Graphs for data science and ML

Machine Learning for graphs and with graphs

P. Borgnat, CNRS, LP ENSL (1)

Data Science & Machine Learning

- I came to ML from Data Science <- Data Processing in fact

The question for this lecture: how to mimic Signal Processing for data on graphs ?

Hence, in 1 slide: **what is Signal processing ?**

$\boxed{\text{Information}} \xrightarrow{\text{Coding}} \boxed{\text{Data}} \xrightarrow{\text{Measuring}} \boxed{\text{Observations}}$

Noise/Missing

e.g.:

x Climate change?	Temperature (t)	Samples at specific times/places
x Gene Expression profiling?	DNA microarray	Images of spots.
x Music (or Speech)	Sound	Noisy sound, sampled, ...

Need to define what is information, how it affects the data, what is interesting, ...

Need a model of measurement and observation

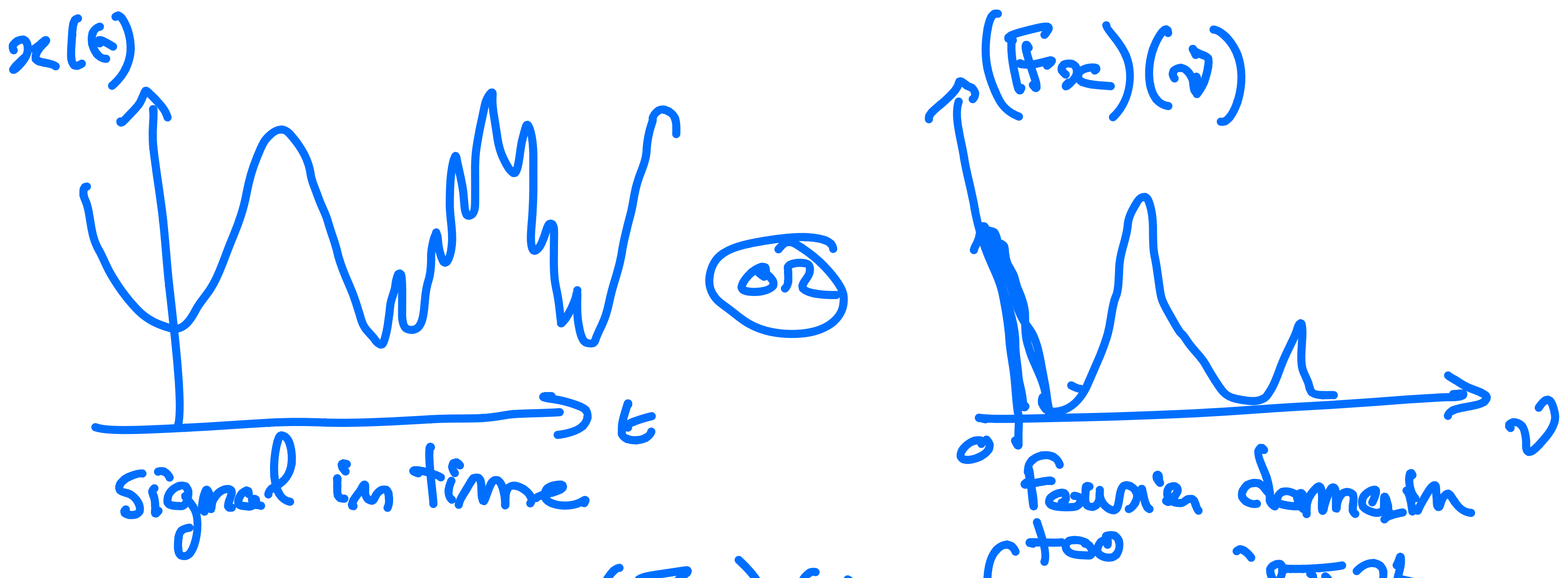
The way back = Signal Processing!

Data Science & Machine Learning

Key lessons from Signal processing:

- **Representation of data is important**

$x(t)$ or $(\mathcal{F}x)(\nu)$ or $x(t,\nu)$ or ...



$$(\mathcal{F}x)(\nu) = \int_{-\infty}^{+\infty} x(t)\, e^{-i2\pi\nu t}\, dt$$

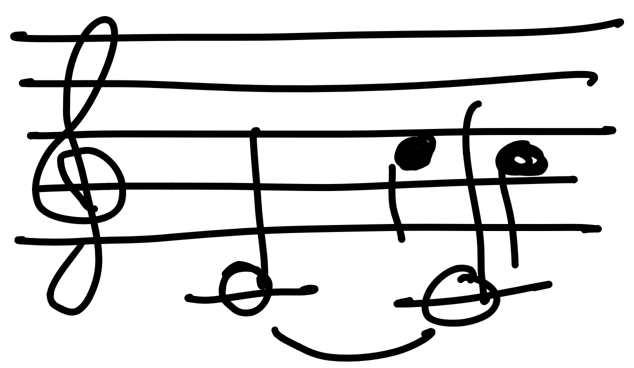
- **Know how to write observation models**

