Network Methods for Big Data: Applications to Astroinformatics

Ognyan Kounchev

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, IZKS-University of Bonn

Supported by the Humboldt Foundation and Project between Serbian and Bulgarian Academy of Sciences

X Serbian-Bulgarian Astronomical Conference, Belgrade, 2016

Main purpose of the talk

- Show recent approach of Signal Processing on Graphs/Networks

Main purpose of the talk

- Show recent approach of Signal Processing on Graphs/Networks
- Using Spectral Graph theory

Main purpose of the talk

- Show recent approach of Signal Processing on Graphs/Networks
- Using Spectral Graph theory
- Spectral Graph Wavelets = **Multiscale** Concepts

Main purpose of the talk

- Show recent approach of Signal Processing on Graphs/Networks
- Using Spectral Graph theory
- Spectral Graph Wavelets = **Multiscale** Concepts
- Importance of **Multiscale** approach in analysis of Astronomical Images

Methods of Signal Processing for Analysis of Big Data

The main points:

- Representation and processing of massive data sets with **irregular** structure in the areas

Methods of Signal Processing for Analysis of Big Data

The main points:

- Representation and processing of massive data sets with **irregular** structure in the areas
- Same methods in: Engineering, Astronomy, Social Networks, Biomolecular research (biochemical and genetics), Commerce (consumer behaviour studies), Security

Methods of Signal Processing for Analysis of Big Data

The main points:

- Representation and processing of massive data sets with **irregular** structure in the areas
- Same methods in: Engineering, Astronomy, Social Networks, Biomolecular research (biochemical and genetics), Commerce (consumer behaviour studies), Security
- Our approach: **Discrete Signal Processing** (DSP) on Networks/Graphs using Spectral theory and Wavelets

The main points:

- Representation and processing of massive data sets with **irregular** structure in the areas
- Same methods in: Engineering, Astronomy, Social Networks, Biomolecular research (biochemical and genetics), Commerce (consumer behaviour studies), Security
- Our approach: **Discrete Signal Processing** (DSP) on Networks/Graphs using Spectral theory and Wavelets
- Challenges: Large-scale **filtering** and **frequency analysis**

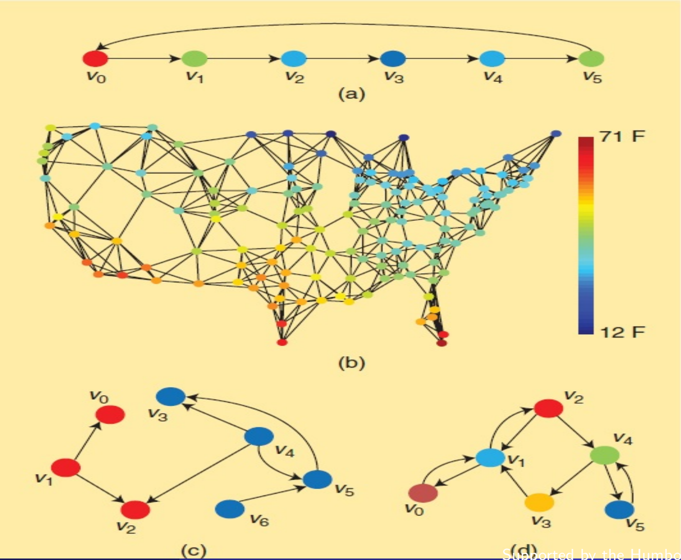
Methods of Signal Processing for Analysis of Big Data

The main points:

- Representation and processing of massive data sets with **irregular** structure in the areas
- Same methods in: Engineering, Astronomy, Social Networks, Biomolecular research (biochemical and genetics), Commerce (consumer behaviour studies), Security
- Our approach: **Discrete Signal Processing** (DSP) on Networks/Graphs using Spectral theory and Wavelets
- Challenges: Large-scale **filtering** and **frequency analysis**
- The **new notions** which generalize those of the classical DSP: **Graph signals**, Graph **filters**, Graph **Fourier transform**, Graph **frequency**, Spectrum **ordering**

3 examples of Signals on graphs

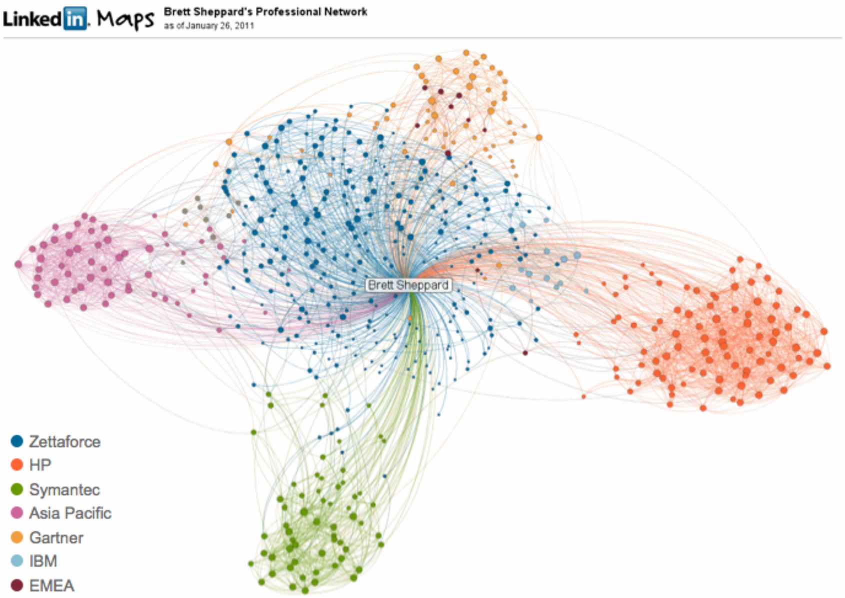
Top: Samples of the signal $\cos(2\pi n/6)$: edges show causality of time;



