidean distance between days
170 based on the wavelet function centered on Touba, before using the Louvain community detection
171 algorithm to identify clusters. Like the analysis of the raw mobility values, this clustering of
172 daily wavelet functions identifies changes in human mobility relating to the dry and wet season.

173 Importantly however, now the presence of relatively short-term and large-scale events were identified
174 throughout the year. These events correspond to a short-term widening of the wavelet function
175 centered on Touba (in network space), and corresponds to the three religious migration events list
176 above. In addition, other events are identified: the Maouloud/Gamou celebration that occurred on
177 January 23, 2013 (day 23) and Independence day that occurred April 4th, 2013 (day 94). There
178 are two extra events that this clustering approach identified, around day 50 and at the end of the
179 year. The former is an unknown event, which speaks to the value of the spectral graph wavelet
180 analysis identifying new and interesting features, and the latter is likely associated with new-years
181 celebrations/holidays.

182 The modes of human mobility found using spectral graph wavelets can be mapped. Averaging
183 the daily wavelet functions associated with each cluster (i.e. producing an average wavelet function
184 centered on Touba for the dry and wet seasons, and for each short-term / large-scale event) reveals
185 stark differences. For example, the absolute difference between the average wavelet function relating
186 to human mobility during the dry and wet seasons (Fig. 4A) identifies change in several coastal
187 towns (along with changes in mobility in and around Touba). These changes reflect two processes.
188 First, in Senegal there is a large and mobile agricultural work-force. People from the coast use
189 Touba as stepping stone to rural locations in the interior of the country. Second, also associated
190 with the start of the rainy season are floods, and in 2013 the coast experienced significant flooding.
191 These differences in the average wavelet functions associated with the dry and wet clusters also
192 reflect the impact of the floods on human mobility.

193 In contrast to these seasonal changes, the absolute difference between the average wavelet function
194 associated with short-term / large-scale events, in this case Eid al-Adha on October 15 (day 287),
195 and the wet season (Fig. 4B) identifies change in mobility to/from many small rural towns in the far
196 east of Senegal. This is indicative of long-distance travel that people make as they go to/from Touba
197 for the religious event. This kind of geographically meaningful information can be used to explore
198 difference in the various modes of human mobility found by clustering daily wavelet functions. In

199 general, this approach more readily reveals heterogenous spatial/network features, relative to results
200 produced from the analysis of the raw mobility values, which simply describe changes in overall
201 traffic to/from Touba.

202 **Discussion**

203 To improve upon current abilities to characterize changes in human mobility over time we have
204 developed and applied a Manifold Learning method based on spectral graph wavelets to a CDR
205 dataset. These data describe the origin-destination mobility patterns for people living in Senegal,
206 daily for the year 2013. This approach is scale explicit, that is it can be used to identify patterns
207 of human mobility that contrast in overall spatial/network scale. This is key to disentangling
208 multiscale patterns with big and rich datasets like those produced by CDRs. Spectral graph wavelets
209 applied to these data allowed us to identify seasonal changes in Senegalese human mobility, as
210 well as punctuated large-scale events relating to religious migration and a national holiday. These
211 short-term but large-scale events were not identified by a standard approach applied to raw mobility
212 values. In doing so, our spectral graph wavelets approach provides a new method for the multi-scale
213 decomposition of human mobility data, and expands the utility of CDR data for anticipating and
214 preparing for changes in modes of human mobility ranging from day-to-day commutes to large-scale
215 migrations and displacements caused by natural disasters.

216 Identifying the dominant modes of human mobility, and the scales at which they occur, is an
217 important part of developing policies for a whole host of operations, ranging from traffic infrastructure
218 to disaster response planning. These kinds of spatial policies have previously been made using far
219 coarser data, both spatially and temporally (REF). Here, making use of relatively high resolution
220 CDR data, we were able to tease out the spatial signal of punctuated large-scale events, in contrast
221 to a standard approach applied to the raw-data. The rate at which new data – always at higher
222 temporal frequency and greater spatial resolution – is every growing, and the quantitative tools used
223 to extract useful information from monthly or even annual datasets are rapidly diminishing in utility.