Observation signal

$$o(t) = x(t-\tau) + n(t)$$

delay noise

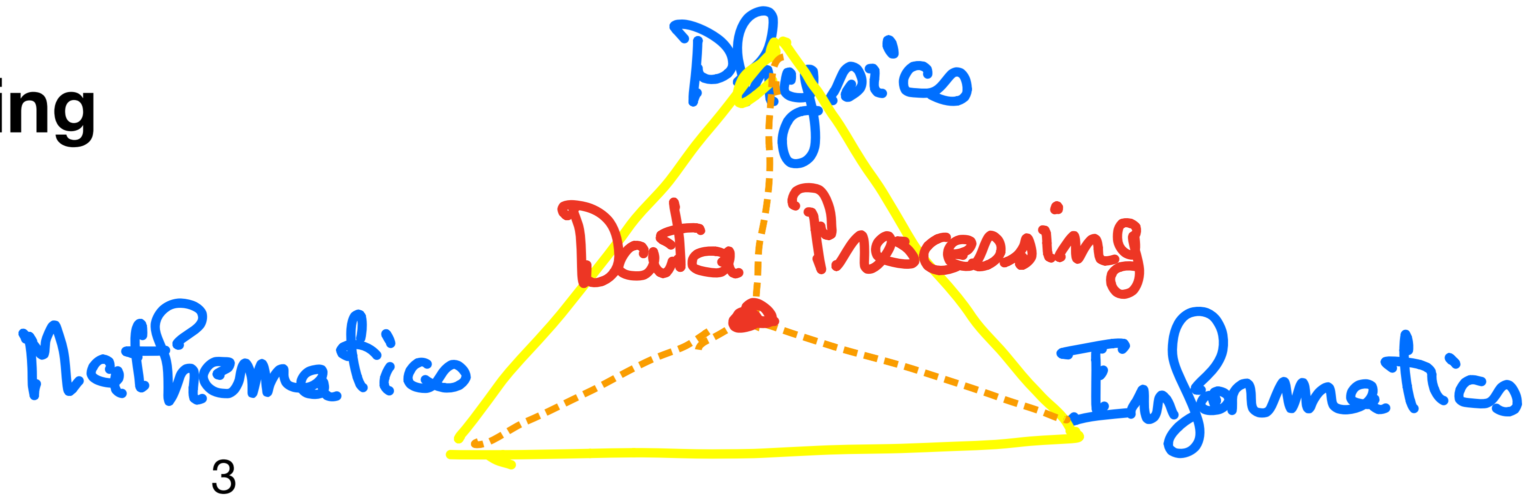
or



?

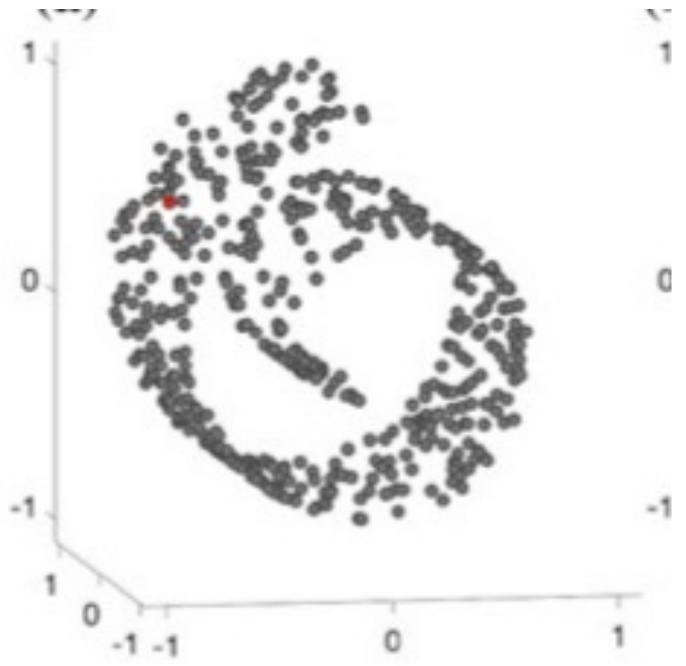
- **Two types of tools are required:**
  - **Exploratory data analysis (know how to better display information)**
  - **Exact tools for inference (know to best extract information, with statistical confidence)**

. The golden triangle of Signal processing



Why dealing with graphs ?

data on nonlinear manifold

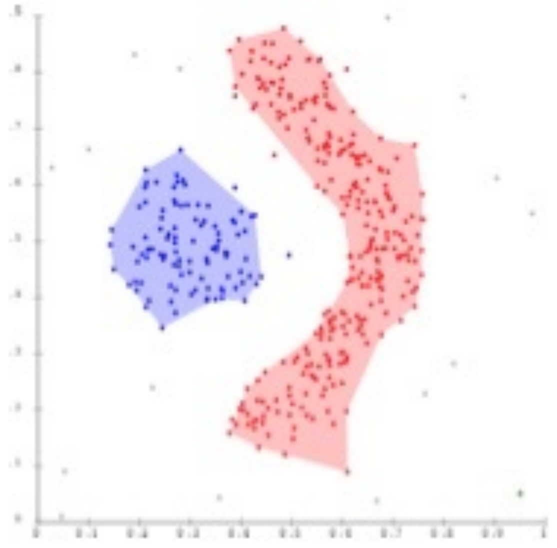


- Data is generally in non-linear spaces

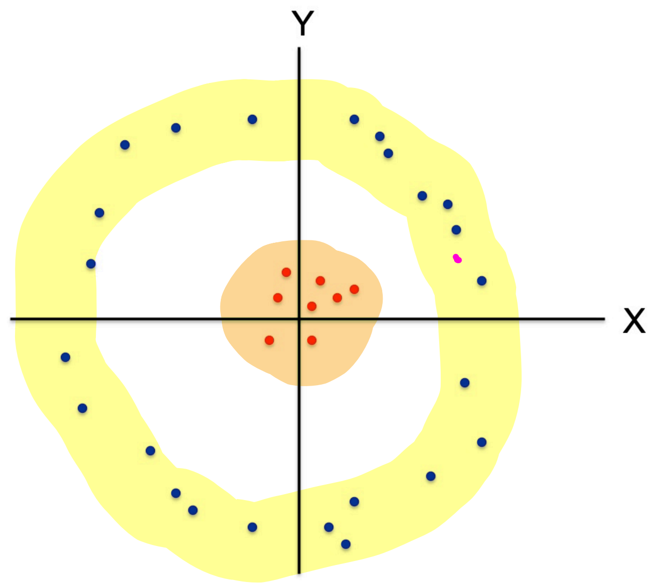
- Some data are first and foremost relational

data on graph

- Often, proximity is important



because proximity is important



because space is sometime better represented by close proximity only

Why dealing with graphs ?

Introduction: on signals and graphs

Why data analysis and processing is useful for networks?

- Many examples of data having both:
  labels and/or attributes (a.k.a. “signals”)
  and structures or relational properties (graphs)

- Many data sets in high dimension, or large dataset, are best encoded with graphs

- Non-trivial estimation issues (e.g., non repeated measures; variables with large distributions (or power-laws); ...)