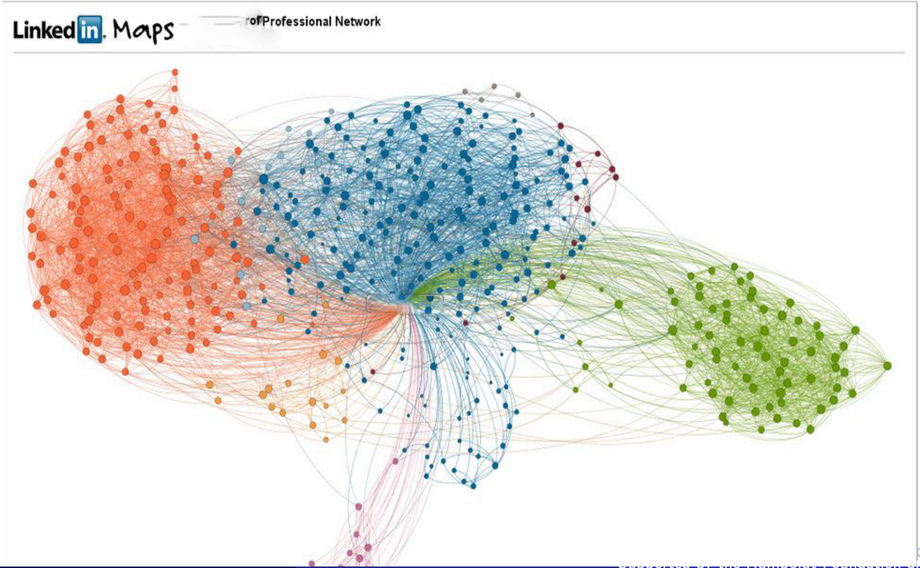
Examples above:

- **Sensor network** – edges show closeness of locations of the sensors;
- **WWW**: nodes = websites with website features, edges = hyperlinks
- **Social Network**: nodes = individuals, edges = connection and strength
- **Astronomical Images** and their Network approximations
- **Network Approximations** have **nodes** which are galaxies, stars, clusters; nodes are intersection of lines along **edges**; closer Nodes represent **similar features** within the data
- **An Examplary OBJECTIVE of OUR WORK**: **Find a method for high speed automatic classification of galaxies**
- **previously done by thousands of volunteers**
- Application to: **images of distant clusters of galaxies that contain several different types of galaxies**
- General approach of **Machine Learning**: This is the art of "teaching" the algorithms the experience of people accumulated for a long time.
- Applications of the **same methods** e.g. in **medicine**, for helping to spot tumours, or in **security**, to find suspicious items in airport scans.