224 To make best use of these new data (i.e. next-gen CDR datasets) new tools are required, and indeed,
225 new tools that can articulate the complex and adaptive nature of human mobility have extreme
226 utility. Like human mobility, other complex adaptive systems are multi-scale by nature, and in
227 general there is a growing need to extract information about the micro-scale agents that comprise
228 the system, from which meso- and macro-scale patterns emerge. Data-analytic tools designed with
229 the multi-scale nature of complex adaptive systems in mind, will help policy makers develop plans
230 that explicitly account for the emergence of patterns over a continuum of scales, like in this case the
231 various modes of human mobility in Senegal, and their associated network/geographic scale.

232 The advances to spectral graph wavelets that we have made allowed us to essentially transform
233 the raw data to a form that better highlights the differences of human mobility patterns. In that
234 spirit, this analysis can be thought of as a process of dimensional reduction or a denoising of the
235 raw data, in a manner that accounts explicitly for network scale. Analyzing the raw mobility data
236 did not provide the same kinds of scale-dependent information because it is noisy. Admittedly, to
237 analyze the raw data we used a very basic approach to classification. There are indeed many other
238 more sophisticated approaches that we could have employed, in order to contrast with the results
239 produced from analysis of the dominant wavelets functions. These approaches vary from traditional
240 dimensional reduction techniques that rely on linear correlations, such as Principal Components
241 Analysis (PCA), to machine learning approaches for feature identification (REF). Indeed, there is
242 great potential for using the wavelet transformed data with these approaches to feature classification.
243 No information is lost in the process and combining our scale-explicit denoising of raw data with
244 machine learning approaches to classification is likely a fruitful vein of research.

245 In contrast to more commonly used approaches to data dimensional reduction, the spectral graph
246 wavelet analysis is nonlinear, multiscale, and localized. The properties of nonlinearity and resolving
247 multiple scales are shared by a number of manifold learning techniques, which can be thought
248 of as nonlinear generalizations of classical linear techniques such as PCA. What wavelet analysis
249 provides that such methods do not is the localization of information and the singling out of the

dominant scales of variation at each vertex. Whether this provides new and more useful information is an important question to ask: many classical approaches (like Principal Components Analysis) often decompose a given dataset using eigenfunctions, which are global in character. For example, PCA-based approaches to the analysis of meteorological and oceanographic datasets will reveal modes of variability that systematically decrease in spatio-temporal scale, and this reflects the eigenspectra produced in the analysis. These methods have been foundational to our understanding of important modes of climate variability such as El Nino, or the North Atlantic Oscillation. However, in practice these kinds of analyses are used to examine only the large-scale features embedded in the data, and the smaller-scale patterns are often assumed to be noise. Moreover, being global in nature, they cannot resolve spatial variability in multiscale structure. In contrast, spectral graph wavelet analysis provides a scale-explicit decomposition of raw data, and does so locally, allowing the user to resolve the differing multiscale structure across spatial locations.

The information that spectral graph wavelets provides can be used in many other ways. Here, we have analyzed human mobility data, but the CDR dataset can also be used to construct human communication networks through time. Performing the same analysis on both sets of data would produce concurrent wavelet functions through time. A comparison of changes in the dominant modes / scales of human communication and mobility might reveal early-warning signals of migration/displacement. Simply put, as people prepare to move they are likely to call their ultimate destination, and this information can help policy makers prepare for changes in population density at specific nodes/places. There is an opportunity to utilize methods from manifold matching (REF) to make these comparisons. Manifold matching has been used in image recognition to match photos of the same person, for example. Here, instead of a set of photos from a person's face, the manifolds that would be compared are those associated with a complex system (the Senegal cellular network) described in two ways (i.e. communications and mobility).