  $\rightarrow$ **advanced statistical approaches**

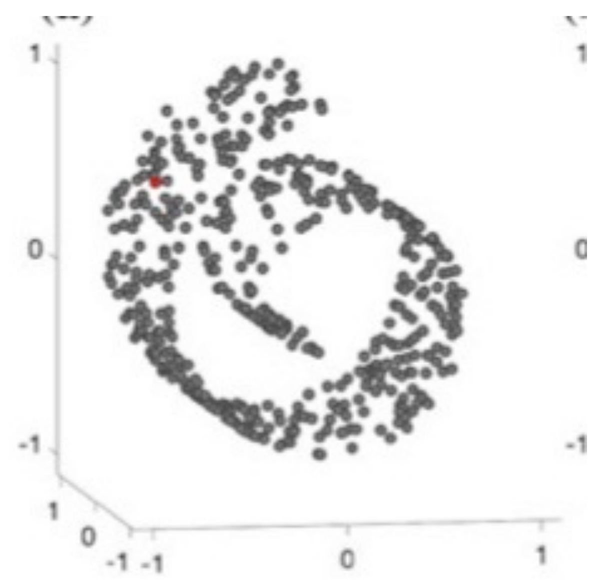
- large networks

  $\rightarrow$ **multiscale approaches**

- dynamical networks

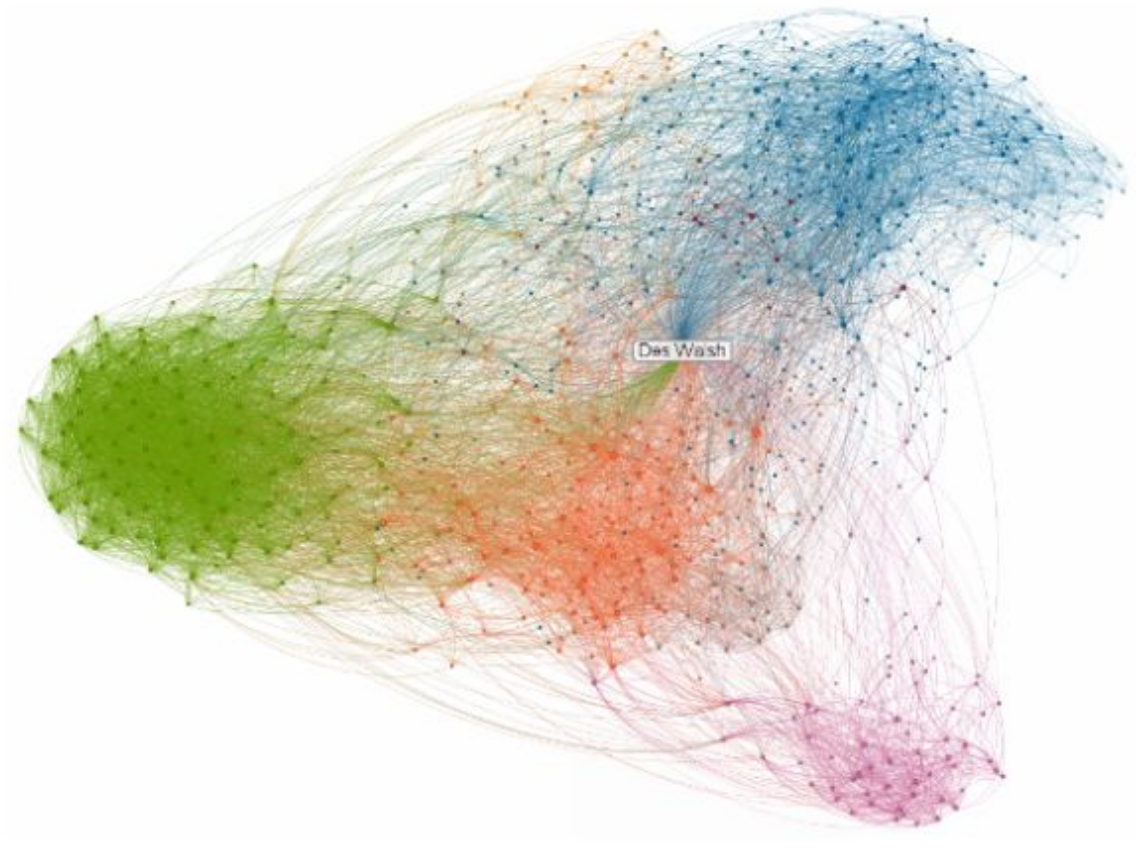
  $\rightarrow$ **nonstationary methods**

