Wavelet-based visualization of time-varying data on graphs

Paola Llerena Valdivia
Advisor: Prof. Dr. Luis Gustavo Nonato

Instituto de Ciências Matemáticas e de Computação
USP São Carlos

ICMC SÃO CARLOS USP

Moens et al., *Mining user generated content*, 2014.

Visual Analytics

Machine

Statistical analysis

Data mining

Machine learning

Compression & filtering

Keim et al., *Visual analytics: Definition, process, and challenges*, 2008.

Visual Analytics

Machine

Statistical analysis

Data mining

Machine learning

Compression & filtering

Human

Human cognition

Perception

Visual intelligence

Decision making theory

Keim et al., *Visual analytics: Definition, process, and challenges*, 2008.

Visual Analytics

Machine

Statistical analysis

Data mining

Machine learning

Compression & filtering

Human-computer interaction

"The best of both sides"

Information visualization

Graphics and rendering

Human

Human cognition

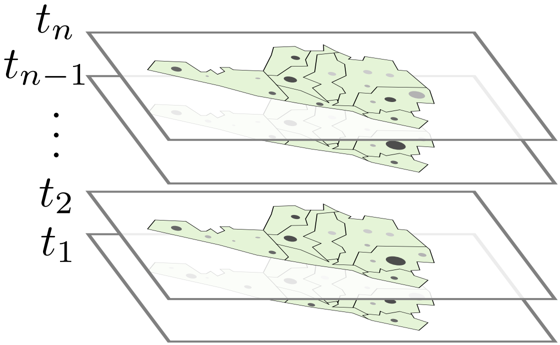
Perception

Visual intelligence

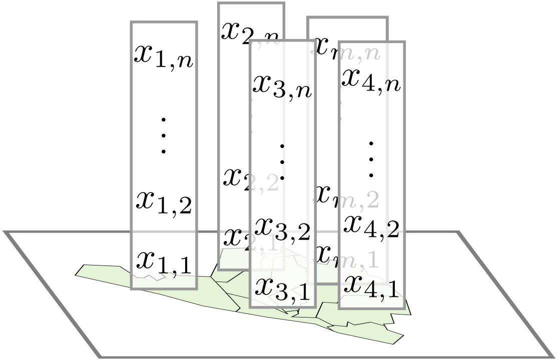
Decision making theory

Keim et al., *Visual analytics: Definition, process, and challenges*, 2008.

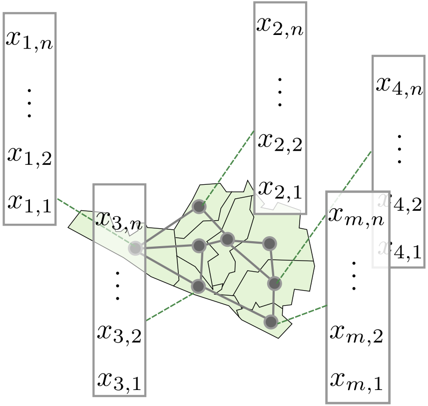
Spatiotemporal data



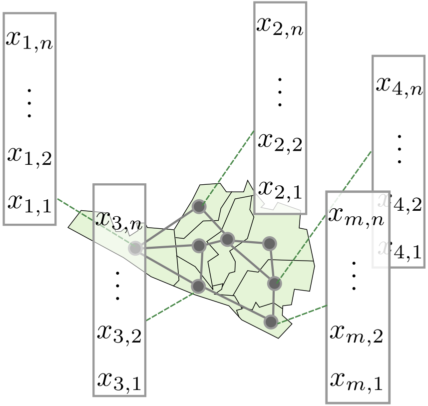
Spatial time series



Spatial time series



Spatial time series



- Take advantage of many graph techniques.
- Particularly, Wavelet theory on graphs, that allows the analysis of local variations .

Aim

Investigate novel visual analytic methodologies for the analysis of spatial time series that relies on the recent developments of graph Wavelet theory.

Aim

Investigate novel visual analytic methodologies for the analysis of spatial time series that relies on the recent developments of graph Wavelet theory.

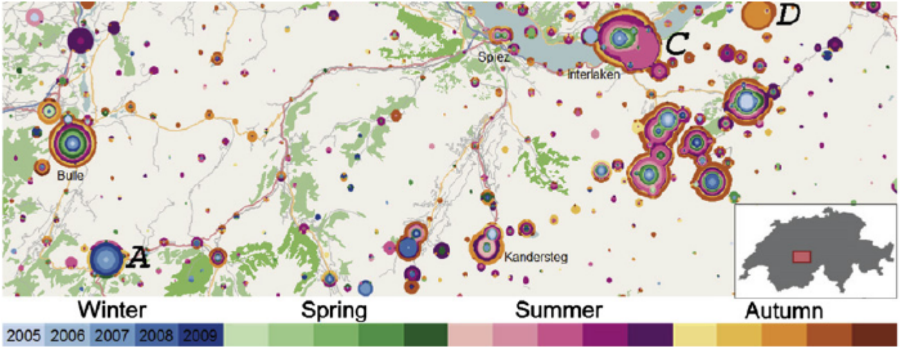
The intended methodology will combine graph wavelets, computational tools and visual elements in order to allow for the analysis of local and global data variation.

Expected contributions

- An efficient and scalable methodology that will guide users in the exploration of spatial time series towards the discovery of hidden patterns.
- An interactive visual framework tailored for the proposed methodology.
- Application in real scenarios such as urban data.

Visualization of spatial time-series

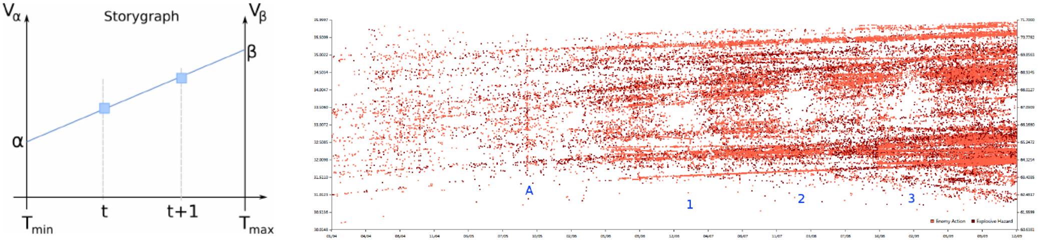
Integrated views



Growth ring maps.

Bak et al., *"Spatiotemporal analysis of sensor logs using growth ring maps"*, 2009.
Andrienko et al., *"A conceptual framework and taxonomy of techniques for analyzing movement"*, 2011.

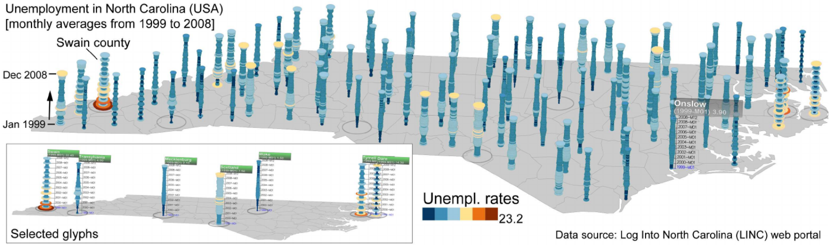
Integrated views



Storygraphs.

Shrestha et al., *“Storygraph: Extracting Patterns from Spatio-temporal Data”*, 2013.

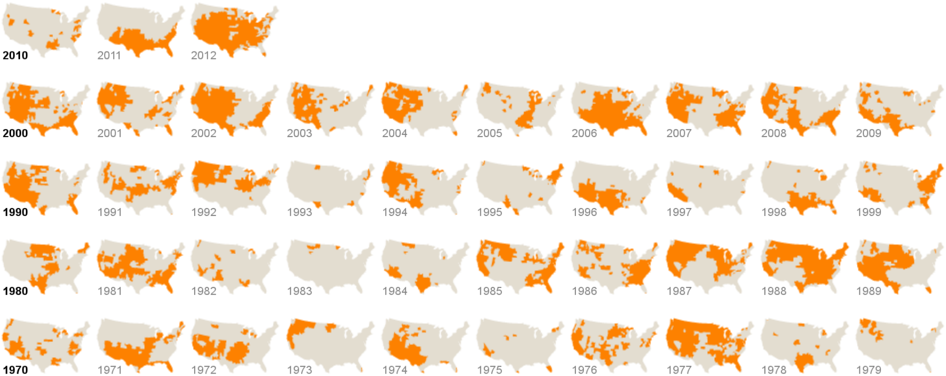
Integrated views in 3D



Data vases

Thakur and Rhyne, *"Data Vases: 2D and 3D Plots for Visualizing Multiple Time Series"*, 2009.

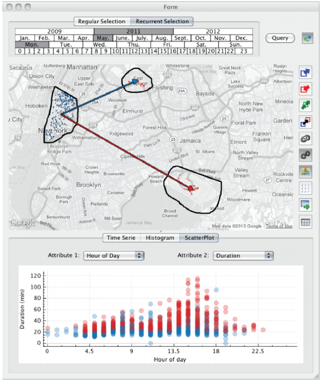
Multiple views



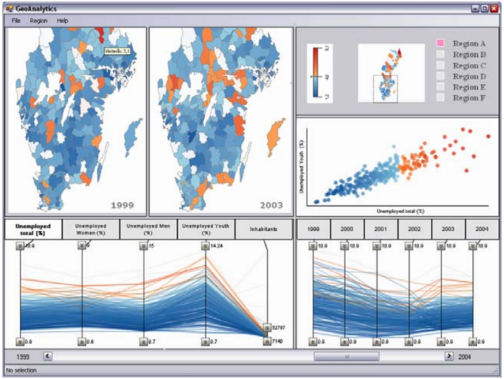
Small multiples.

Park and Quealy, *Drought's Footprint*, 2012.
New York Times, http://www.nytimes.com/interactive/2012/07/20/us/drought-footprint.htm. Accessed May, 2015.

Multiple views



A)



B)

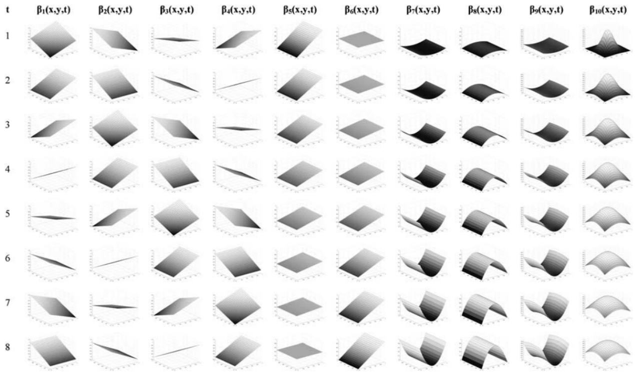
Coordinated multiple views.

A) Ferreira et al., *“Visualization and Computer Graphics, IEEE Transact”*, 2013.

B) Jern and Franzén, *““GeoAnalytics” - Exploring spatio-temporal and multivariate data”*, 2006.

Visual analytics of time series

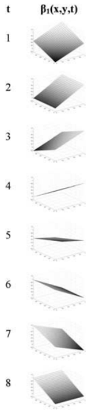
Geographically weighted regression

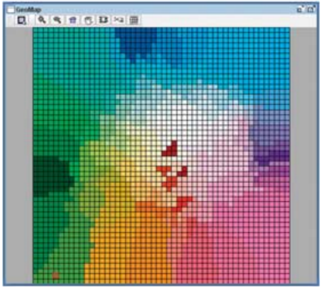


Regression parameters.

Demšar et al., *"Exploring the spatio-temporal dynamics of geographical processes with geographically weighted regression and geovisual analytics"*, 2008.