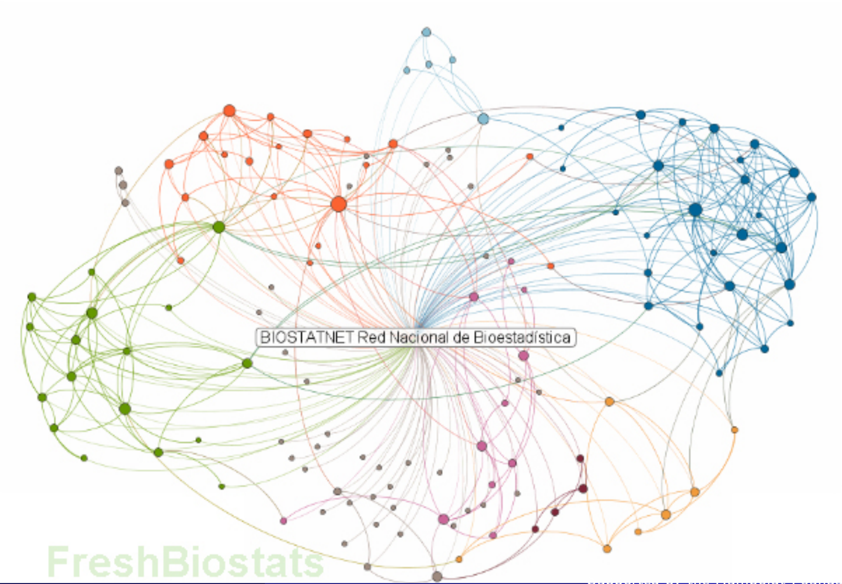
Example of Network - LinkedIn



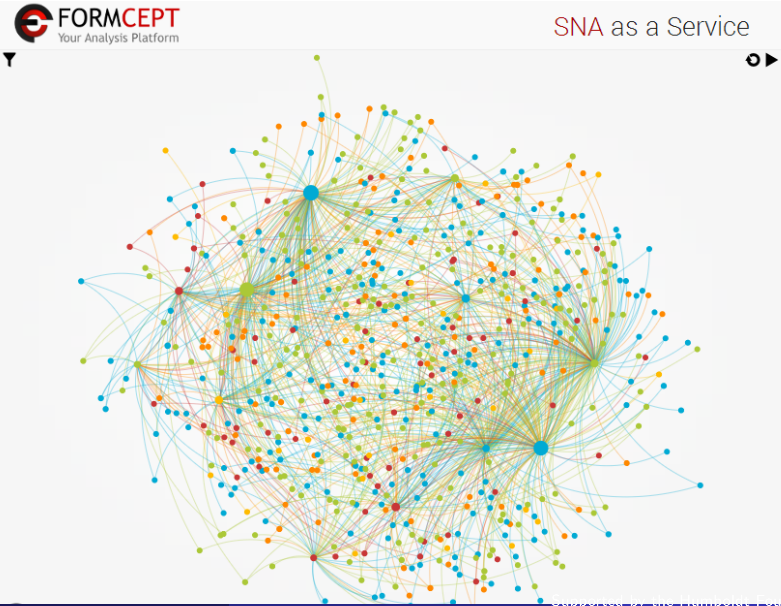
Example - LinkedIn



Biostatistics - example graph



Influencer Network-graph



General principles of SP on graphs

- DSP is applied to data sets in which the elements are related by **dependency**, **similarity**, **physical porximity**, etc.

General principles of SP on graphs

- DSP is applied to data sets in which the elements are related by **dependency**, **similarity**, **physical porximity**, etc.
- data elements $v_j \in G$ = nodes of the graph/network, for $j = 1, 2, ..., N$

General principles of SP on graphs

- DSP is applied to data sets in which the elements are related by **dependency**, **similarity**, **physical porximity**, etc.
- data elements $v_j \in G =$ nodes of the graph/network, for $j = 1, 2, ..., N$
- $A = (a_{ij})$ is the weighted adjacency matrix where a_{ij} shows the **similarity/closeness** between nodes v_i and v_j

General principles of SP on graphs

- DSP is applied to data sets in which the elements are related by **dependency**, **similarity**, **physical porximity**, etc.
- data elements $v_j \in G =$ nodes of the graph/network, for $j = 1, 2, ..., N$
- $A = (a_{ij})$ is the weighted adjacency matrix where a_{ij} shows the **similarity/closeness** between nodes v_i and v_j
- Signal $S = (s_j)_{j=1}^{N}$ is a sequence of numbers which are valued on the nodes v_j

General principles of SP on graphs

- DSP is applied to data sets in which the elements are related by **dependency**, **similarity**, **physical porximity**, etc.
- data elements $v_j \in G$ = nodes of the graph/network, for $j = 1, 2, ..., N$
- $A = (a_{ij})$ is the weighted adjacency matrix where a_{ij} shows the **similarity/closeness** between nodes v_i and v_j
- Signal $S = (s_j)_{j=1}^{N}$ is a sequence of numbers which are valued on the nodes v_j
- The "neighbouring" nodes for the node ν_j is the set

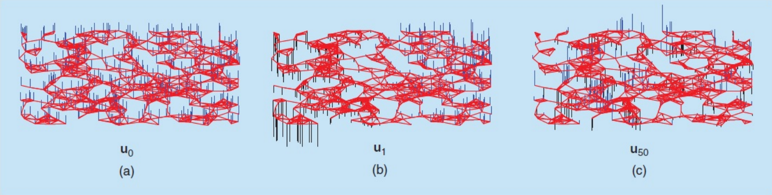
$$N_j$$

Multiscale structure Analysis of Network Signals

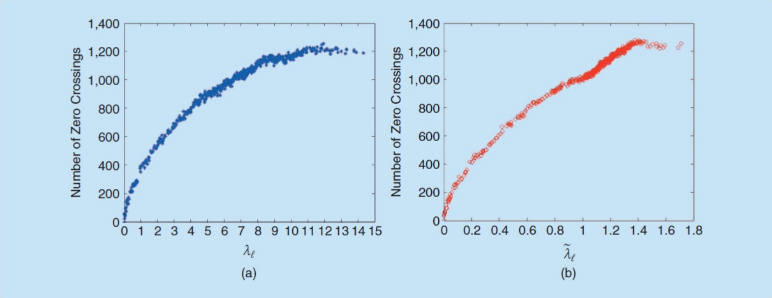
1. Graph shift - generalizes the usual DSP **time delay**; taking average over the Neighbours
$$\overline{s}_n = \sum_{m \in N_n} A_{n,m} s_m$$
2. Graph Fourier transform needs the **Graph Laplace operator** (we assume undirected graph): for a signal f on the graph we define
$$Lf_m := \sum_{n \in N_m} a_{m,n} \left(f_m - f_n \right)$$
3. L is a real symmetric matrix: we denote the N (orthonormal) **eigenvectors and eigenvalues** of L by
$$\chi^{(n)} = \left(\chi_1^{(n)}, ..., \chi_N^{(n)} \right) \in \mathbb{R}^N,$$
$$\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{N-1}.$$