Mathematical theory suggests that in complex adaptive systems, such as dynamic human mobility networks, there may exist early-warning signals of multi-scale change. Many of these early-warning

signals are based on bifurcation theory (REF) and are measured by changes in the variance and autocorrelation in macroscopic variables. In the context of human mobility CDR data, there is an opportunity to advance new early-warning signals of multi-scale change. Specifically, the approach of spectral graph wavelets is one way to learn the geometry of the manifold on which dynamics occur. Changes in a system's geometry are known to be associated with the classical bifurcation-based early-warning signals (Colon REF), but go further to directly address how these kinds of large and abrupt changes in complex systems are related to changes in the behavior of micro-scale agents (i.e. in this case, how individual people move from place to place).

Identifying the dominant modes of human mobility, and potentially early-warning signals of changes between them, is of principal interest of groups tasked with managing human communities, as they go about their everyday lives as well as respond to infrequent but impactful events like a natural hazard. In Senegal, flooding is a persistent problem and indeed in 2013 the capital Dakar was severely hit. These kinds of events can lead to the permanent displacement of people from their homes, and there is value to identify where and when this displacement occurs. Indeed, displacement is not necessarily instantaneous with regards to the perturbation, but it may take a relatively long time for people to “realize” their displacement. Multi-scale methods like spectral graph wavelets applied to CDR data can help distinguish this additional mode of human mobility, and further, methods from manifold matching are likely to be useful. These approaches are used in image classification to associate similar features, for example faces from photograph albums. In this context manifold matching would be used to associate similar (or dissimilar) parts of human mobility graphs, as they undergo change created by a perturbation like a flood.

In sum, we have advanced the use of spectral graph wavelets to analyze CDR human mobility data. Our approach extracts useful information that is scale-explicit: we identified seasonal changes in human mobility as well as punctuated large-scale mobility events associated with religious celebrations and a national holiday. Spectral graph wavelets essentially recast the original mobility data in terms of the dominant modes of human mobility, as a function of scale. Here, we have focused

302 our multi-scale analysis on one place in Senegal - Touba - a place of religious significance. However,
303 the spectral graph wavelets analysis produces information for all nodes in the network, and there is
304 rich scale-explicit information in the full wavelet transform of the raw data. While the growth in
305 data obtained for complex adaptive systems is daunting, the application of scale-explicit dimensional
306 reduction techniques like we have explored here, greatly improves our ability to characterize and
307 predict multi-scale change. An ability to do so is vital if we are to maintain welfare from the complex
308 system in which we are embedded.

309 **Materials and Methods**

310

311 **Mathematical Motivation.** The set of mobility network data described above was analyzed using a new form
312 of Spectral Graph Wavelets, specifically Hermitian Graph Wavelets. Below we provide describe SGWs
313 and our advance to them, and then describe how they were applied to the set of daily Senegal mobility
314 matrices. The general framework of Spectral Graph Wavelets depends on a choice of wavelet kernel (i.e.
315 the shape of the wavelet) that materially affects the properties and outputs of the wavelet-based analysis.
316 Other studies have relied on wavelet kernels constructed in a somewhat ad hoc manner broadly aimed
317 at similarity to kernels used in classical analyses. Critically, in the setting of graphs, it is not clear how
318 much such general wavelet kernels resolve the geometric structure of a graph, and in fact the theorems
319 establishing the desirable properties of wavelets in the classical case can fail for such general wavelet kernels
320 on graphs. One of our key contributions is the establishment and proof of such analogous theorems for
321 graphs, guaranteeing our heat kernel-based wavelets do in fact properly quantify and reflect the multiscale
322 geometry of the underlying graph, in this case corresponding to patterns of human mobility.