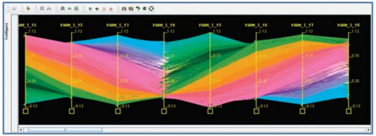
Geographically weighted regression





geoMap

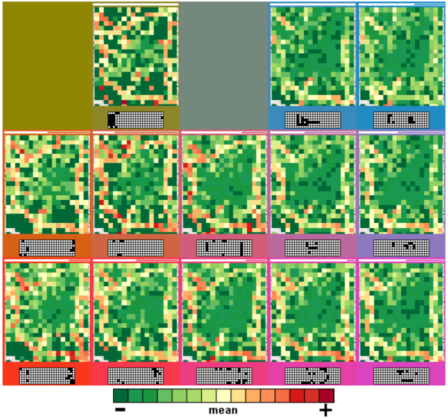
SOM



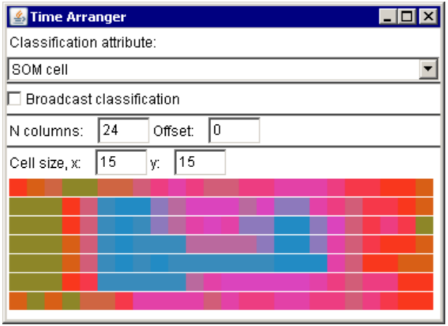
Temporal PCP

Demšar et al., *"Exploring the spatio-temporal dynamics of geographical processes with geographically weighted regression and geovisual analytics"*, 2008.

Space-in-time SOMs



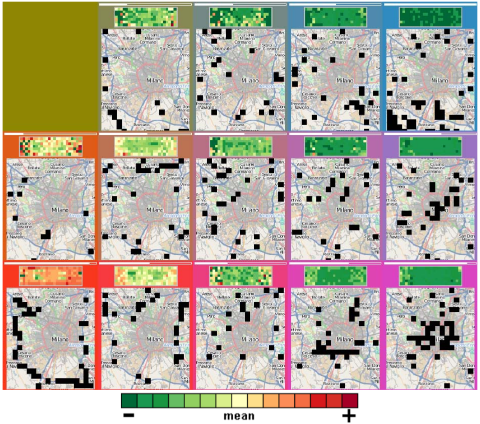
SOM matrix.



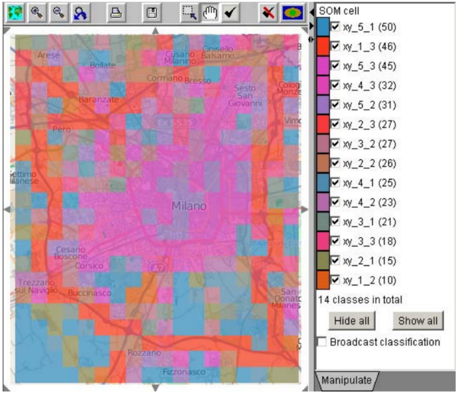
Temporal patterns

Andrienko et al., *"Space-in-Time and Time-in-Space Self-Organizing Maps for Exploring Spatiotemporal Patterns"*, 2010.

Time-in-space SOMs



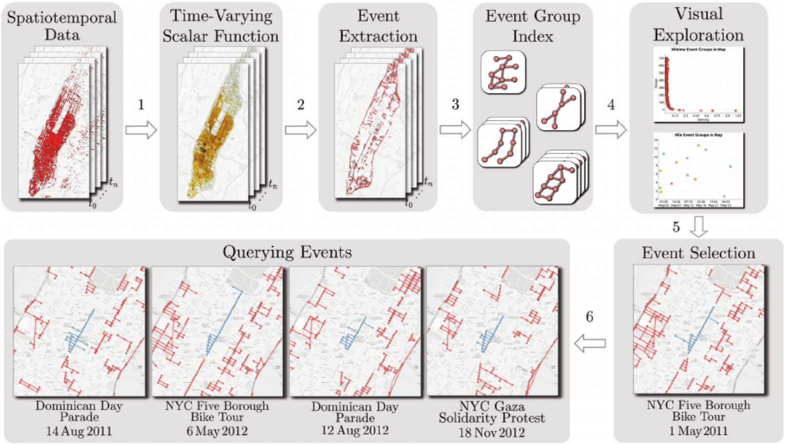
SOM matrix.



Spatial patterns

Andrienko et al., *"Space-in-Time and Time-in-Space Self-Organizing Maps for Exploring Spatiotemporal Patterns"*, 2010.

Event-guided exploration



Doraiswamy et al., *"Using Topological Analysis to Support Event-Guided Exploration in Urban Data"*, 2014.

Graph Basics

Notation

Weighted graph

$$\mathcal{G}^w = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$$

Notation

Weighted graph

$$\mathcal{G}^w = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$$

$\mathcal{V} = \{\tau_1, \tau_2, \ldots, \tau_N\}$	node set
$\mathcal{E}$	link set
$\mathbf{W}$	weight matrix
$w_{ij} \neq 0$ if $\tau_i \sim \tau_j$	weigth matrix entry
$d^w(\tau_i) = \sum_{\tau_j \sim \tau_i} w_{ij}$	node degree

Graph function

$f : \mathcal{V} \to \mathbb{R}$

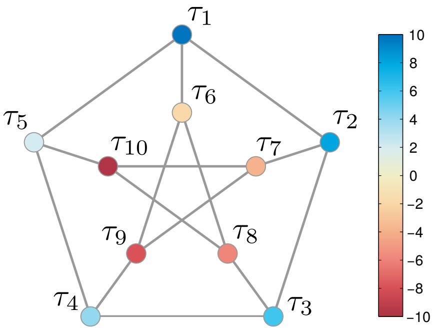
$\mathbf{f} \in \mathbb{R}^N$

Graph function

$$f : \mathcal{V} \to \mathbb{R}$$

$$\mathbf{f} \in \mathbb{R}^N$$

$f(i)$ is the scalar value in the node τ_i .



$$\mathbf{f} = \begin{bmatrix} 10 \\ 8 \\ 6 \\ 4 \\ 2 \\ -2 \\ -4 \\ -6 \\ -8 \\ -10 \end{bmatrix}$$

Function defined on the Petersen graph.

Unnormalized Laplacian matrix

$$\mathbf{L} := \mathbf{D} - \mathbf{W}$$

$\mathbf{D}$	diagonal degree matrix
$\mathbf{W}$	weight matrix
$D_{ii} = d^w(\tau_i)$	diagonal degree matrix entries

- $\mathbf{L}$ is a real symmetric, positive-semidefinite matrix;
- real non-negative eigenvalues $\sigma(\mathbf{L}) = \{\lambda_\ell\}_{\ell=0,1,\ldots,N-1}$, which can be ordered $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{N-1} := \lambda_{\max}$
- corresponding orthogonal eigenvectors $\{\mathbf{u}_\ell\}_{\ell=0,1,\ldots,N-1}$.

Graph Fourier transform

Fourier transform

$f : \mathbb{R} \to \mathbb{R}$

$$\begin{aligned}\widehat{f}(\xi) &= \langle f, e^{2\pi i \xi t} \rangle \\ &= \int_{\mathbb{R}} f(t) e^{-2\pi i \xi t} dt\end{aligned}$$

$f : \mathcal{V} \to \mathbb{R}$

$$\begin{aligned}\widehat{f}(\lambda_\ell) &= \langle \mathbf{f}, \mathbf{u}_\ell \rangle \\ &= \sum_{\tau_i \in \mathcal{V}} f(i) \mathbf{u}_\ell^\intercal(i)\end{aligned}$$

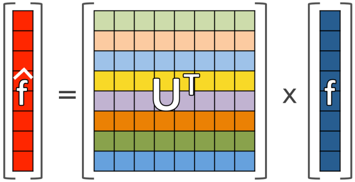
Fourier transform

$$f : \mathbb{R} \to \mathbb{R}$$

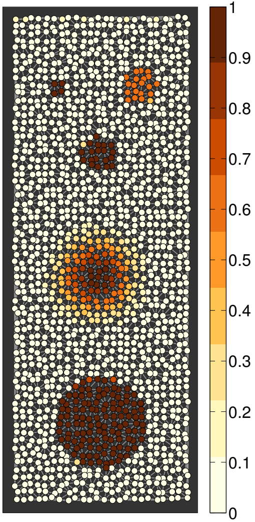
$$\begin{aligned}\widehat{f}(\xi) &= \langle f, e^{2\pi i \xi t} \rangle \\ &= \int_{\mathbb{R}} f(t) e^{-2\pi i \xi t} dt\end{aligned}$$

$$f : \mathcal{V} \to \mathbb{R}$$

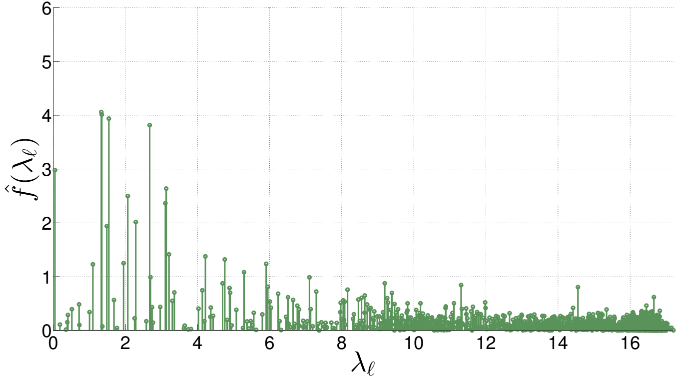
$$\begin{aligned}\widehat{f}(\lambda_\ell) &= \langle \mathbf{f}, \mathbf{u}_\ell \rangle \\ &= \sum_{\tau_i \in \mathcal{V}} f(i) \mathbf{u}_\ell^{\mathsf{T}}(i)\end{aligned}$$



Example



$f : \mathcal{V} \to \mathbb{R}$



$\hat{f} : \sigma(\mathbf{L}) \to \mathbb{R}$

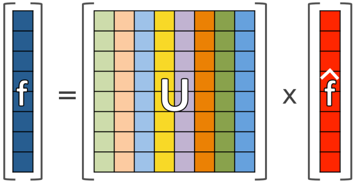
Inverse Fourier transform

$f : \mathbb{R} \to \mathbb{R}$

$$f(t) = \int_{\mathbb{R}} \widehat{f}(\xi) e^{2\pi i \xi t} d\xi$$

$f : \mathcal{V} \to \mathbb{R}$

$$f(i) = \sum_{\ell=0}^{N-1} \widehat{f}(\lambda_\ell) \mathbf{u}_\ell(i)$$



Frequency domain filtering on graphs

$$\widehat{f}_{out}(\lambda_l) = \widehat{f}_{in}(\lambda_l)\widehat{g}(\lambda_l)$$
$$f_{out}(i) = \sum_{l=0}^{N-1} \widehat{f}_{out}(\lambda_l)\mathbf{u}_l(i)$$