They are the analogue to e^{ikt}, since $-\frac{d^2}{dt^2} e^{ikt} = k^2 e^{ikt}$

Example of eigenvectors



[FIG2] (a)–(c) Three graph Laplacian eigenvectors of a random sensor network graph. The signals' component values are represented by the blue (positive) and black (negative) bars coming out of the vertices. Note that $\mathbf{u}_{50}$ contains many more zero crossings than the constant eigenvector $\mathbf{u}_0$ and the smooth Fiedler vector $\mathbf{u}_1$.



[FIG3] The number of zero crossings, $|\mathcal{Z}_{\mathcal{G}}(\mathbf{u}_\ell)|$ in (a) and $|\mathcal{Z}_{\mathcal{G}}(\tilde{\mathbf{u}}_\ell)|$ in (b), of the unnormalized and normalized graph Laplacian eigenvectors for the random sensor network graph of Figure 2, respectively (the latter of which is defined in the section "Other Graph

- Hence, we have Fourier Analysis

$$f_m = \sum \widehat{f}_n \chi_m^{(n)}$$

Fourier and Wavelet Analysis on Graphs

- Hence, we have Fourier Analysis

$$f_m = \sum \widehat{f}_n \chi_m^{(n)}$$

- One may define **Wavelet Analysis** by using appropriate filter function g, as follows:

$$W_f(p) = \sum_{n=0}^{N-1} g(\lambda_n) \widehat{f}_n \chi_p^{(n)}$$

Fourier and Wavelet Analysis on Graphs

- Hence, we have Fourier Analysis

$$f_m = \sum \widehat{f}_n \chi_m^{(n)}$$

- One may define **Wavelet Analysis** by using appropriate filter function g, as follows:

$$W_f\left(p\right) = \sum_{n=0}^{N-1} g\left(\lambda_n\right) \widehat{f}_n \chi_p^{(n)}$$

- We may define the scale t by putting

$$W_f\left(p; t\right) = \sum_{n=0}^{N-1} g\left(t\lambda_n\right) \widehat{f}_n \chi_p^{(n)}$$

Fourier and Wavelet Analysis on Graphs

- Hence, we have Fourier Analysis

$$f_m = \sum \widehat{f}_n \chi_m^{(n)}$$

- One may define **Wavelet Analysis** by using appropriate filter function g, as follows:

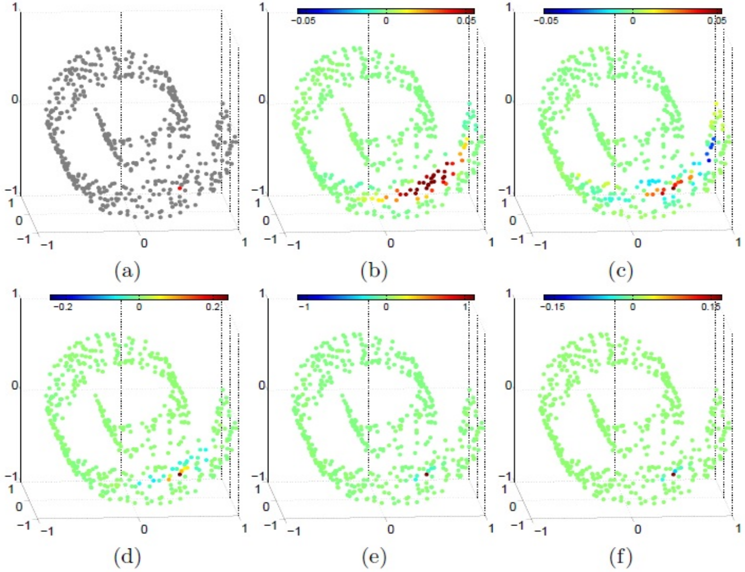
$$W_f\left(p\right) = \sum_{n=0}^{N-1} g\left(\lambda_n\right) \widehat{f}_n \chi_p^{(n)}$$

- We may define the scale t by putting

$$W_f\left(p;t\right) = \sum_{n=0}^{N-1} g\left(t\lambda_n\right) \widehat{f}_n \chi_p^{(n)}$$

- The role of t - the **scale** of the localized at **node** a wavelet $\psi_{t,a}$ where $\left(a \in G\right)$.