323 **Mathematical Advances to Spectral Graph Wavelets.** Wavelets on graphs are a generalization of classical
324 wavelet analysis, used commonly in time-series analysis (REF), to graphs/networks. Like in classical
325 applications of wavelets to timeseries analysis, spectral graph wavelets (REF) provide a multiscale analysis
326 on networks in terms of spectral resolutions of graph Laplacians. As with classical wavelets, there is a

choice of which wavelet kernel to use. Here, we construct Hermitian graph wavelets derived from the heat kernel, which have well known properties and applications to revealing network structure [(REF)]. Hermitian graph wavelets, as we shall show here, possess similar localization properties to classical wavelets and can thus be used to analyze the structure of networks locally and at various scales. However, while the heat kernel provides a local, multiscale representation of network structure, it does not provide for any choice of scale. What are the important scales for a given network? Wavelets possess cancelation properties that allow for multiscale decompositions in terms of basis functions that are able to distinguish among scales and highlight which scales are most important in terms of network structure at a given vertex. Moreover, our wavelets being derived from the heat kernel allows for direct interpretation and understanding of these dominant scales in terms of well understood geometric properties of the heat kernel. Thus, the tools we present here allow for localized, multiscale decomposition of networks into dominant scales, identifying the most important scales for each vertex. In this way, our methods provide a localized multiscale learning of network geometry, similar to other well-known (but global) multiscale learning algorithms, for example Principal Components Analysis (REF), and Diffusion Maps (REF).

Applying Spectral Graph Wavelets to Human Mobility Data. Each daily mobility matrix is an *undirected*, *weighted* graph. The weights on the edges $w_{i,j} : E \rightarrow [0, \infty)$ are non-negative real numbers, and where $w_{i,j} = 0$ if $e_{i,j} = 0$. The set of edges and their weights define a given *affinity matrix* A where $A_{i,j} = w_{i,j}$, for a given day of the year in 2013. Given a network represented by a graph, often one also has additional data in the form of a function $f : V \rightarrow \mathbb{R}$, which assigns to each vertex a value. In our setting for example, f could be the user density at each tower on a given day. Such a function is often referred to as a *signal* on the network in analogy with classical signal processing. We can consider the set of all signals on G as a vector space $L^2(V)$ with inner product given by

$$\langle f, g \rangle = \sum_{x \in V} f(x)g(x),$$

and where as usual we define $\|f\|^2 = \langle f, f \rangle$. The graph Laplacian is defined as an operator on this space of signals as follows: Let D be the row (or column) sum of A , meaning that $D_{i,j} = 0$ for $i \neq j$ and

352 $D_{i,i} = \sum_j A_{i,j}$. The Laplacian is given by

353
$$\Delta = D - A.$$

354 Δ is symmetric, and if we let $\{\phi_k\}$ and $\{\lambda_k\}$ denote the eigenvectors and eigenvalues respectively of Δ , we
355 have as usual that $\langle \phi_i, \phi_j \rangle = 1$ if $i = j$ and 0 otherwise. Also, $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$. We can then define
356 the *graph Fourier transform* of a signal $f \in L^2(V)$ by $\hat{f}(k) = \langle f, \phi_k \rangle$ and have the usual inversion formula
357 $f = \sum_k \hat{f}(k)\phi_k$. (Note that we have suppressed any mention of time or dynamics in the above, however the
358 weights $w_{i,j}$ and all other quantities above may depend on time as well and evolve dynamically. For now,
359 we continue to suppress this time dependence, dealing with all values as fixed, and will return to dynamics
360 later.)

361 Following the general framework of Spectral Graph Wavelets (REF) we construct our wavelets as kernels
362 of operators given by $g(s\Delta)$ where $s \in (0, \infty)$ is a continuous scale parameter and $g : [0, \infty) \rightarrow \mathbb{R}$ satisfies
363 certain conditions. For a fixed scale s and $x \in V$ our wavelet kernel takes the form

364
$$\psi_{s,x}(y) = \sum_{k=0}^N g(s\lambda_k)\phi_k(x)\phi_k(y).$$

365 **Using the Heat Kernel.** One of the features of primary importance in classical wavelet analysis is the ability
366 of the wavelet kernel to localize in frequency as well as in space. In the classical setting of the real axis this
367 is readily achieved using the Fourier transform and its transformation under rescaling. In the setting of a
368 graph where no such obvious rescaling exists. For this we turn to the heat kernel, defined as the integral
369 kernel of the heat semigroup, $e^{-t\Delta}$ (REF)