Frequency domain filtering on graphs

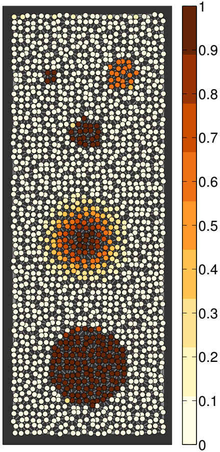
$$\widehat{f}_{out}(\lambda_l) = \widehat{f}_{in}(\lambda_l)\widehat{g}(\lambda_l)$$

$$f_{out}(i) = \sum_{l=0}^{N-1} \widehat{f}_{out}(\lambda_l)\mathbf{u}_l(i)$$

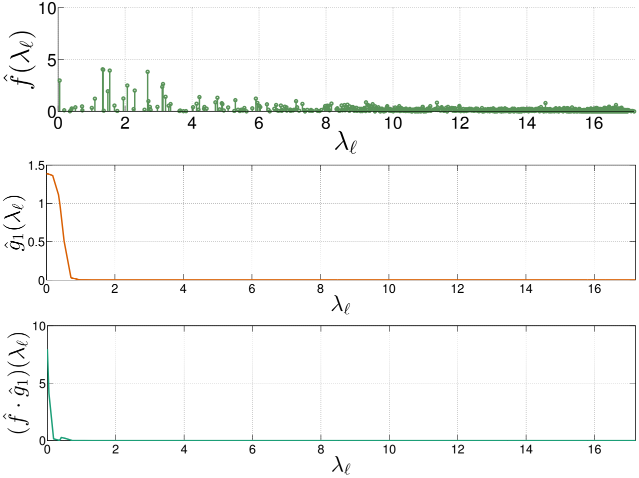
$$\mathbf{f}_{out} = \mathbf{U} \underbrace{\begin{bmatrix} \widehat{g}(\lambda_0) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \widehat{g}(\lambda_{N-1}) \end{bmatrix}}_{\mathbf{G}} \mathbf{U}^{\intercal}\mathbf{f}_{in}$$

$$\mathbf{f}_{out} = \mathbf{U}\mathbf{G}\mathbf{U}^{\intercal}\mathbf{f}_{in}$$

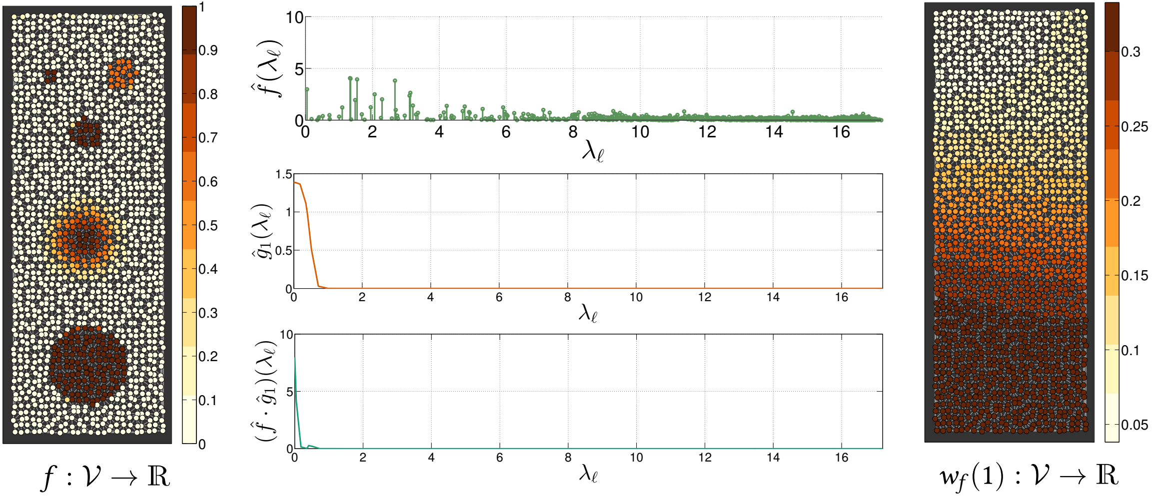
Example



$f : \mathcal{V} \to \mathbb{R}$

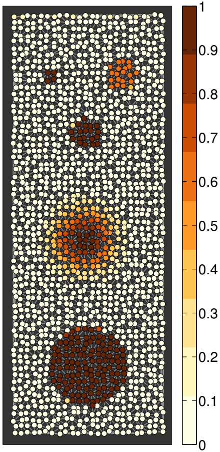


Example

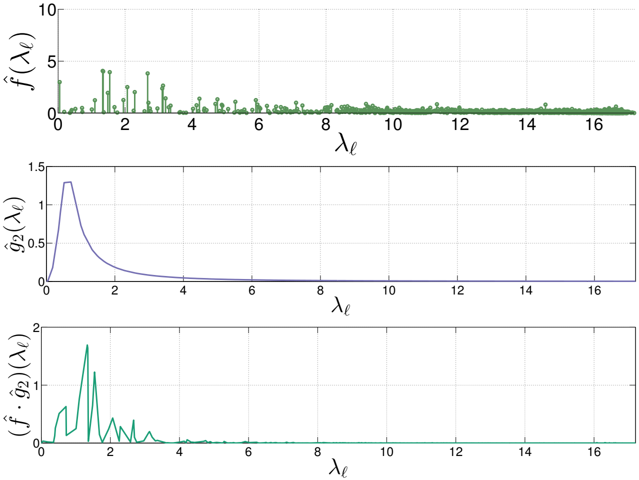


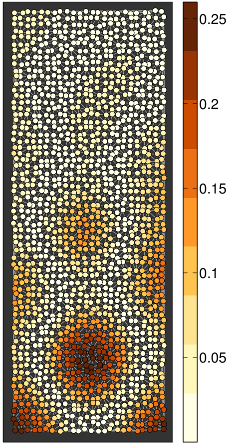
$f : \mathcal{V} \to \mathbb{R}$
$\hat{f}(\lambda_\ell)$
$\hat{g}_1(\lambda_\ell)$
$(\hat{f} \cdot \hat{g}_1)(\lambda_\ell)$
λ_ℓ
$w_f(1) : \mathcal{V} \to \mathbb{R}$

Example



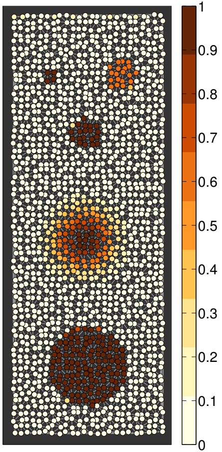
$f : \mathcal{V} \to \mathbb{R}$



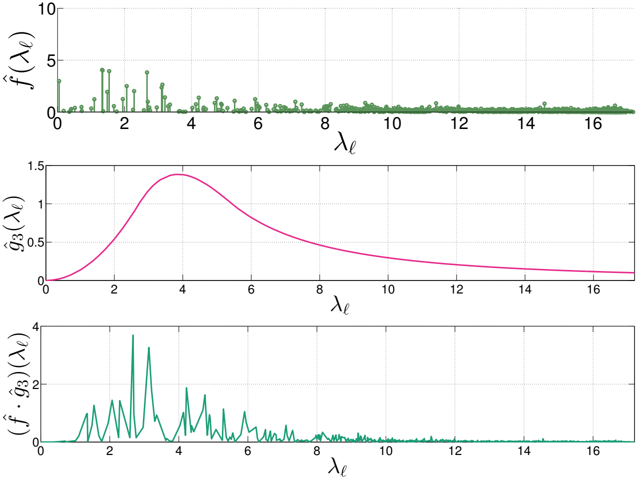


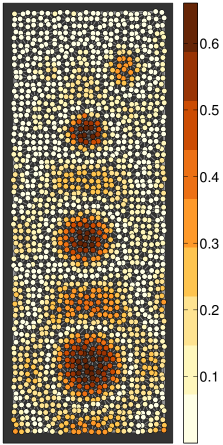
$w_f(2) : \mathcal{V} \to \mathbb{R}$

Example



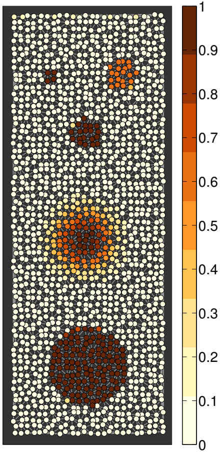
$f : \mathcal{V} \to \mathbb{R}$



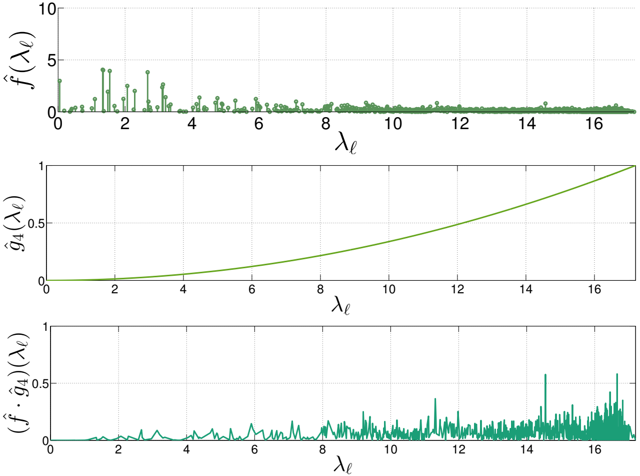


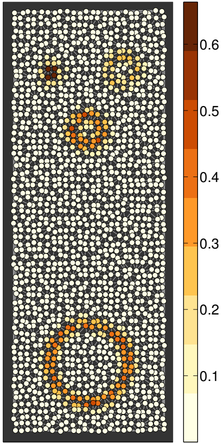
$w_f(3) : \mathcal{V} \to \mathbb{R}$

Example



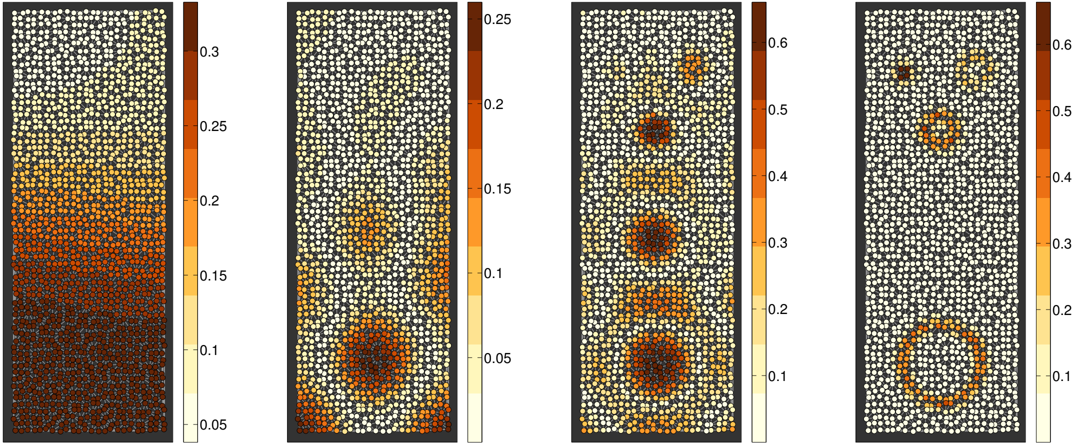
$f : \mathcal{V} \to \mathbb{R}$





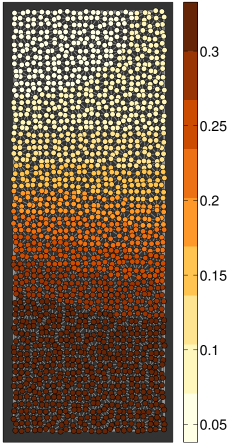
$w_f(4) : \mathcal{V} \to \mathbb{R}$

Example

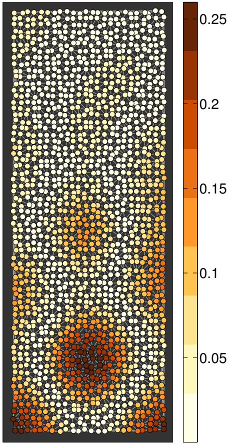


$w_f(1) : \mathcal{V} \rightarrow \mathbb{R}$ $w_f(2) : \mathcal{V} \rightarrow \mathbb{R}$ $w_f(3) : \mathcal{V} \rightarrow \mathbb{R}$ $w_f(4) : \mathcal{V} \rightarrow \mathbb{R}$

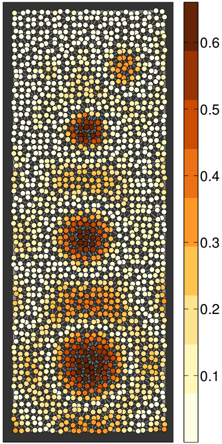
Example



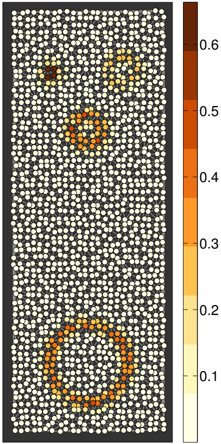
$w_f(1) : \mathcal{V} \to \mathbb{R}$



$w_f(2) : \mathcal{V} \to \mathbb{R}$



$w_f(3) : \mathcal{V} \to \mathbb{R}$



$w_f(4) : \mathcal{V} \to \mathbb{R}$

Wavelet coefficients for f.