Examples of localizations – Swiss roll



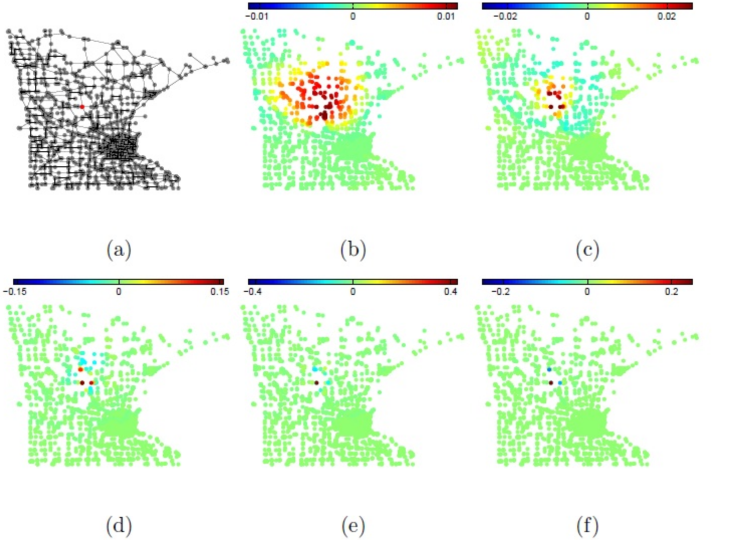


Figure 4: Spectral graph wavelets on Minnesota road graph, with $K = 100$, $J = 4$ scales. (a) vertex at which wavelets are centered (b) scaling function (c)-(f) wavelets, scales 1-4.

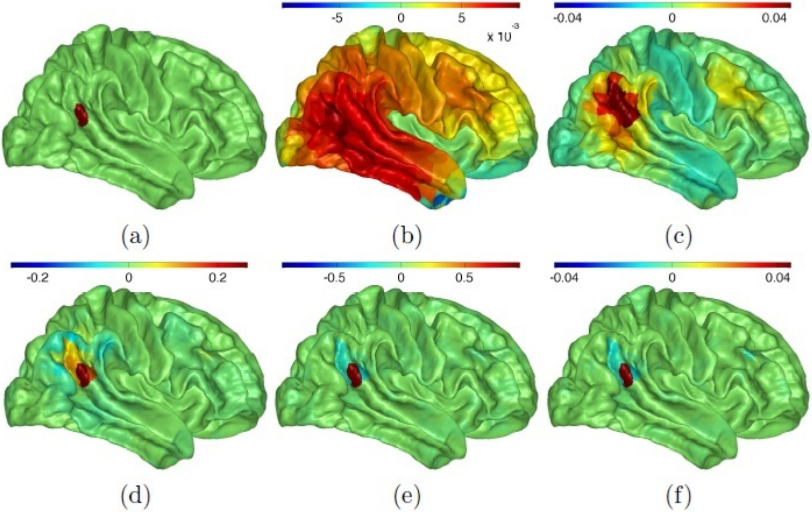


Figure 5: Spectral graph wavelets on cerebral cortex, with $K = 50$, $J = 4$ scales. (a) ROI at which wavelets are centered (b) scaling function (c)-(f) wavelets, scales 1-4.