data on nonlinear manifold

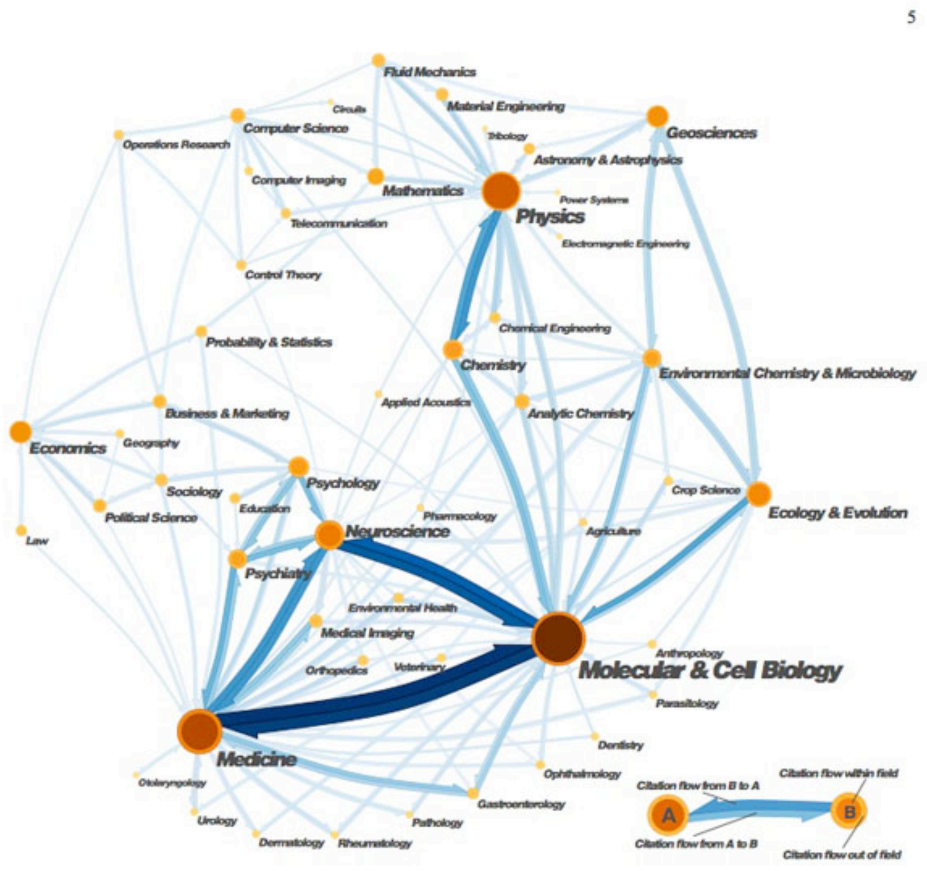


data on graph

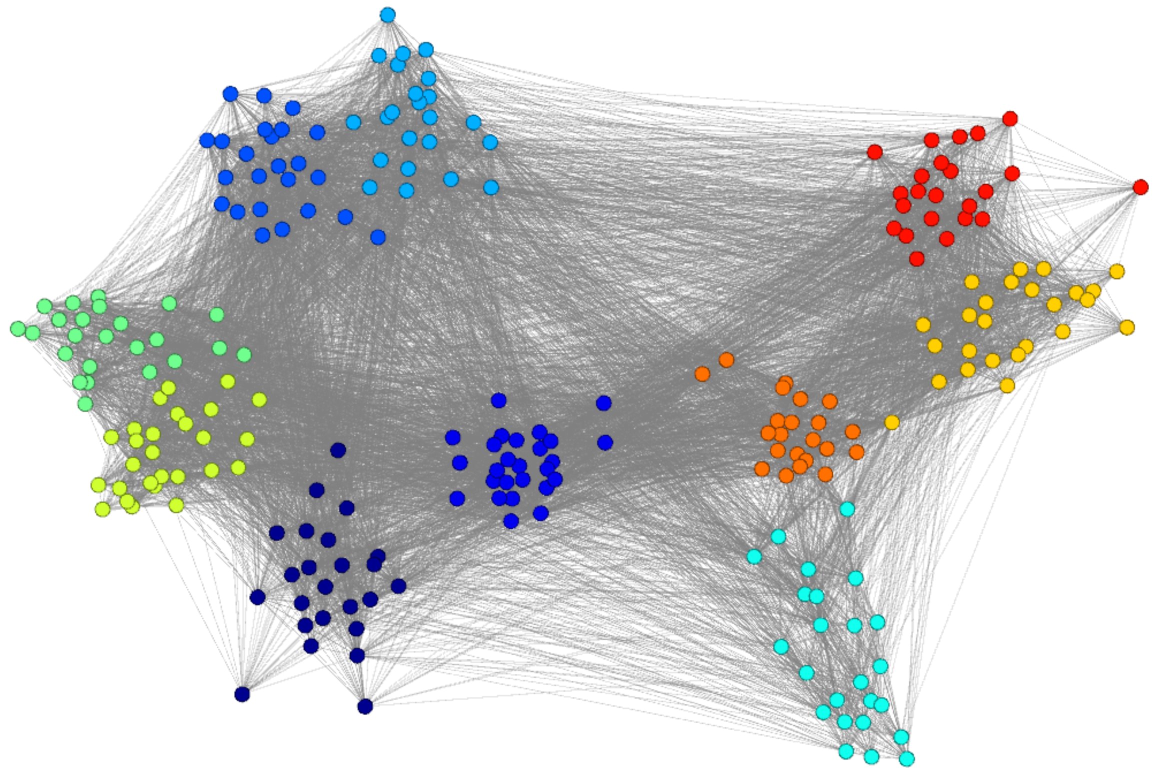
Examples of networks from our digital world