Spectral graph Wavelets

Spectral graph Wavelets

Graph wavelet transform (Hammond et al., 2011)

Decompose a function f in terms of basis functions $\{\psi_{s,1}, \ldots, \psi_{s,n}\}$ where each $\psi_{s,i}$ depends on a scale s and a location τ_i.

Spectral graph Wavelets

Graph wavelet transform (Hammond et al., 2011)

Decompose a function f in terms of basis functions $\{\psi_{s,1}, \ldots, \psi_{s,n}\}$ where each $\psi_{s,i}$ depends on a scale s and a location τ_i.

Projecting the function f onto spaces with different scales is equivalent to band-pass filtering the function f.

Spectral graph Wavelets

Graph wavelet transform (Hammond et al., 2011)

Decompose a function f in terms of basis functions $\{\psi_{s,1}, \ldots, \psi_{s,n}\}$ where each $\psi_{s,i}$ depends on a scale s and a location τ_i.

Projecting the function f onto spaces with different scales is equivalent to band-pass filtering the function f.

$$\widehat{T_g f}(\lambda_\ell) := \widehat{g}(\lambda_\ell)\widehat{f}(\lambda_\ell).$$

$$(T_g f)(i) = \sum_{\ell=0}^{N} \widehat{g}(\lambda_\ell)\widehat{f}(\lambda_\ell)\mathbf{u}_\ell(i).$$

Graph wavelet transform (Hammond et al., 2011)

Decompose a function f in terms of basis functions $\{\psi_{s,1}, \ldots, \psi_{s,n}\}$ where each $\psi_{s,i}$ depends on a scale s and a location τ_i.

Projecting the function f onto spaces with different scales is equivalent to band-pass filtering the function f.

$$\widehat{T_g f}(\lambda_\ell) := \widehat{g}(\lambda_\ell)\widehat{f}(\lambda_\ell).$$

$$(T_g f)(i) = \sum_{\ell=0}^{N} \widehat{g}(\lambda_\ell)\widehat{f}(\lambda_\ell)\mathbf{u}_\ell(i).$$

$$\psi_{s,i} = T_g^s \boldsymbol{\delta}_i,$$

$$\psi_{s,i} = \mathbf{U}\mathbf{G}\mathbf{U}^\intercal \boldsymbol{\delta}_i.$$

$$\mathbf{G} = \begin{bmatrix} \widehat{g}(s\lambda_0) & & 0 \\ & \ddots & \\ 0 & & \widehat{g}(s\lambda_{N-1}) \end{bmatrix}$$

Graph wavelet coefficients (Hammond et al., 2011)

Wavelet coefficients at scale s and node τ_i are obtained through the dot product

$$\omega_f(s, i) = \psi_{s,i}^{\mathsf{T}} f.$$

Graph wavelet coefficients (Hammond et al., 2011)

Wavelet coefficients at scale s and node τ_i are obtained through the dot product

$$\omega_f(s, i) = \psi_{s,i}^{\intercal} f.$$

$$w_f(s) = \mathbf{U}\mathbf{G}\mathbf{U}^{\intercal}\mathbf{f},$$

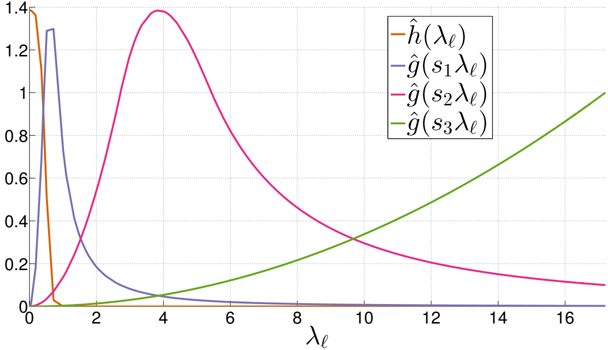
$$\mathbf{G} = \begin{bmatrix} \widehat{g}(s\lambda_0) & & 0 \\ & \ddots & \\ 0 & & \widehat{g}(s\lambda_{N-1}) \end{bmatrix}.$$

Filter functions

$$\widehat{h}(x) = \gamma \exp\left(\left(\frac{x}{0.6\lambda_{\min}}\right)^4\right)$$

$$\widehat{g}(x) = \begin{cases} x_1^{-2}x^2 & \text{for } x < x_1 \\ p(x) & \text{for } x_1 \le x \le x_2 \\ x_2^2 x^{-2} & \text{for } x > x_2 \end{cases}$$

$$p(x) = -5 + 11x - 6x^2 + x^3$$

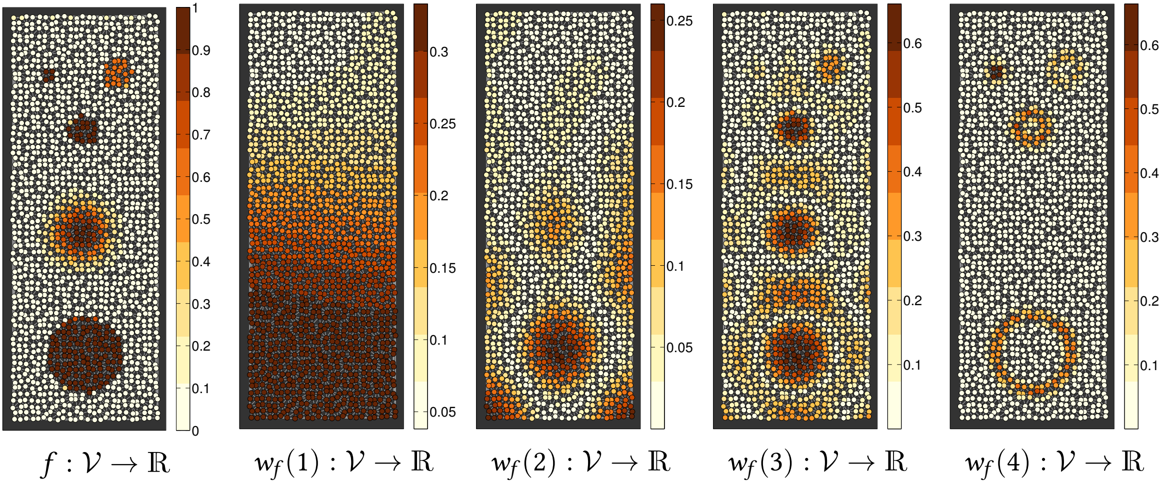


The scales are logarithmically sampled between

$$s_{\min} = s_1, s_2, \ldots, s_m = s_{\max},$$
$$s_{\min} = x_2/\lambda_{\max}, \text{ and}$$
$$s_{\max} = 20x_2/\lambda_{\max}.$$

Hammond et al., *"Wavelets on graphs via spectral graph theory"*, 2011.

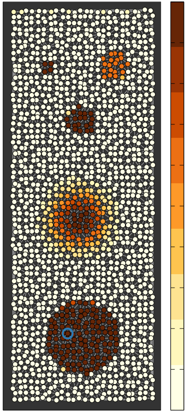
Example



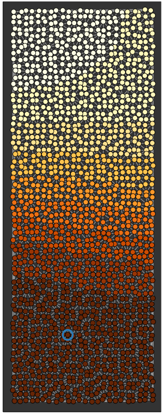
Wavelet coefficients for f.

Spatiotemporal visual analytics

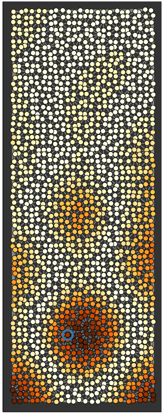
Example



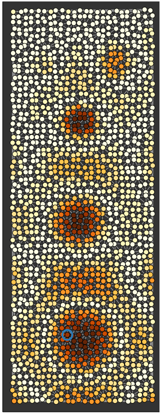
$f : \mathcal{V} \to \mathbb{R}$



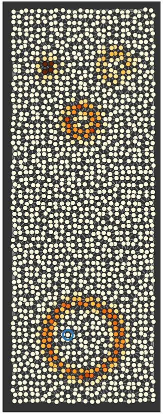
$w_f(1) : \mathcal{V} \to \mathbb{R}$



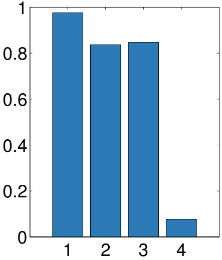
$w_f(2) : \mathcal{V} \to \mathbb{R}$



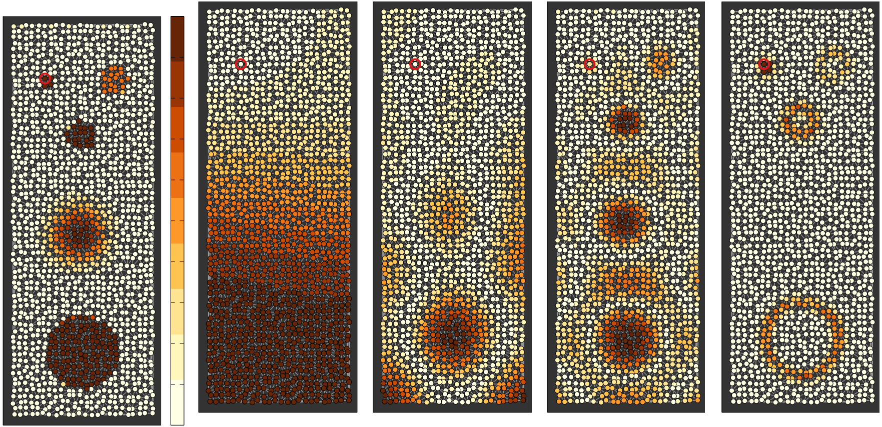
$w_f(3) : \mathcal{V} \to \mathbb{R}$



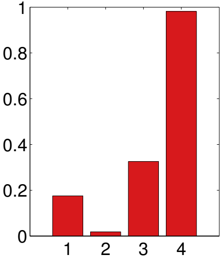
$w_f(4) : \mathcal{V} \to \mathbb{R}$



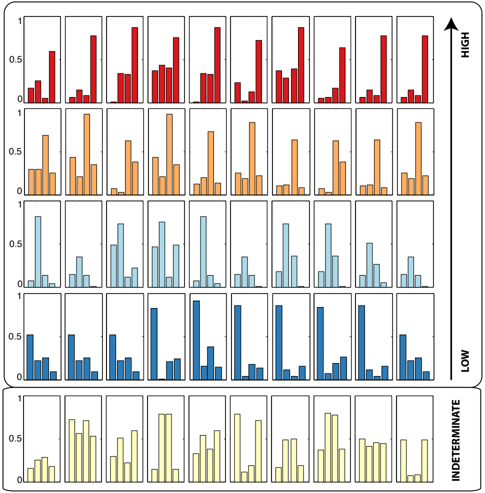
Example



$f : \mathcal{V} \to \mathbb{R}$ $\quad w_f(1) : \mathcal{V} \to \mathbb{R}$ $\quad w_f(2) : \mathcal{V} \to \mathbb{R}$ $\quad w_f(3) : \mathcal{V} \to \mathbb{R}$ $\quad w_f(4) : \mathcal{V} \to \mathbb{R}$

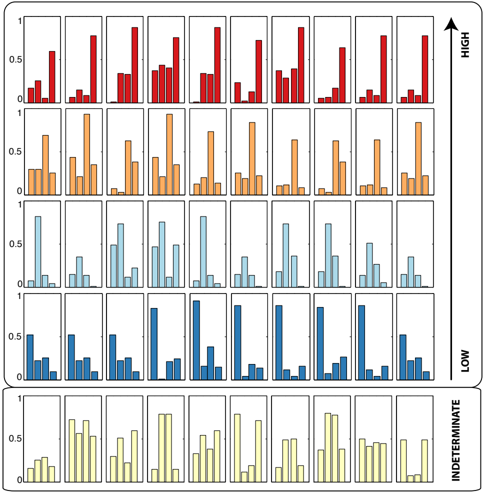


Node classification



Dimensional feature vectors characterizing four frequency classes.