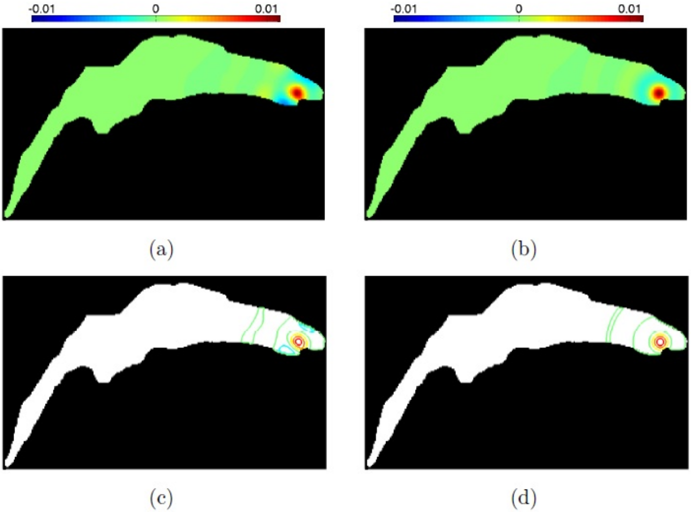
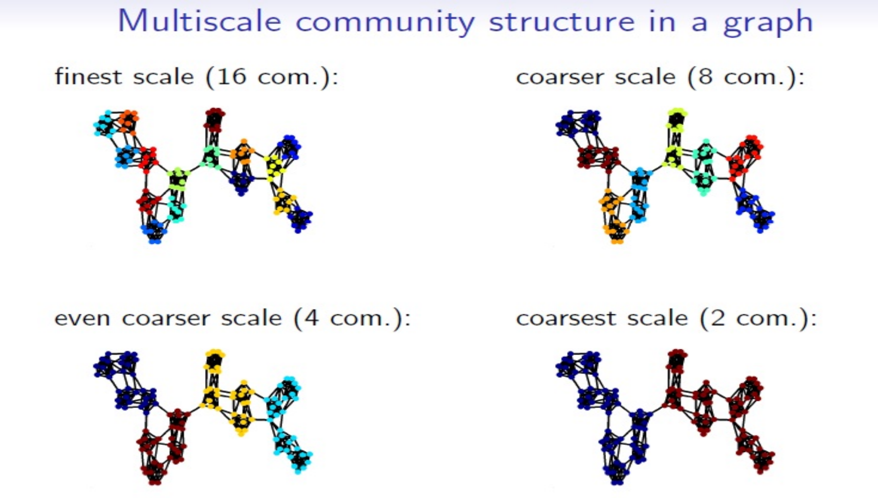
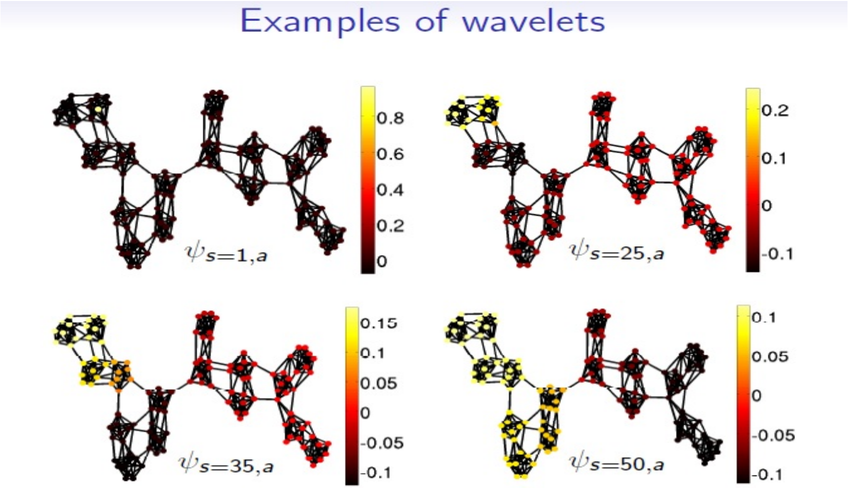


Figure 6: Spectral graph wavelets on lake Geneva domain, (spatial map (a), contour plot (c)); compared with truncated wavelets from graph corresponding to complete mesh (spatial map (b), contour plot (d)). Note that the graph wavelets adapt to the geometry of the domain.

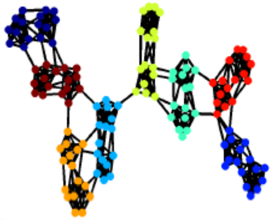
Main Objective: Clusterization (Community Detection on graphs)




Examples of wavelets
$\psi_{s=1,a}$
0.8
0.6
0.4
0.2
0
$\psi_{s=25,a}$
0.2
0.1
0
-0.1
$\psi_{s=35,a}$
0.15
0.1
0.05
0
-0.05
-0.1
$\psi_{s=50,a}$
0.1
0.05
0
-0.05
-0.1

Comparison with other methods for Community Detection

Classical community detection algorithm do not have this "scale-vision" of a graph. Modularity optimisation finds:



Possible Applications outside and in Astronomy

- Remote Sensing Network for Time-Sensitive Detection of Fine Scale Damage to Transportation Infrastructure

Possible Applications outside and in Astronomy

- Remote Sensing Network for Time-Sensitive Detection of Fine Scale Damage to Transportation Infrastructure
- Expoiting existing ground-based remote sensing Networks to improve High-resolution weather forecasts.

Possible Applications outside and in Astronomy

- Remote Sensing Network for Time-Sensitive Detection of Fine Scale Damage to Transportation Infrastructure
- Expoiting existing ground-based remote sensing Networks to improve High-resolution weather forecasts.
- Neural Network Applications in High-Resolution Atmospheric Remote Sensing

Possible Applications outside and in Astronomy

- Remote Sensing Network for Time-Sensitive Detection of Fine Scale Damage to Transportation Infrastructure
- Expoiting existing ground-based remote sensing Networks to improve High-resolution weather forecasts.
- Neural Network Applications in High-Resolution Atmospheric Remote Sensing
- Development of a Real Time Remote Sensing Network to Monitor Small Fluctuations in Urban Air Quality

Possible Applications outside and in Astronomy

- Remote Sensing Network for Time-Sensitive Detection of Fine Scale Damage to Transportation Infrastructure
- Expoiting existing ground-based remote sensing Networks to improve High-resolution weather forecasts.
- Neural Network Applications in High-Resolution Atmospheric Remote Sensing
- Development of a Real Time Remote Sensing Network to Monitor Small Fluctuations in Urban Air Quality
- Network clustering by graph coloring: An application to **astronomical** images