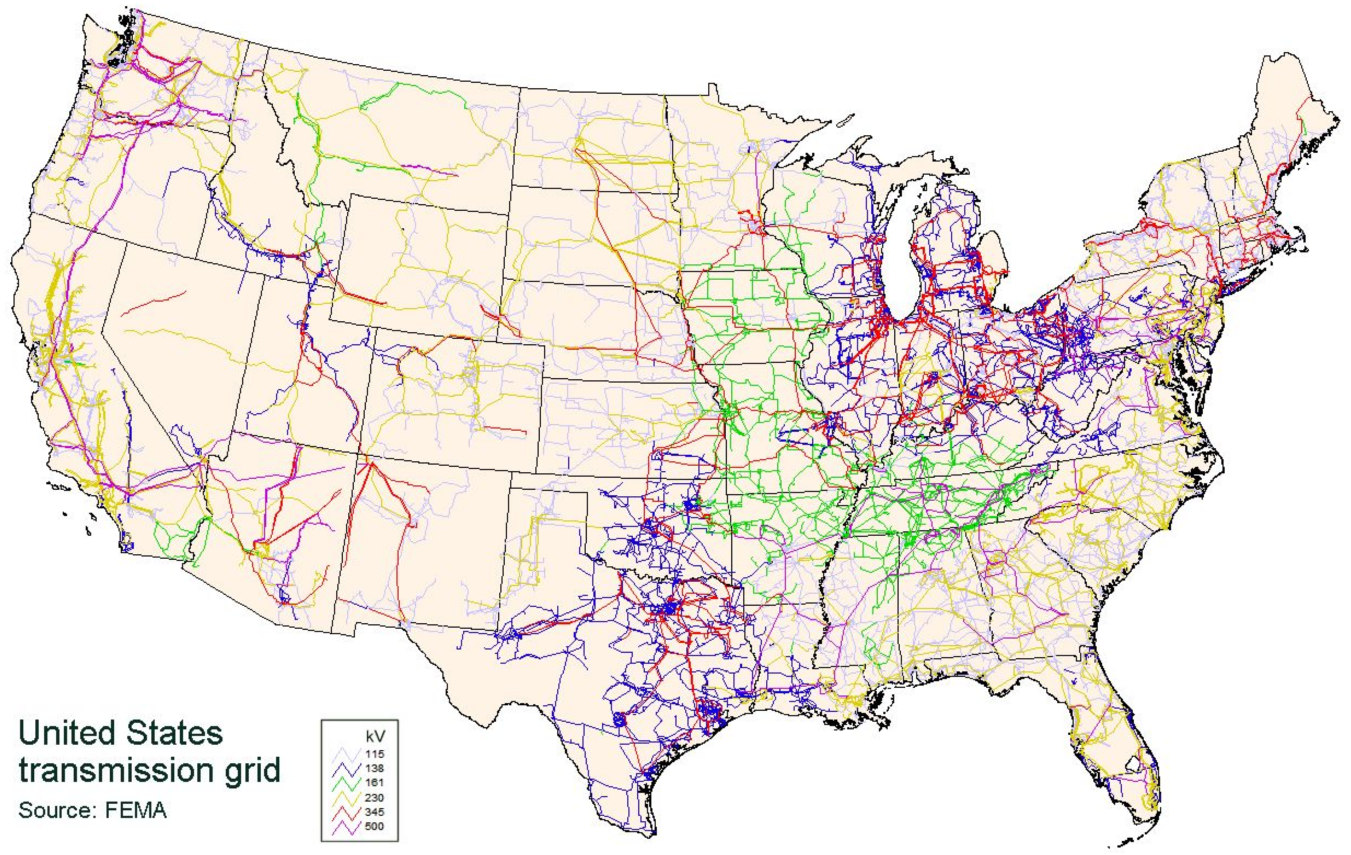
LinkedIn Network



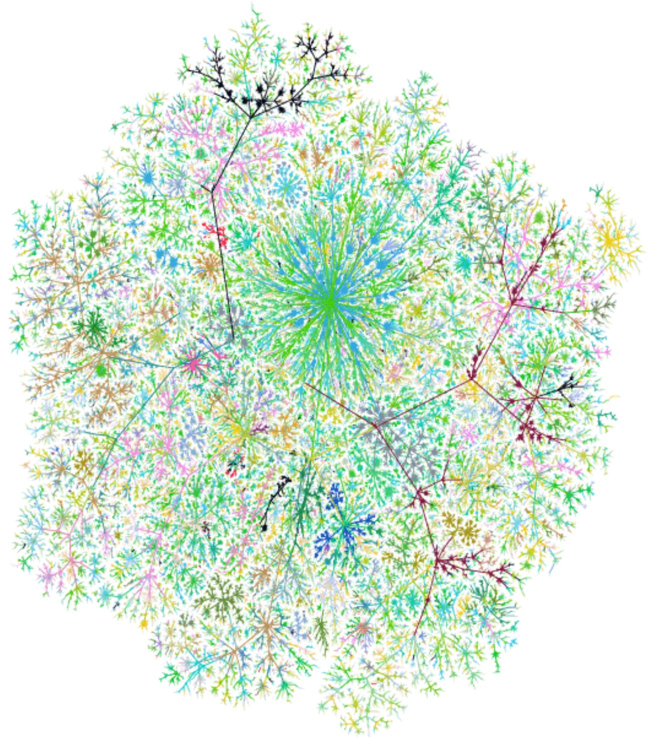
Citation Graph



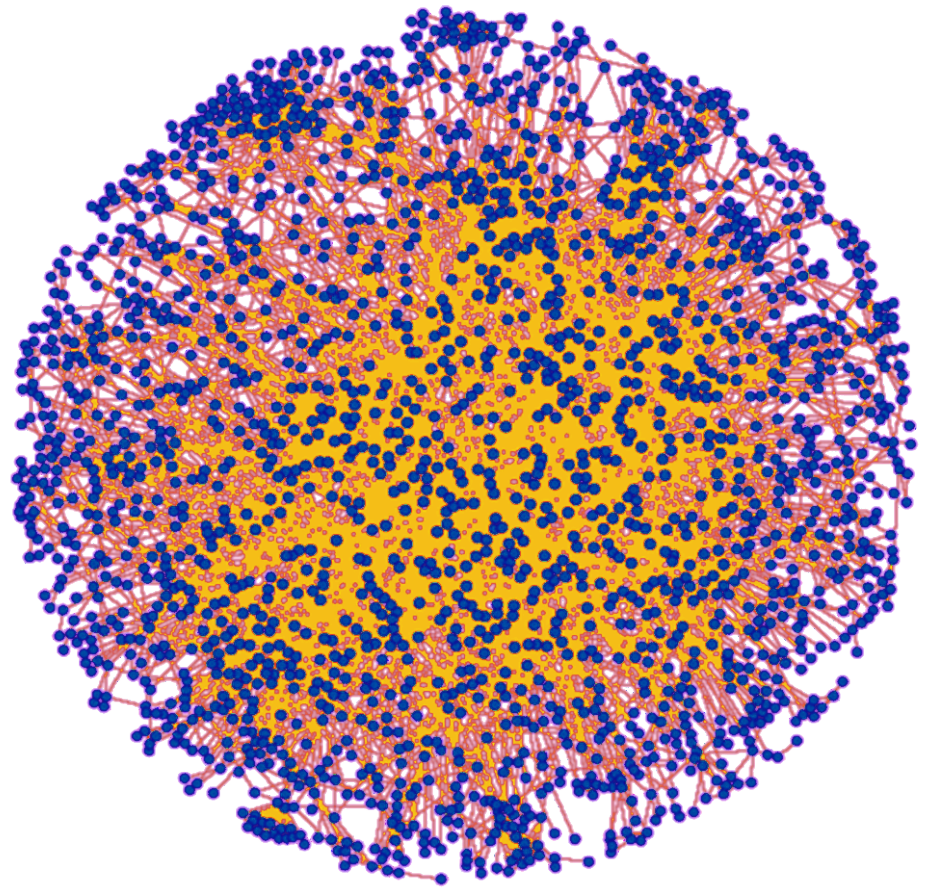
Sociopatterns graphs



USA Power grid