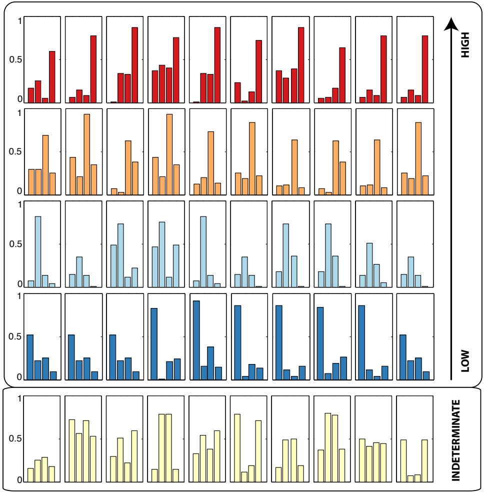
Node classification



Dimensional feature vectors characterizing four frequency classes.

- Supervised classification method.

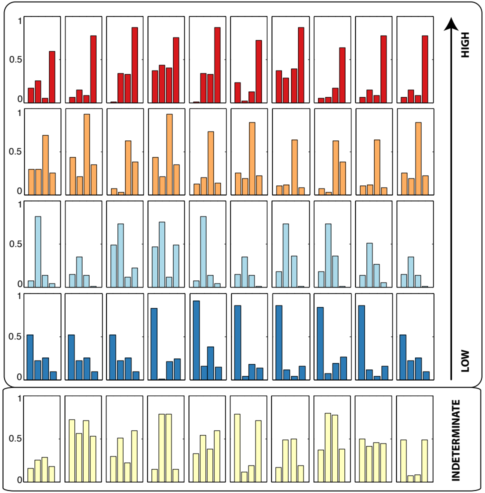
Node classification



Dimensional feature vectors characterizing four frequency classes.

- Supervised classification method.
- Pattern recognition neural network (PRNN).

Node classification

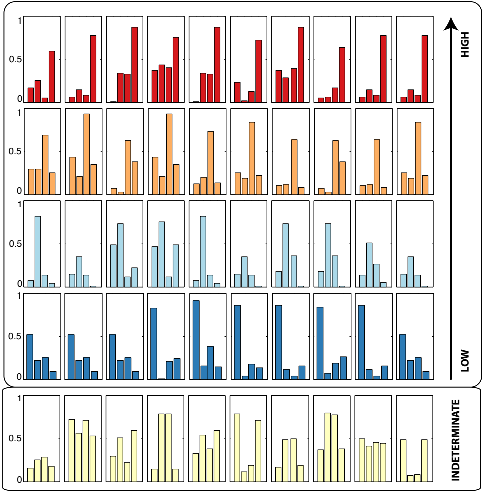


Dimensional feature vectors characterizing four frequency classes.

- Supervised classification method.
- Pattern recognition neural network (PRNN).
- Training function: Scaled conjugate gradient [1].

[1] Møller, *"A scaled conjugate gradient algorithm for fast supervised learning"*, 1993.

Node classification

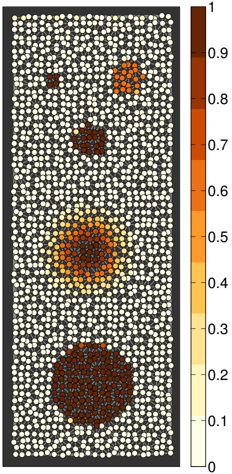


Dimensional feature vectors characterizing four frequency classes.

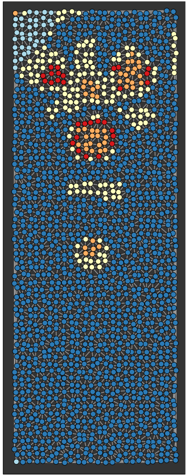
- Supervised classification method.
- Pattern recognition neural network (PRNN).
- Training function: Scaled conjugate gradient [1].
- 25 units in the hidden layer, 5 units in the output layer.

[1] Møller, *"A scaled conjugate gradient algorithm for fast supervised learning"*, 1993.

Example



Graph function.



Wavelet coefficients classification.

High frequencies

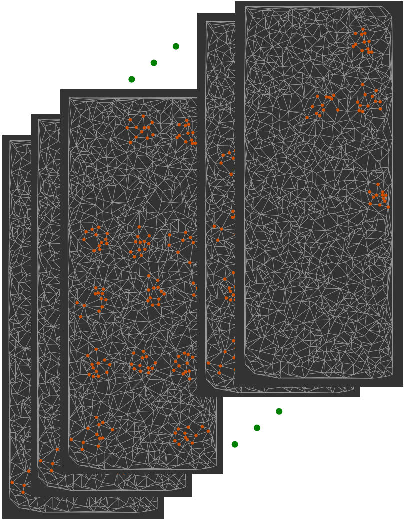
Mid-high frequencies

Indeterminate

Mid-low frequencies

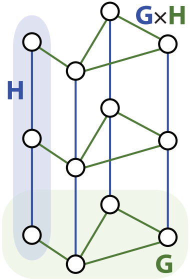
Low frequencies

Time-varying data



$f : \mathcal{V}_{\mathcal{G}} \times [a, b] \to \mathbb{R}$, for every time t in the interval $[a, b]$, assigns a real scalar to each node $\tau_i \in \mathcal{V}$.

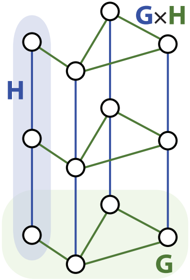
Time-varying data



Cartesian product $\mathcal{G} \times \mathcal{H}$.

Graph structure

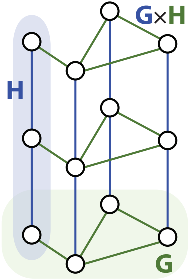
- $\mathcal{G} = (\mathcal{V}_\mathcal{G}, \mathcal{E}_\mathcal{G})$, with nodes τ_i.
- $\mathcal{H} = (\mathcal{V}_\mathcal{H}, \mathcal{E}_\mathcal{H})$, linear graph with nodes ι_j.
- $\mathcal{G} \times \mathcal{H} = (\mathcal{V}_\mathcal{H}, \mathcal{E}_\mathcal{H})$. $\mathcal{V}_\mathcal{H} = \{\iota_1, \iota_2, \ldots, \iota_r\}$ and edges $\mathcal{E}_\mathcal{H} = \{e_{ii+1} = \iota_i \iota_{i+1}\}_{i=1,\ldots,r-1}$.
- $\mathcal{G} \times \mathcal{H}$ can be seen as copies of $\mathcal{G}$ stacked according to the nodes of $\mathcal{H}$.



Cartesian product $\mathcal{G} \times \mathcal{H}$.

Graph function

- $f : \mathcal{V}_\mathcal{G} \times [a, b] \to \mathbb{R}$, for every time t in the interval $[a, b]$, assigns a real scalar to each node $\tau_i \in \mathcal{V}$.
- The time-varying function f can naturally be extended to $\mathcal{G} \times \mathcal{H}$ through the function $f_{\mathcal{G}\times\mathcal{H}} : \mathcal{V}_\mathcal{G} \times \mathcal{V}_\mathcal{H} \to \mathbb{R}$ such that $f_{\mathcal{G}\times\mathcal{H}}((\tau_i, \iota_j)) = f(\tau_i, t_j)$.

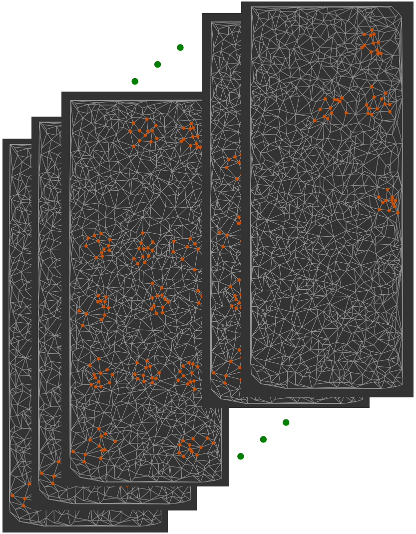


Cartesian product
$\mathcal{G} \times \mathcal{H}$.

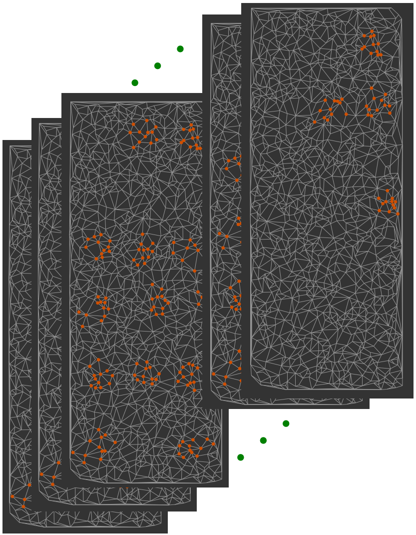
Spectrum of $\mathcal{G} \times \mathcal{H}$

- $\mathbf{u}_j$, λ_j be the eigenvectors and eigenvalues of $\mathcal{G}$ and
- $\mathbf{v}_k$, μ_k be the eigenvectors and eigenvalues of $\mathcal{H}$, then

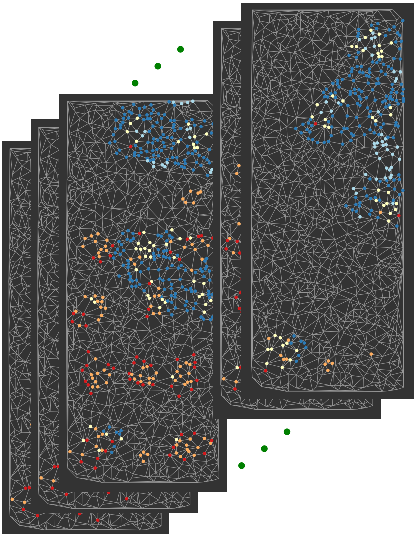
- $\lambda_j + \mu_k$ is an eigenvalue of $\mathcal{G} \times \mathcal{H}$ and
- $\mathbf{w}_{jk} = \mathbf{u}_j \otimes \mathbf{v}_k$ the corresponding eigenvector.



$$f_{\mathcal{G}\times\mathcal{H}} : \mathcal{V}_{\mathcal{G}} \times \mathcal{V}_{\mathcal{H}} \to \mathbb{R}$$



$$f_{\mathcal{G}\times\mathcal{H}} : \mathcal{V}_{\mathcal{G}} \times \mathcal{V}_{\mathcal{H}} \to \mathbb{R}$$



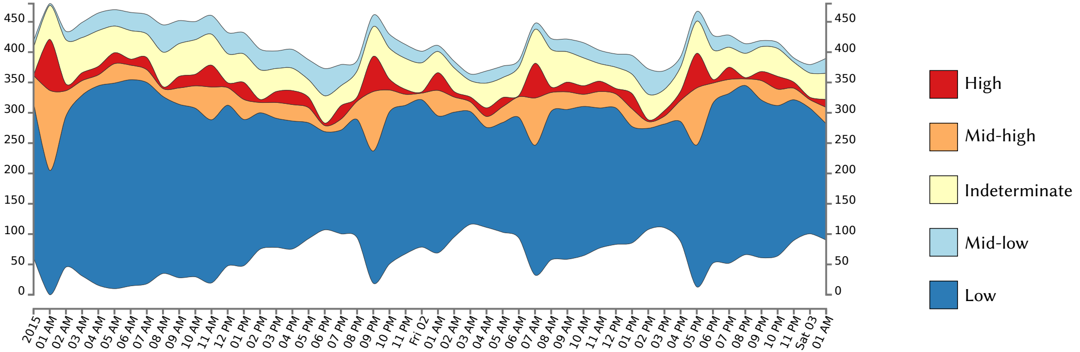
$$w_{\mathcal{G}\times\mathcal{H}} : \mathcal{V}_{\mathcal{G}} \times \mathcal{V}_{\mathcal{H}} \to \mathbb{R}$$

Stacked graph

For each time t_i, we can count the number of nodes that belong to a specific class in that time.