370
$$H_t(x, y) = \sum_{k=0}^N e^{-t\lambda_k}\phi_k(x)\phi_k(y).$$

371 The term "heat" refers to the fact that the heat semigroup when applied to any function $f \in L^2(G)$ solves
372 the heat equation:

373
$$\frac{d}{dt}e^{-t\Delta}(f) = \Delta e^{-t\Delta}(f).$$

374 When f is taken to be a point mass δ_x , one obtains the heat kernel.

375 One of the most important features of the heat kernel is that its decay reflects the geometry of G , as is

376 made precise by the theory of ([cite yau]), as follows: First, suppose we are given a metric on G , that is a
 377 map $\rho : V \times V \rightarrow [0, \infty)$ that satisfies the triangle inequality. Such a metric is called *intrinsic* if for each
 378 $x \in V$ we have $\sum_{y \sim x} w_{x,y} \rho^2(x, y) < 1$. Given an intrinsic metric, we define its *jump size* to be $\sup_{x \sim y} \{\rho(x, y)\}$,
 379 meaning the supremum over all pairs of points with edges with nonzero weight. With these definitions in
 380 hand we now state the main result of ([cite yau]): If we let

$$381 \quad \zeta_s(t, r) = \frac{1}{s^2} \left(rs \cdot \operatorname{arcsinh} \frac{rs}{t} - \sqrt{t^2 + r^2 s^2} + t \right),$$

382 then we have

$$383 \quad H_t(x, y) \leq e^{-\zeta_s(t, \rho(x, y))}$$

384 where s is the jump size of the intrinsic metric ρ . Note the bounds dependence on the choice of intrinsic
 385 metric. [Plot compared to Gaussian]

386 **Wavelet Localization.** Returning to our construction of our wavelets, by analogy with classical harmonic
 387 analysis we choose for our g the time derivative of the heat kernel $H_t(x, y)$: $g(\lambda) = \lambda e^{-\lambda}$, so that in what
 388 follows

$$389 \quad \psi_{s,x}(y) = \sum_{k=0}^N s \lambda_k e^{-s \lambda_k} \phi_k(x) \phi_k(y).$$

390 Using the above bounds we can obtain analogous bounds for our wavelet kernel by following the method of
 391 proof in [cite Davies], in particular one can obtain the following bounds (with the same notation as above
 392 and letting $r = \rho(x, y)$):

$$393 \quad |\psi_{t,x}(y)| \leq \left[\frac{r^2}{t} \left(1 + \frac{s}{\sqrt{t^2 + s^2 r^2}} \right) \left(\frac{1}{sr + \sqrt{t^2 + s^2 r^2}} \right) - \left(\frac{t}{\sqrt{t^2 + s^2 r^2}} + 1 \right) \right] e^{-\xi_s(t, r)}$$

394 This bound, which to our knowledge has not appeared in the literature before, gives precise quantitative
 395 information on the decay of $|\psi_{t,x}(y)|$ as $\rho(x, y) \rightarrow \infty$ and ensures sharp localization of wavelets. Note
 396 however the proof depends essentially upon our choice of g and its relation to $H_t(x, y)$, which was a primary
 397 motivation for our choice.