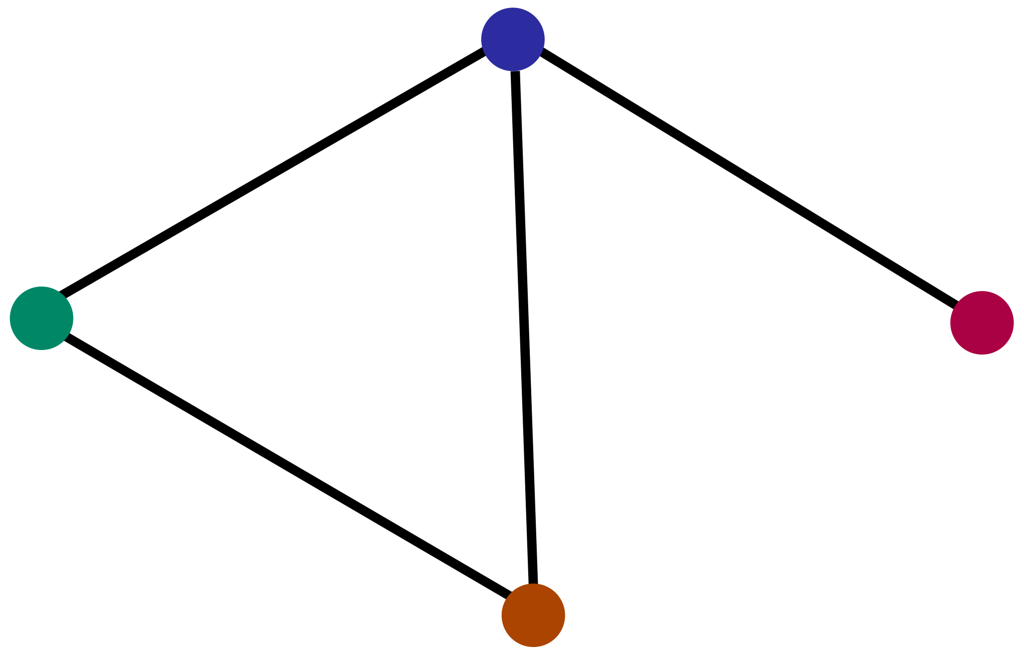
Web Graph



Protein Network

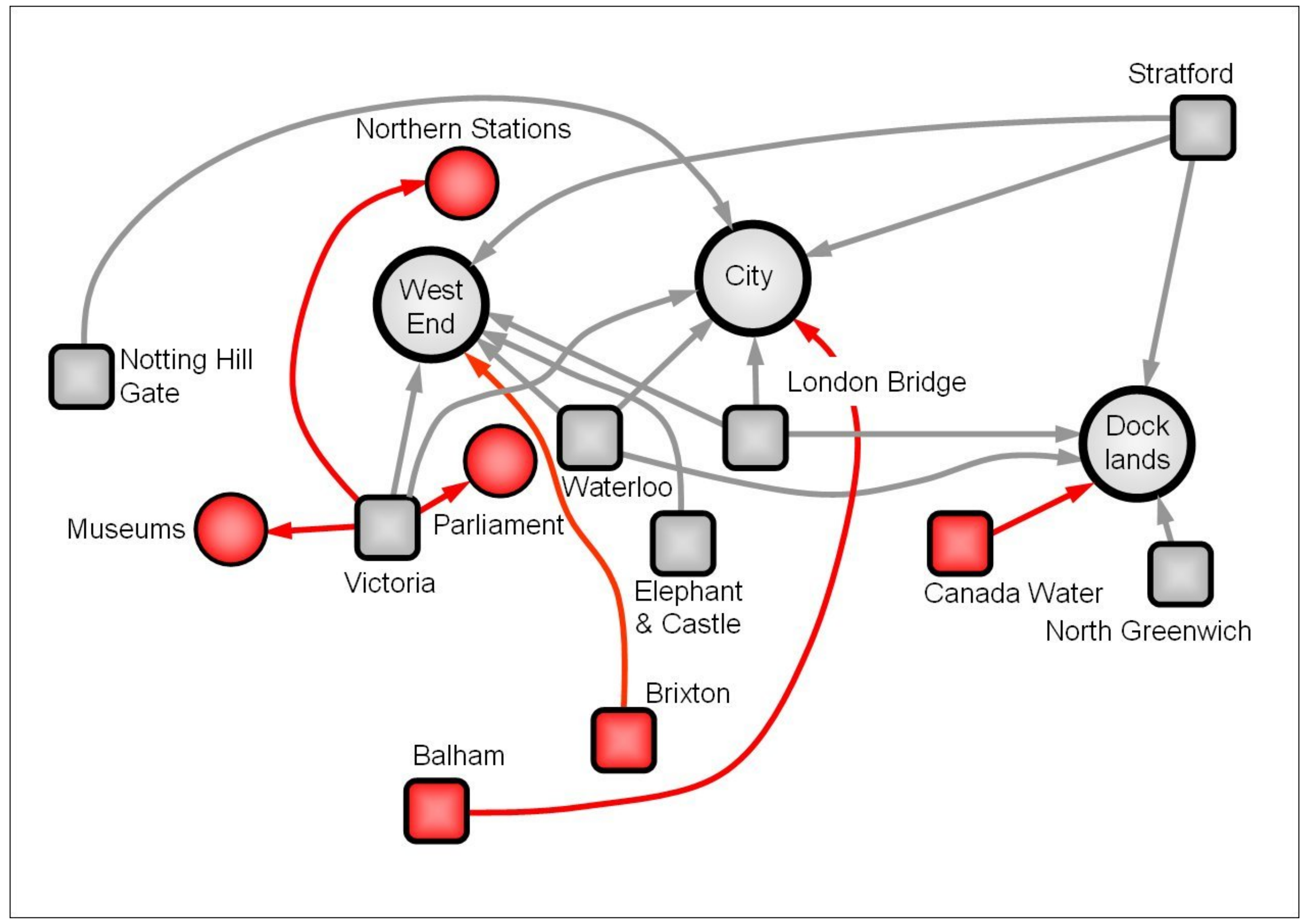
Data as graphs

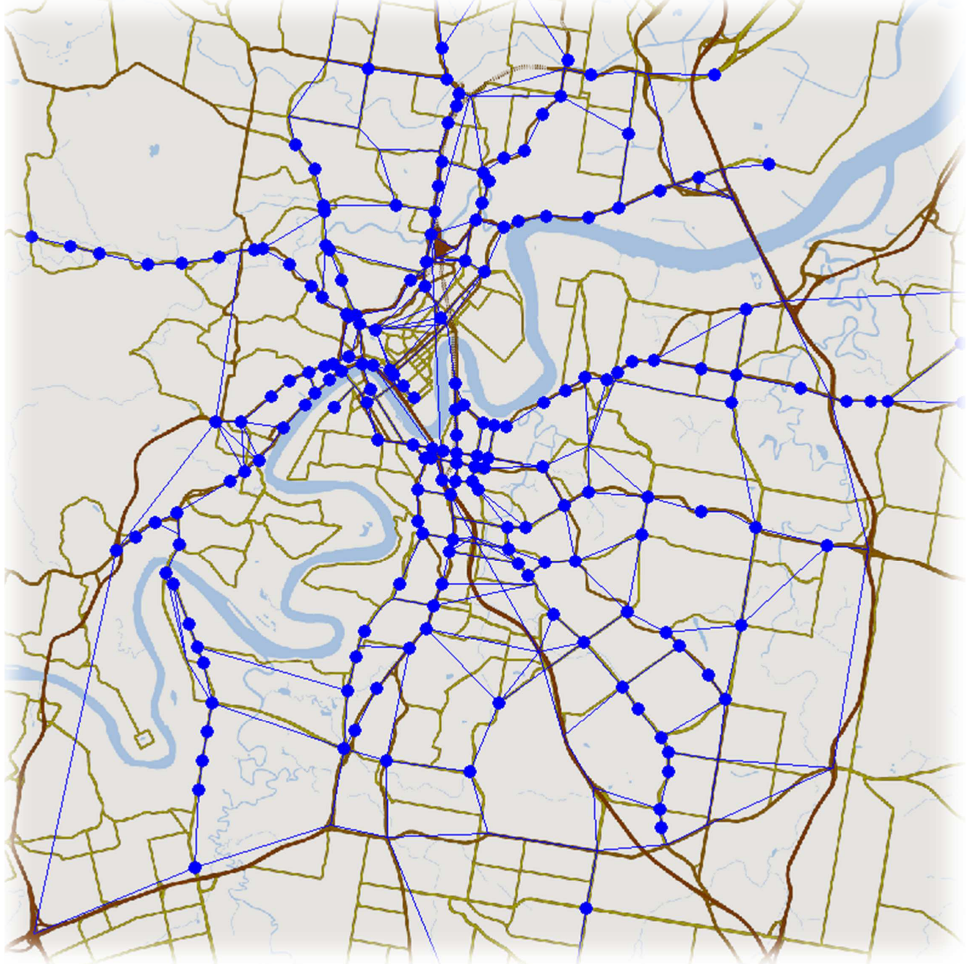
- A graph $G = (V, E)$, set of nodes in V and edges in E



$$V = \{blue, green, orange, red\} \text{ and}$$
$$E = \{(b, g), (g, o), (o, b), (b, r)\}$$

- Good to represent relations ($\in E$) between entities ($\in V$)