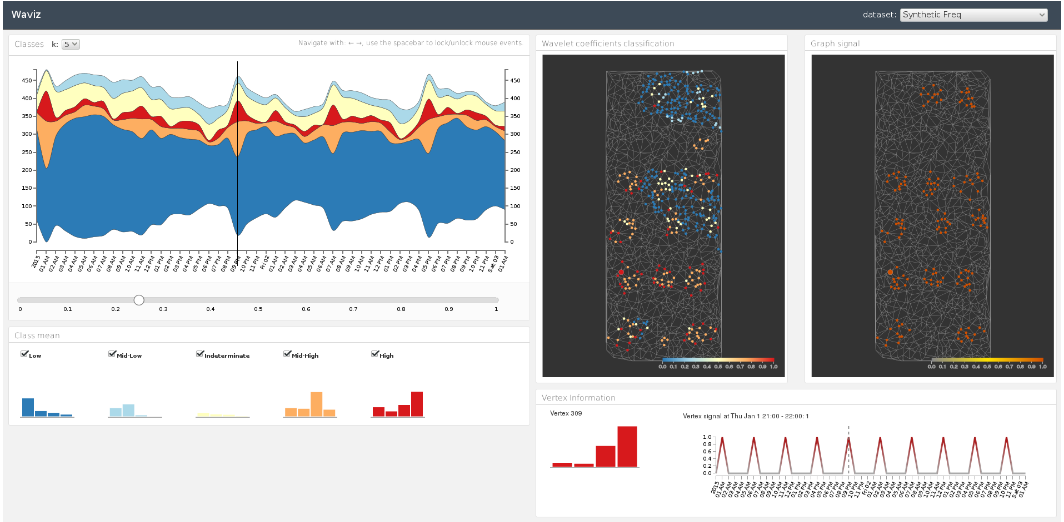
Stacked graph

For each time t_i, we can count the number of nodes that belong to a specific class in that time.



Stacked graph illustrating how each of the five classes evolve over time.

Visual elements and linked views

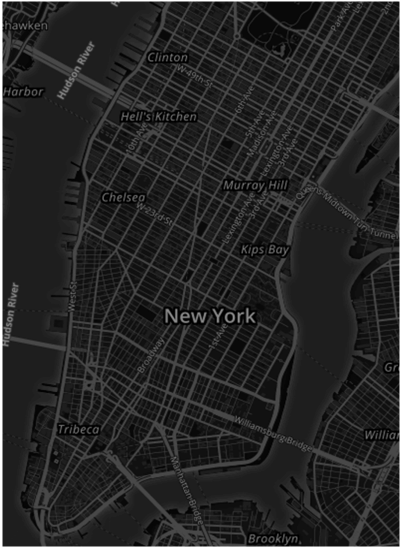


Pipeline

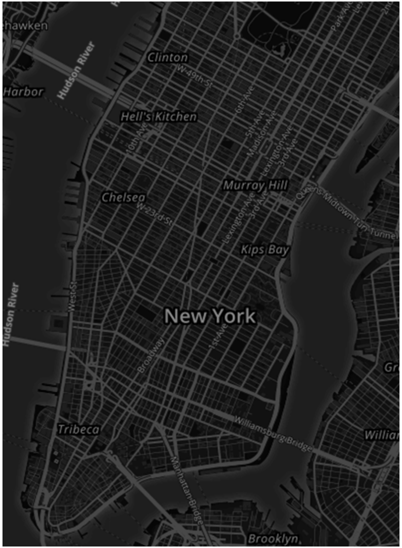

WAVELET TRANSFORM
FEATURE VECTOR
FEATURE CLASSIFICATION
VISUAL ANALYSIS

Case studies

NY graph


Clinton
Hudson River
Harbor
Hell's Kitchen
Chelsea
Murray Hill
Kips Bay
New York
Tribeca
Williamsburg Bridge
Manhattan Bridge
Brooklyn
Queens Midtown Tunnel
Broadway

NY graph


hawken
Clinton
W 49th St
Harbor
Hudson River
Hell's Kitchen
Chelsea
W 23rd St
Murray Hill
Kips Bay
New York
Tribeca
Williamsburg Bridge
Brooklyn
Manhattan Bridge
Broadway
Queens Midtown Tunnel


hawken
Harbor
Hudson River
Brooklyn

NY Taxi

The dataset contains 14 attributes:

Information about the taxi and the driver: medallion, hack_license, vendor_id, rate_code, store_and_fwd_flag.

Information about the time of the trip: pickup_datetime , dropoff_datetime, trip_time_in_secs.

Geographical start and end of each trip: pickup_longitude, pickup_latitude, dropoff_longitude, dropoff_latitude.

Other information: passenger_count, trip_distance.

NY Taxi

The dataset contains 14 attributes:

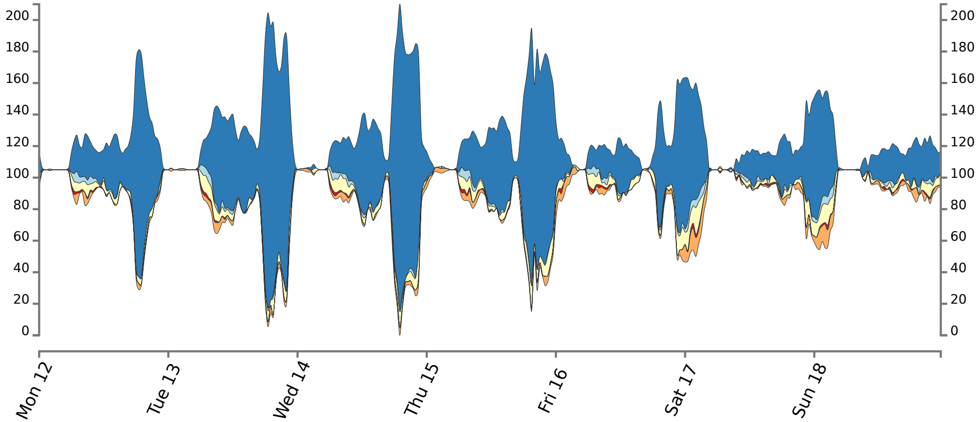
Information about the taxi and the driver: medallion, hack_license, vendor_id, rate_code, store_and_fwd_flag.

Information about the time of the trip: **pickup_datetime**, dropoff_datetime, trip_time_in_secs.

Geographical start and end of each trip: **pickup_longitude, pickup_latitude**, dropoff_longitude, dropoff_latitude.

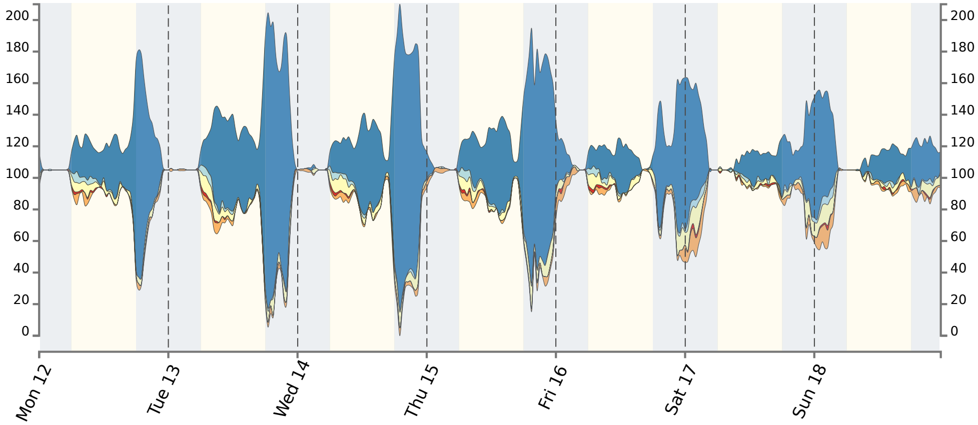
Other information: passenger_count, trip_distance.

NY Taxi



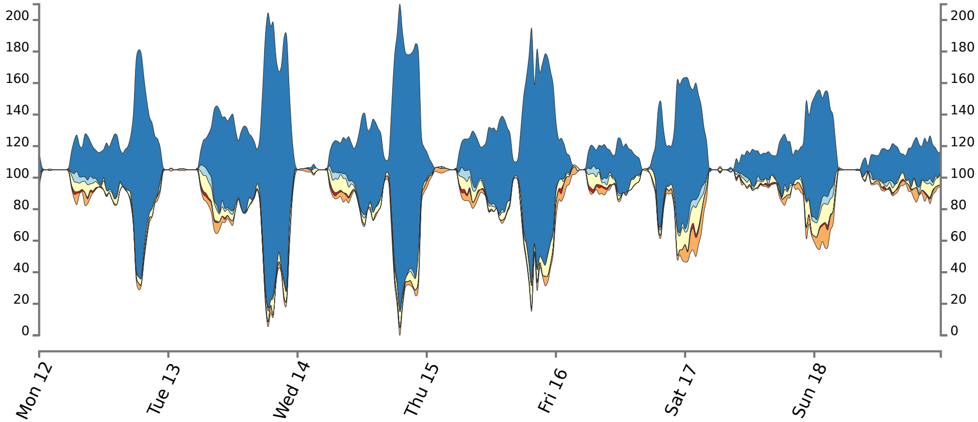
Stacked graph of taxi pickups from August 12th to 18th, 2013 ($\sim$ 3M records).
Data aggregated in periods of thirty minutes: $\mathcal{G} \times \mathcal{H}$ with $\sim 1,5$M nodes.

NY Taxi



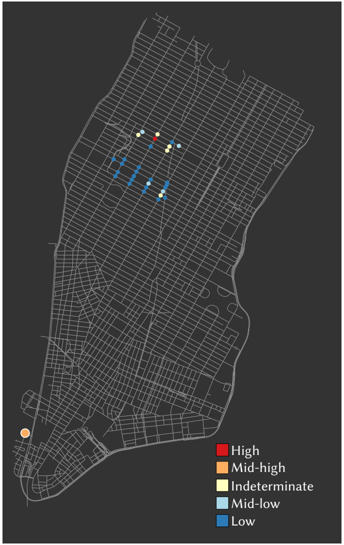
- Pattern of activity of each day.
- Difference between day and night.

NY Taxi

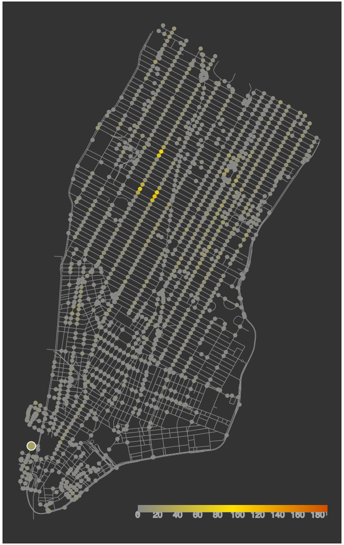


- Predominance of low frequency.
- Temporal variation occurs as a large, uniform event, without many peaks of activity.