398 Another benefit of our choice of wavelet kernel is the following interpretation of scale it provides: As

399 mentioned above the heat kernel provides multiscale structure locally but does not distinguish between
400 which scale is most important. With our choice of wavelet, for each vertex x we define the dominant scale
401 as the value of t that maximizes $\|\psi_{t,x}\|^2$. This is possible as one can show, e.g. using the above bounds or
402 from reasoning directly from the heat equation, that $\lim_{t \rightarrow +0} |\psi_{0,x}(y)| = 0$ for all x, y and $\lim_{t \rightarrow \infty} |\psi_{t,x}(y)| = 0$,
403 which together with the fact that ψ is continuous on $[0, \infty) \times G \times G$ tells us it must achieve a maximum.
404 That this maximum is unique follows from the maximum principle of the heat equation. We can interpret
405 this scale in the following way: $\|\psi_{t,x}\|^2$ measures the total energy of the derivative of $H_t(x, \cdot)$ over the
406 network. When this is maximum is when heat is maximally flowing from vertex x to the rest of the graph,
407 i.e., is the scale at which it is most strongly influencing the rest of the network.

408 **ACKNOWLEDGMENTS.** The authors would like to acknowledge the DARPA Young Faculty Award
409 YFA N66001-17-1-4038, and the Orange D4D competition for providing the data.

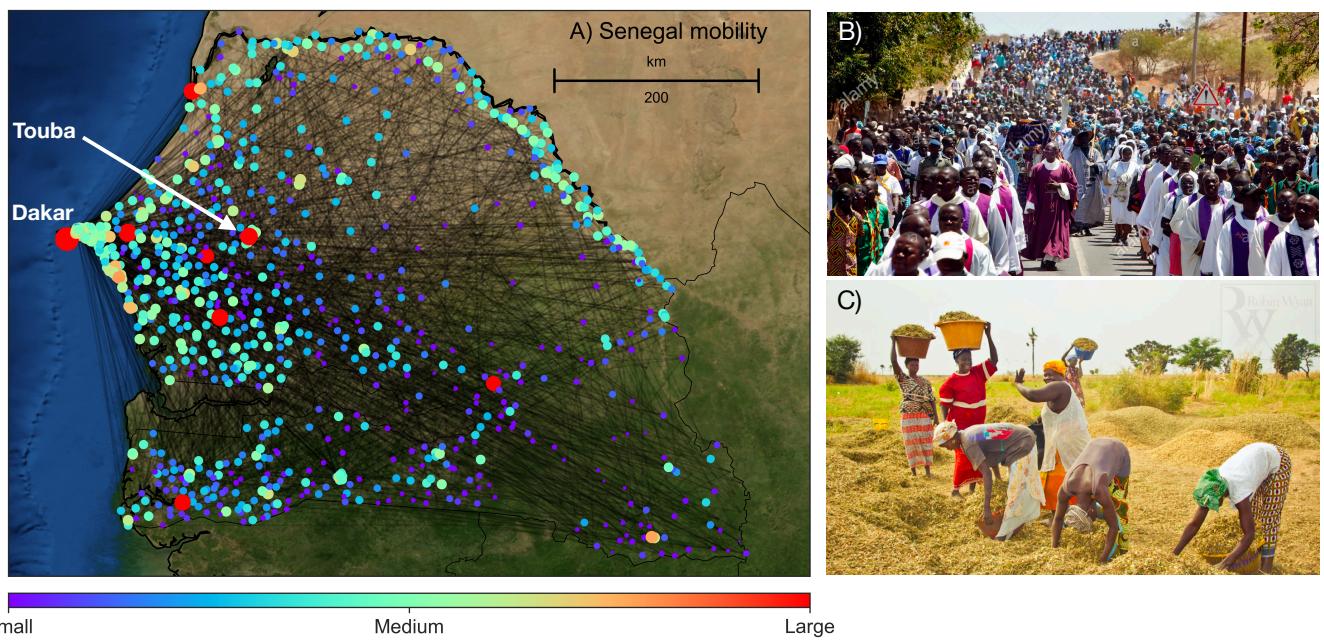


Fig. 1. A) Map of Senegal with major cellular communication nodes color coded by relative population density (i.e. town/city size) and edges connecting them highlighting the density and complexity of human communication and mobility networks. These networks are changing over time, and in Senegal there are large-scale modes of mobility relating to religious events (often held at Touba) and relating to seasonal changes and agriculture (B and C).