Possible Applications outside and in Astronomy

- Remote Sensing Network for Time-Sensitive Detection of Fine Scale Damage to Transportation Infrastructure
- Expoiting existing ground-based remote sensing Networks to improve High-resolution weather forecasts.
- Neural Network Applications in High-Resolution Atmospheric Remote Sensing
- Development of a Real Time Remote Sensing Network to Monitor Small Fluctuations in Urban Air Quality
- Network clustering by graph coloring: An application to **astronomical** images
- Astronomical image segmentation by Networks and wavelets

Possible Applications outside and in Astronomy

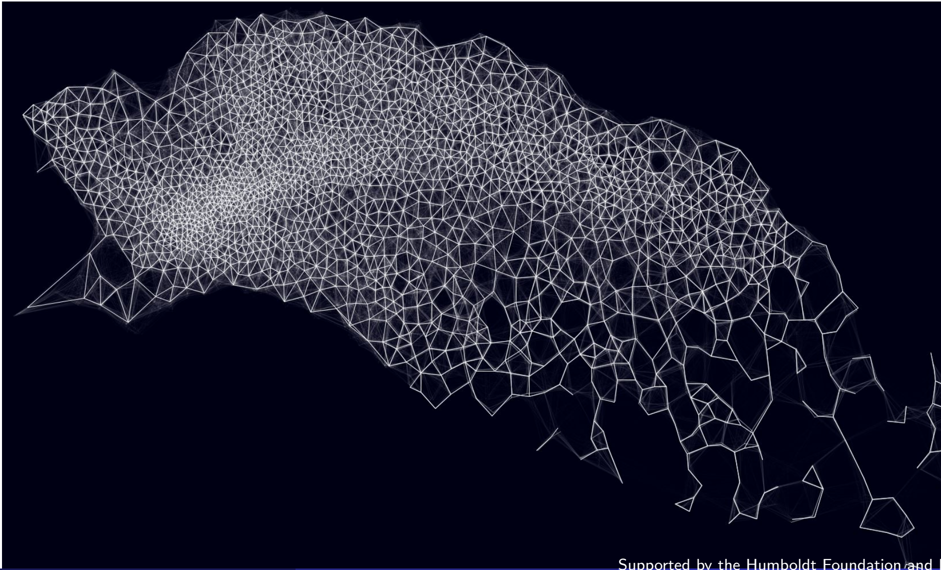
- Remote Sensing Network for Time-Sensitive Detection of Fine Scale Damage to Transportation Infrastructure
- Expoiting existing ground-based remote sensing Networks to improve High-resolution weather forecasts.
- Neural Network Applications in High-Resolution Atmospheric Remote Sensing
- Development of a Real Time Remote Sensing Network to Monitor Small Fluctuations in Urban Air Quality
- Network clustering by graph coloring: An application to **astronomical** images
- Astronomical image segmentation by Networks and wavelets
- Machine learning method to analyse galaxy images

Pictures to:

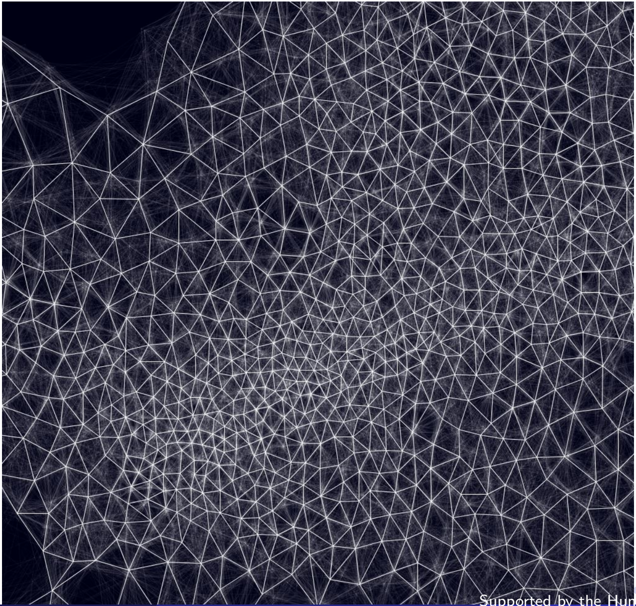
Machine learning: to "teach" a machine to analyse galaxy images

1. Picture 1: Hubble Space Telescope image of the **cluster** of galaxies MACS0416.1-2403, one of the Hubble 'Frontier Fields'. Bright yellow 'elliptical' galaxies can be seen, surrounded by numerous blue **spiral** and amorphous (star-forming) galaxies
2. Picture 2: Visualisation of the **(neural) network** representing the 'brain' of the machine learning algorithm. The intersections of lines are called nodes, and these represent a map of the input data. Nodes that are closer to each other represent similarity
3. Picture 3: A **zoom-in of part** of the network described above. Credit: J. Geach / A. Hocking
4. Picture 4: Image showing the MACS0416.1-2403 **cluster**, highlighting parts of the image that the algorithm has identified as '**star-forming**' galaxies. Credit: NASA / ESA / J. Geach / A. Hocking
5. Picture 5: Image showing the MACS0416.1-2403 **cluster**, highlighting parts of the image that the algorithm has identified as '**elliptical**' galaxies. Credit: NASA / ESA / J. Geach / A. Hocking