NY Taxi



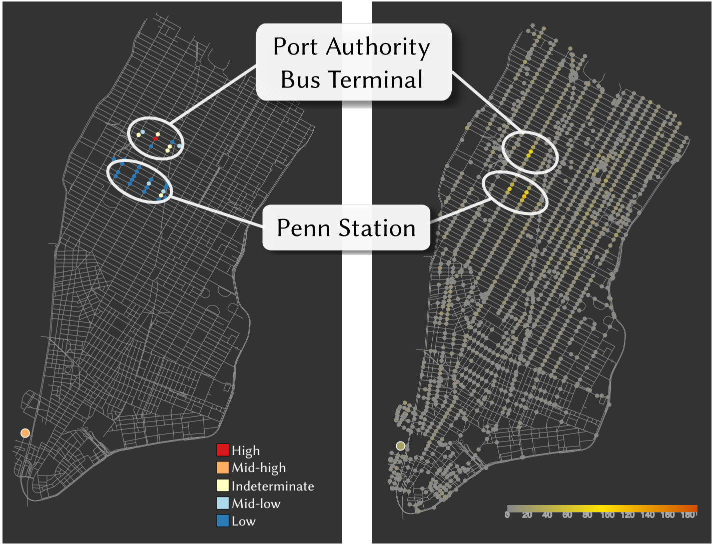
Wavelets Coefficients



Density graph

7:30am to 8am, Monday
August 12, 2013

NY Taxi

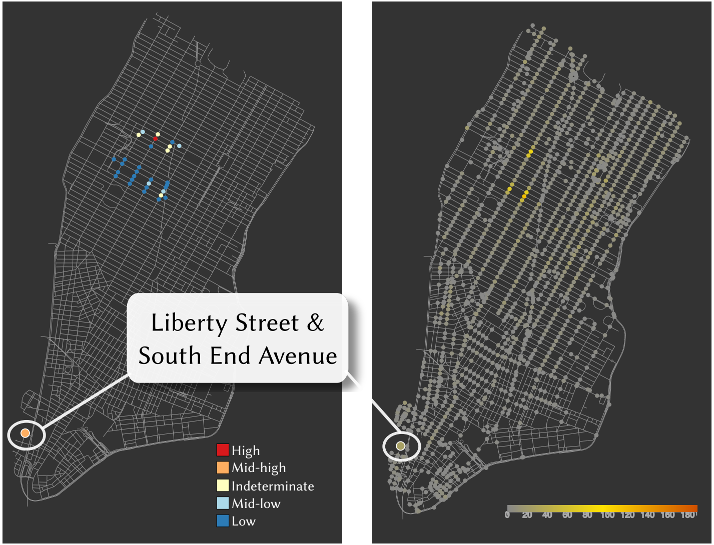


Wavelets Coefficients

Density graph

7:30am to 8am, Monday August 12, 2013

NY Taxi

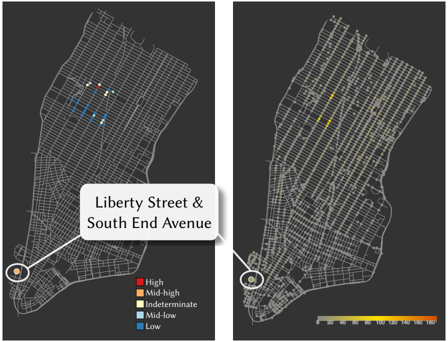


Wavelets Coefficients

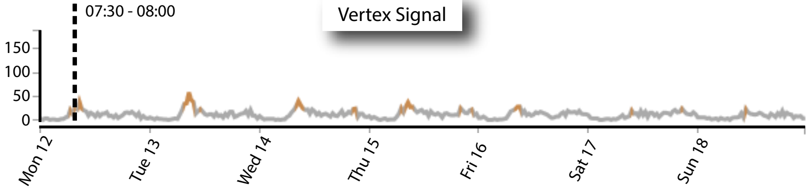
Density graph

7:30am to 8am, Monday
August 12, 2013

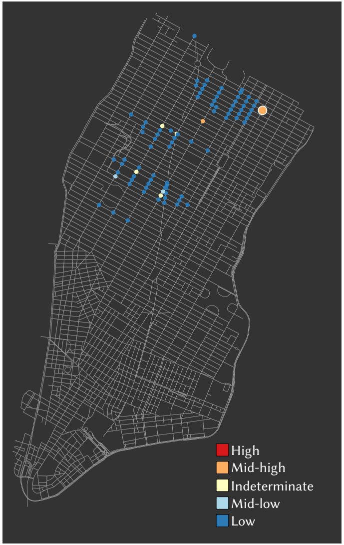
NY Taxi



7:30am to 8am, Monday
August 12, 2013



NY Taxi

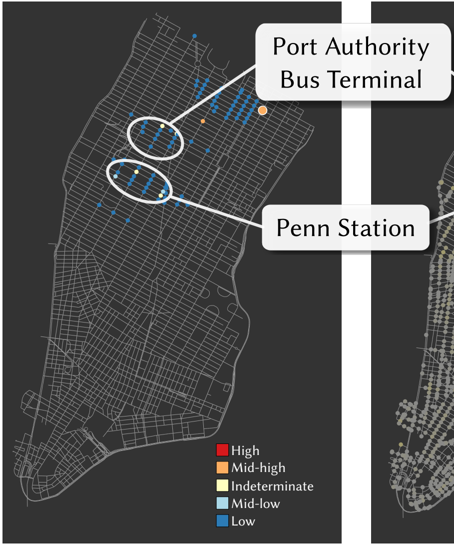


Wavelets Coefficients

Density graph

7:30pm to 8pm, Friday
August 16, 2013

NY Taxi

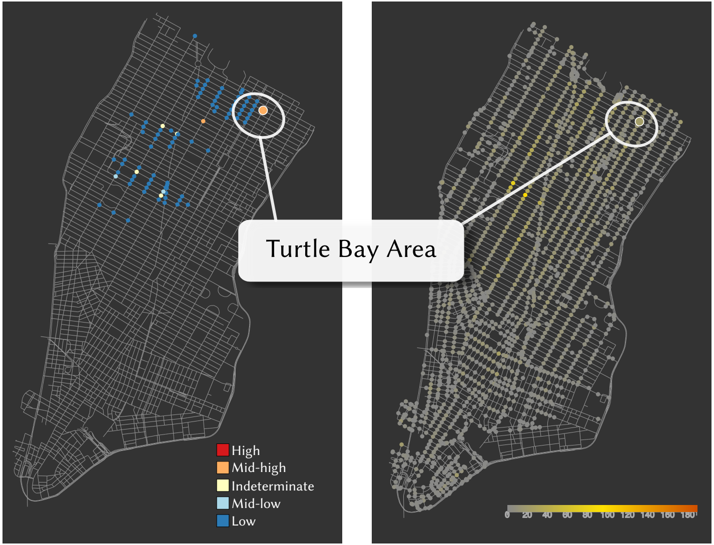


Wavelets Coefficients

Density graph

7:30pm to 8pm, Friday
August 16, 2013

NY Taxi

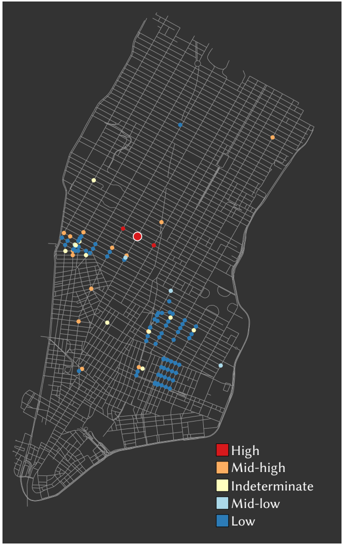


Wavelets Coefficients

Density graph

7:30pm to 8pm, Friday
August 16, 2013

NY Taxi



Wavelets Coefficients

Density graph

1:30am to 2am, Saturday
August 17, 2013

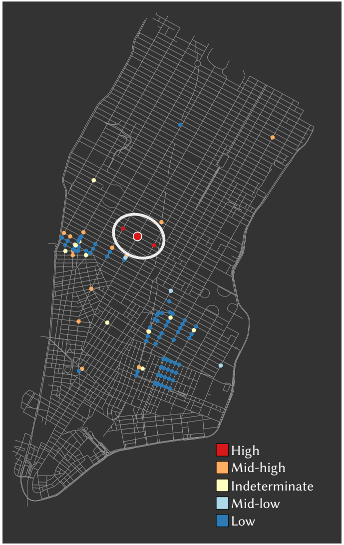
NY Taxi

Wavelets Coefficients

Density graph

1:30am to 2am, Saturday
August 17, 2013

NY Taxi

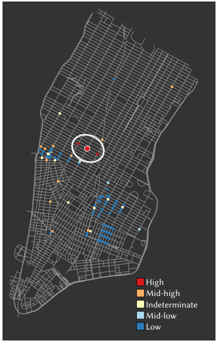


Wavelets Coefficients

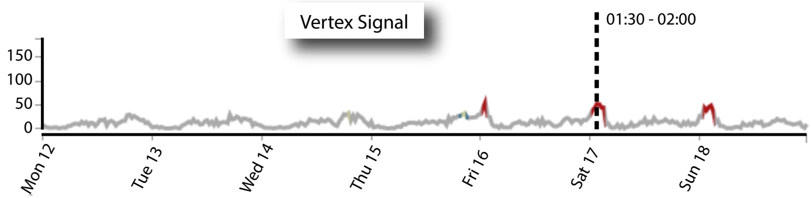
Density graph

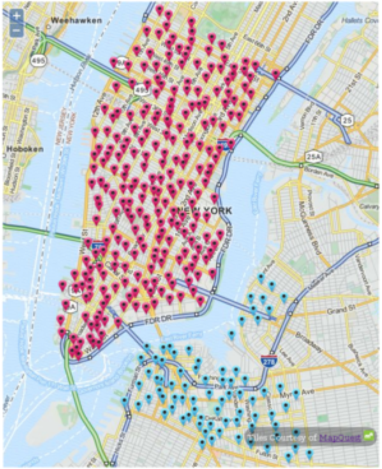
1:30am to 2am, Saturday
August 17, 2013

NY Taxi



1:30am to 2am, Saturday
August 17, 2013





Citi Bike NY station map.

Stations from downtown Manhattan (pink markers) are considered in this study.

Citi Bike NY

The dataset contains 15 attributes:

Information about the bike: bikeid.

Information about the user: usertype, birth_year, gender.

Information about the time of the trip: starttime, stoptime , tripduration.

Information about the start station of each trip: start_station_id, start_station_name, start_station_latitude, start_station_longitude.

Information about the end station of each trip: end_station_id, end_station_name, end_station_latitude, end_station_longitude.

Citi Bike NY

The dataset contains 15 attributes:

Information about the bike: bikeid.

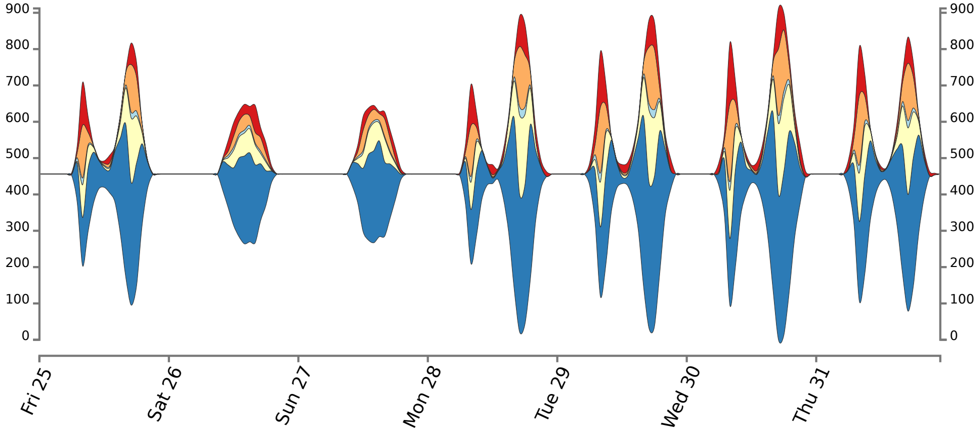
Information about the user: usertype, birth_year, gender.

Information about the time of the trip: **starttime, stoptime**, tripduration.

Information about the start station of each trip: start_station_id, start_station_name, **start_station_latitude, start_station_longitude**.