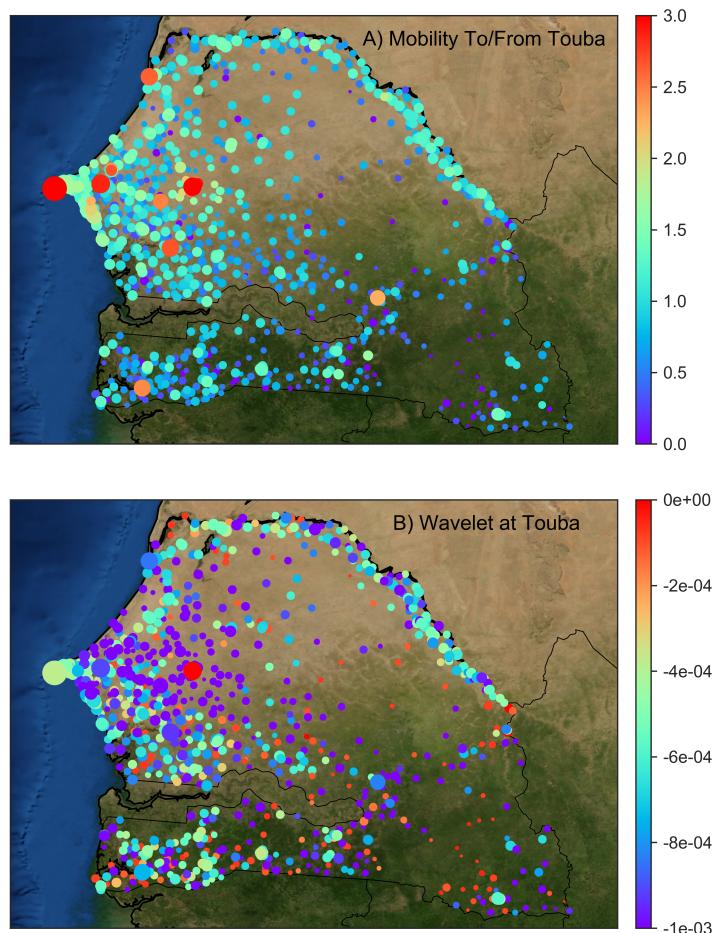


Fig. 2. A) Mobility nodes in Senegal color coded by the \log_{10} number of people moving to/from Touba on a random day in 2013. B) The dominant wavelet function centered on Touba for the same day. The two maps reveal very different information. In (A) mobility to/from the major urban hubs are identified and in (B) the shape of the wavelet function is highlighted (positive at Touba, decreasing to large negative values in nearby towns, before becoming less negative at far-off towns).