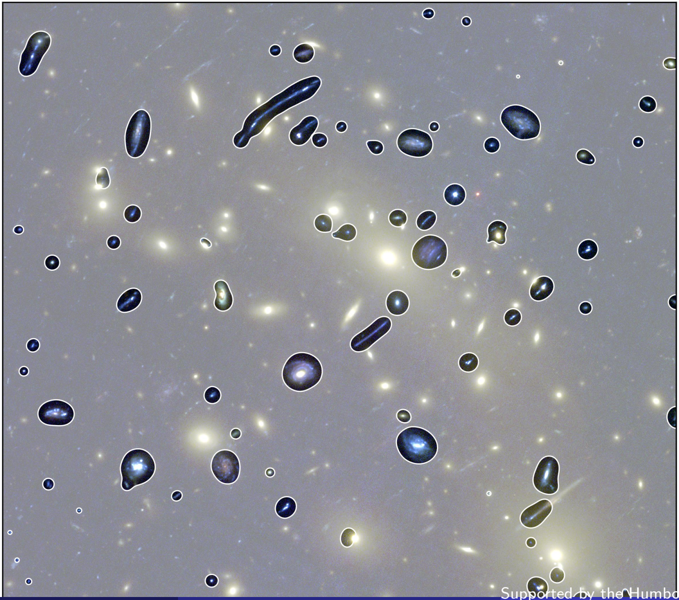
Continued - Network approximation



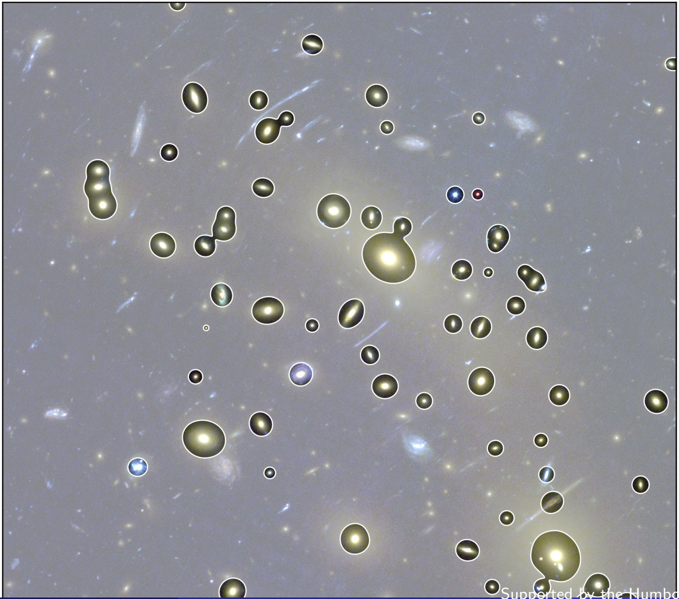
Cont'd - zoomed part



Cont'd - "star forming" galaxies



Cont'd – "elliptical" galaxies



Deep Learning and Wavelets approach

References and Credits

- Hammond, Vandergheynst, R. Gribonval, *Wavelets on graphs via spectral graph theory*, 2009 in arxiv; 2011 in IEEE.
- R. Lambiotte, J.-C. Delvenne and M. Barahona, *Laplacian Dynamics and Multiscale Modular Structure in Networks*, 2009, arxiv.
- R. Lambiotte, J.-C. Delvenne and M. Barahona, Random Walks, *Markov Processes and the Multiscale Modular Organization of Complex Networks*, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, VOL. 1, NO. 2, JULY-DECEMBER, 2014
- Leonardi, N., D. Van De Ville, *Tight wavelet frames on multislice graphs*, IEEE, 2013
- M. Mitrovic and B. Tadic, "*Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneities*," Phys. Rev. E, vol. 80, p. 026123, Aug 2009.
- Raif M. Rustamov & Leonidas Guibas, *Wavelets on Graphs via Deep Learning*, 2013

Credits

- A. Sandryhaila and J. Moura, *Big data analysis with signal processing on graphs*, IEEE Signal processing Magazine, September, 2014.
- D. Shuman, Ch. Wiesmeyr, *N*. Hoighaus, P. Vandergheynst, *Spectrum-adapted tight graph wavelet and vertex-frequency frames*, IEEE, 2015.
- D. Shuman, Pierre Vandergheynst, and Pascal Frossard, *Chebyshev Polynomial Approximation for Distributed Signal Processing*, IEEE, 2011

Credits

- David I Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst, *The Emerging Field of Signal Processing on Graphs*, IEEE SIGNAL PROCESSING MAGAZINE, May 2013.
- Tremblay, N., P. Borgnat, *Graph wavelets for multiscale community mining*, IEEE, 2014.
- Tremblay, N., P. Borgnat, *Multiscale community mining in networks using spectral graph wavelets*, 2013.
- R. Carmona, W.L. Hwang, and B. Torresani, *Practical Time-frequency analysis*, Academic press, 1998.