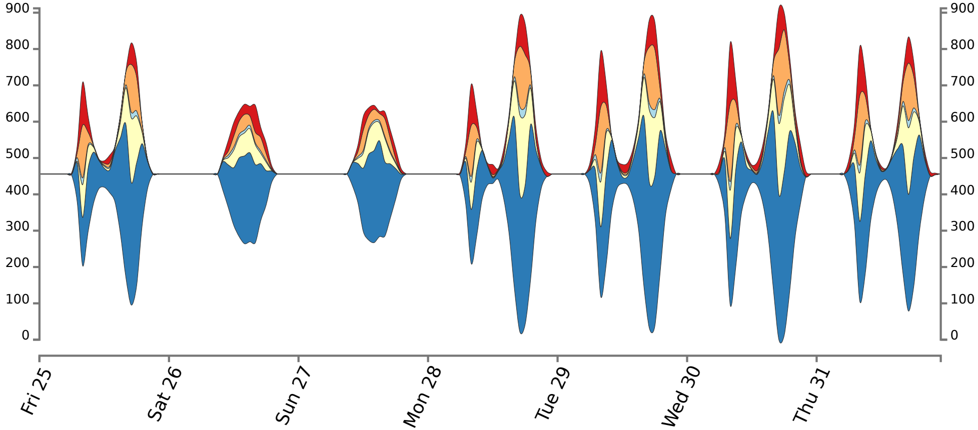
Information about the end station of each trip: end_station_id, end_station_name, **end_station_latitude, end_station_longitude**.

Citi Bike NY



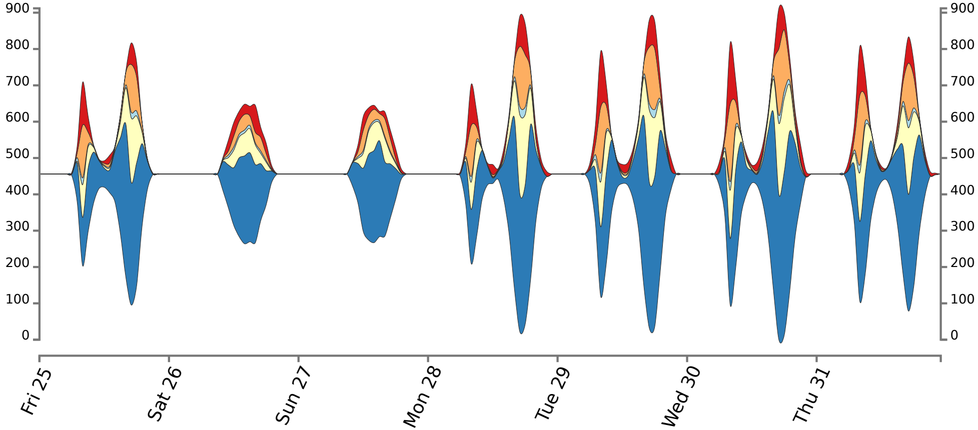
Stacked graph of bike activities from October 25th to 31st, 2013.
Data aggregated in periods of one hour: $\mathcal{G} \times \mathcal{H}$ with $\sim$ 788k nodes.

Citi Bike NY



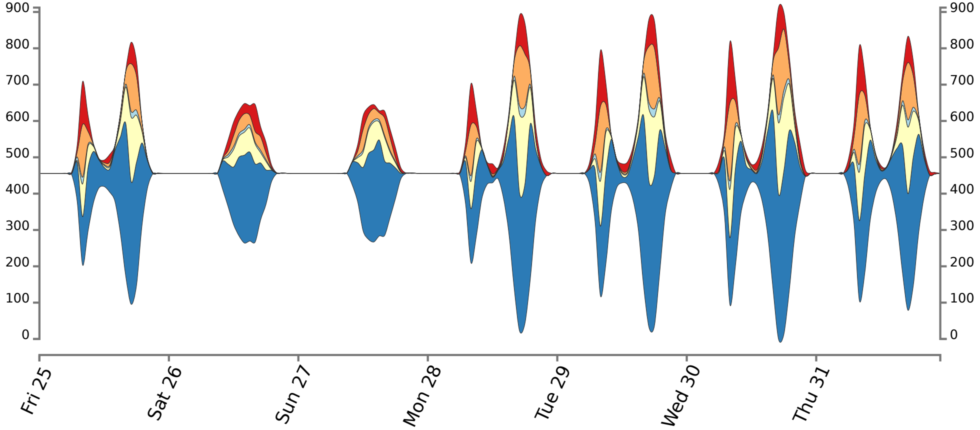
- Pattern of activity of each day.
- Difference between working days and weekend.

Citi Bike NY



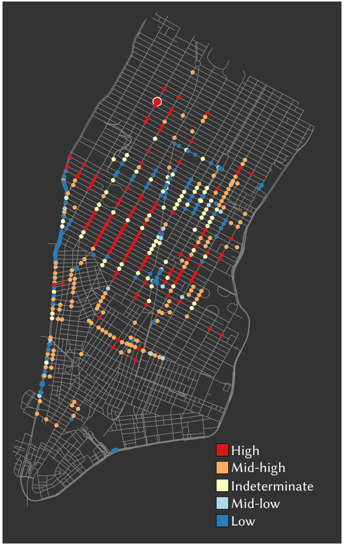
- Working days: more activity during afternoon.
- Low frequency predominant in afternoons.

Citi Bike NY

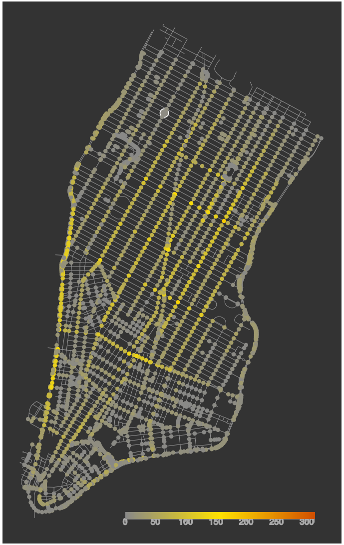


- Thursday 31st, morning and afternoon peaks are almost of the same size.

Citi Bike NY



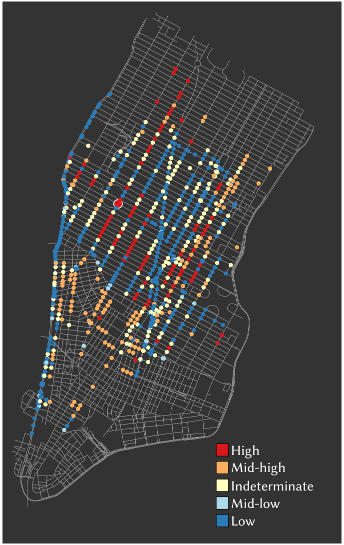
Wavelets Coefficients



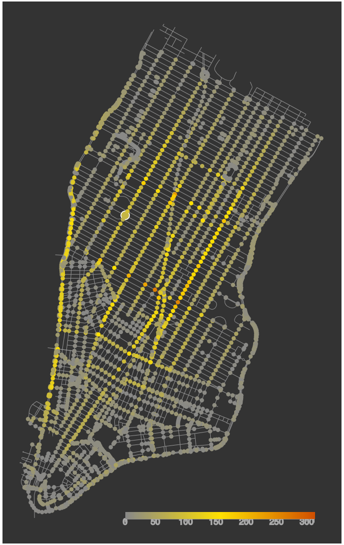
Density graph

8am to 9am, Monday
October 28, 2013

Citi Bike NY



Wavelets Coefficients

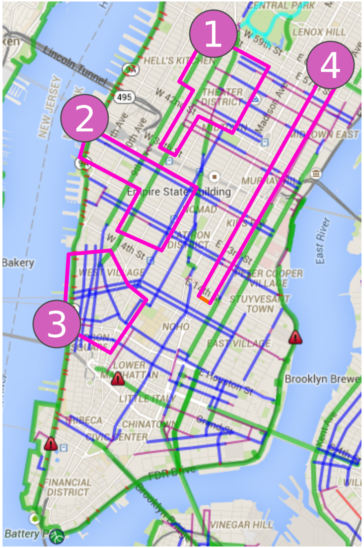


Density graph

5pm to 6pm, Monday
October 28, 2013

Citi Bike NY

Wavelets Coefficients



Bike lanes map

Regions 1, 2, and 3 correspond to areas with a high concentration of bike lanes.

Region 4 is next to the main bike lines.

Discussion and limitations

- Large amounts of information.

Discussion and limitations

- Large amounts of information.
- Visualization of spatiotemporal events, regardless of their size.

Discussion and limitations

- Large amounts of information.
- Visualization of spatiotemporal events, regardless of their size.
- Algorithm to approximate routes.

Discussion and limitations

- Large amounts of information.
- Visualization of spatiotemporal events, regardless of their size.
- Algorithm to approximate routes.
- Classification scheme turned out satisfactory.

Discussion and limitations

- Large amounts of information.
- Visualization of spatiotemporal events, regardless of their size.
- Algorithm to approximate routes.
- Classification scheme turned out satisfactory.
- Training data.

Discussion and limitations

- Large amounts of information.
- Visualization of spatiotemporal events, regardless of their size.
- Algorithm to approximate routes.
- Classification scheme turned out satisfactory.
- Training data.
- Stacked graph plot.

Discussion and limitations

- Large amounts of information.
- Visualization of spatiotemporal events, regardless of their size.
- Algorithm to approximate routes.
- Classification scheme turned out satisfactory.
- Training data.
- Stacked graph plot.
- Discriminating spatial and temporal changes.

Final considerations

A novel visual analytic methodology for analyzing time-varying data that combines graph wavelet theory, pattern classification, and stacked graph visual metaphor in a linked view visualization environment was described.

Final considerations

A novel visual analytic methodology for analyzing time-varying data that combines graph wavelet theory, pattern classification, and stacked graph visual metaphor in a linked view visualization environment was described.

The described methodology has been conditionally accepted by the IEEE VAST 2015 Program Committee to be presented at IEEE VIS into the conference-only track.

Research Planning

Future research

Event mining

- Perform a deeper study of the resulting Wavelets coefficients in order to distinguish different events.
  - Identify and manually label events.
  - Use a clustering method, e.g., SOM, rather than the employed classification method.
- Discriminate spatial from temporal events. First extracting temporal information and then including the spatial dimensions could lead to a better understanding of the events.

Future research

Data modeling

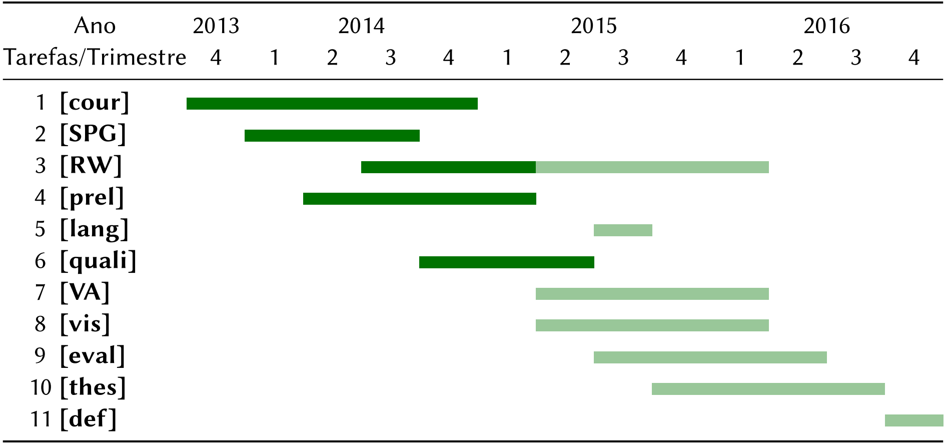
- Consider multivariate time series.
- Consider data from another context, such as LBSNs.

Future research

Visual elements

- Another visual metaphor, in which classes with few elements are not hidden, to allow finding events in time-slices.
- Separation of spatial from temporal events will also have an impact in the visual elements employed.

Proposed timeline



Obrigada!

paolalv@icmc.usp.br

Visual and Geometry Processing Group

ICMC SÃO CARLOS USP