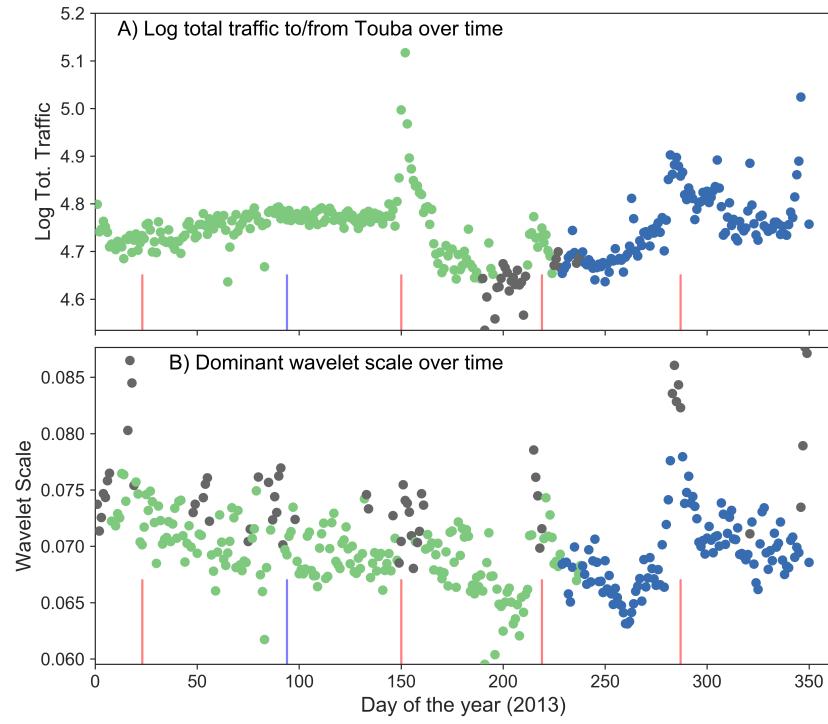


Fig. 3. A) Time-series of the \log_{10} total traffic to/from Touba for 2013, colored by group ID produced from clustering days by their raw to/from Touba data. The cluster analysis identifies mobility associated with the dry (green) and wet (blue) seasons (with an transitional phase in grey), and there are peaks in total traffic to/from Touba that correspond with major religious events at day 150, 219 and 287; however, the cluster analysis does not recognize these events as distinct because in these data only the magnitude of traffic changes, with little information about changes in where people move from/to. B) In contrast, changes over time in the dominant wavelet scale better reveals the punctuated and large-scale mobility events related to religious celebrations (vertical red lines) and Senegal's independence day (blue vertical line), in addition to distinguishing the dry and wet seasons.

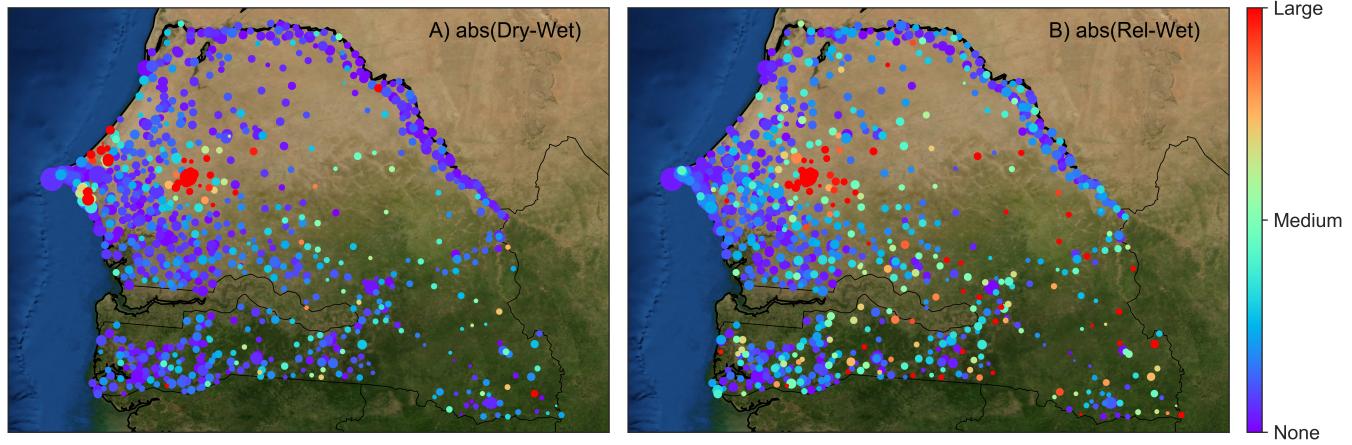


Fig. 4. Dominant modes of human mobility can be identified by averaging the wavelet functions associated with the cluster groups shown in Fig. 3B. Then, these modes of human mobility can be compared by calculating their absolute difference and plotting as a map. A) The absolute difference in the average wavelet function (centered on Touba) associated with the dry and wet seasons highlights changes in mobility to/from the coast and Touba. B) The absolute difference in the average wavelet function (centered on Touba) associated with religious events and the wet season reveals changes in mobility in far-off towns in the east of Senegal, and Touba.