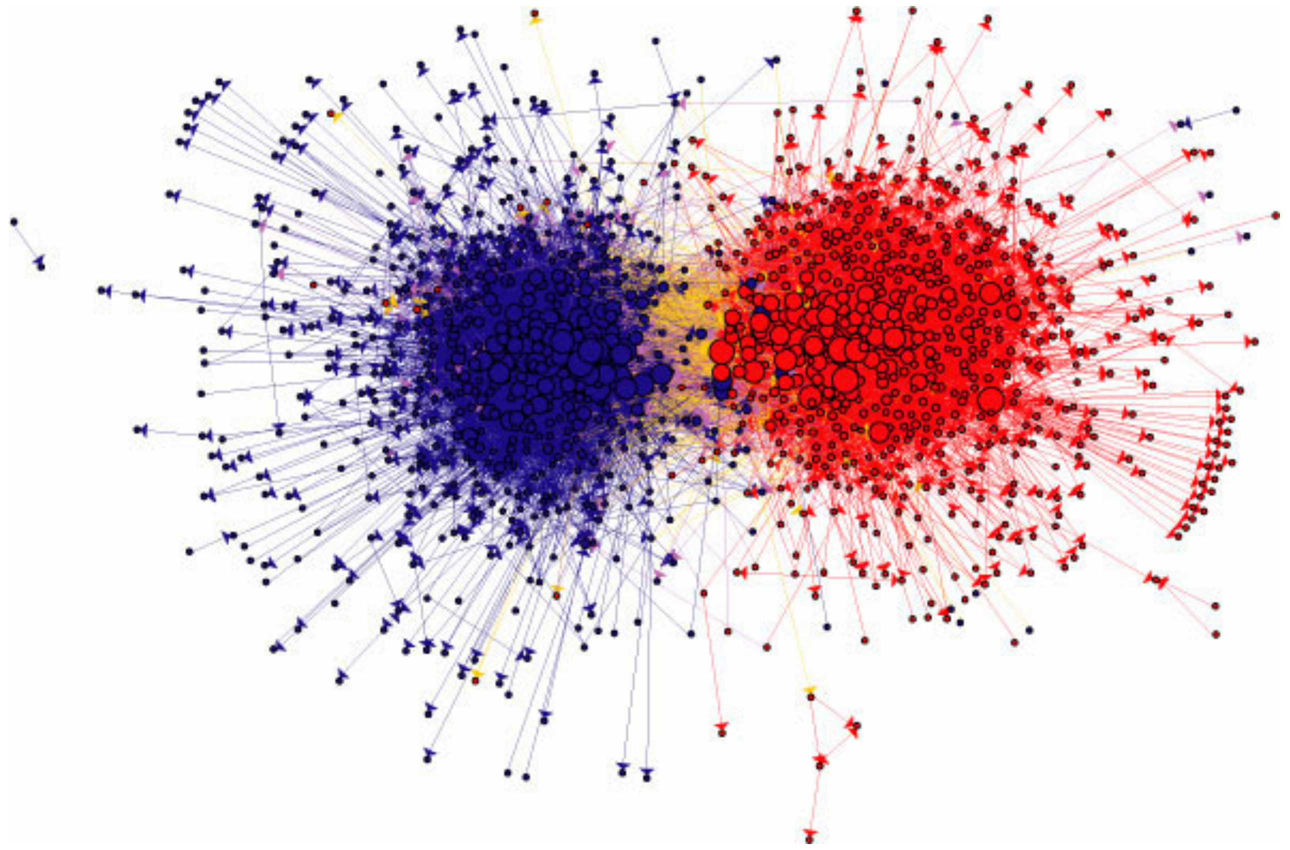




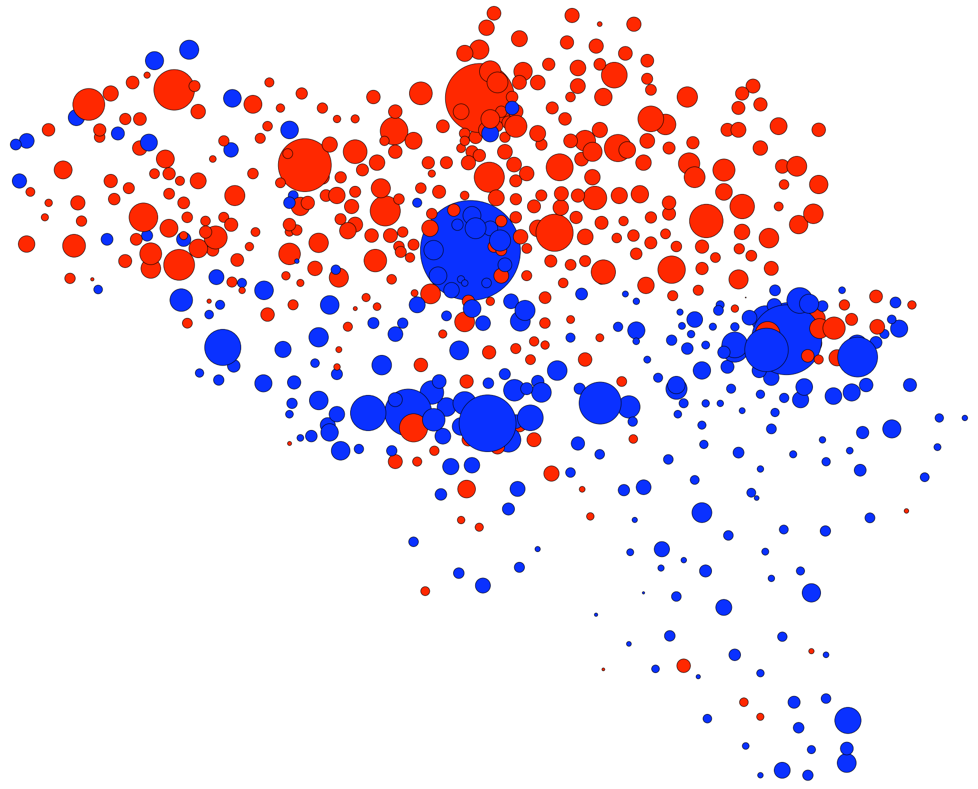
[Roth et al., 2011] [Michau et al., 2017]

Data as graphs: many uses

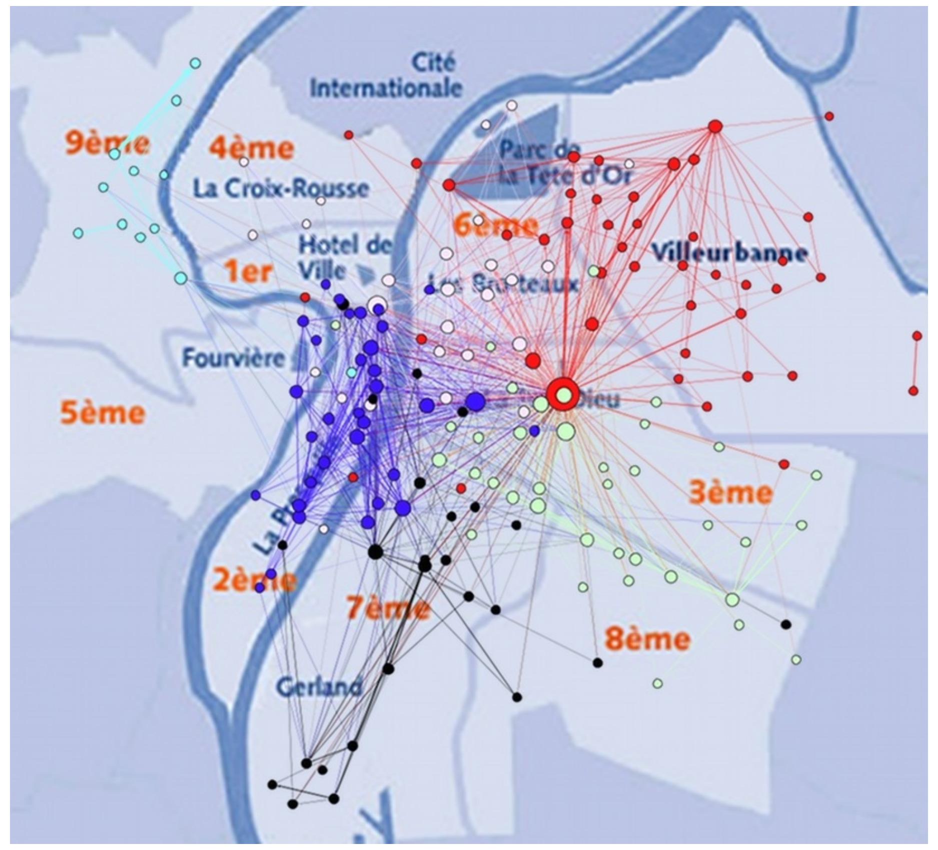
- Good to detect groups in the data ($\simeq$ clustering)



Blogosphere US 2004
[Adamic et al. 2005]

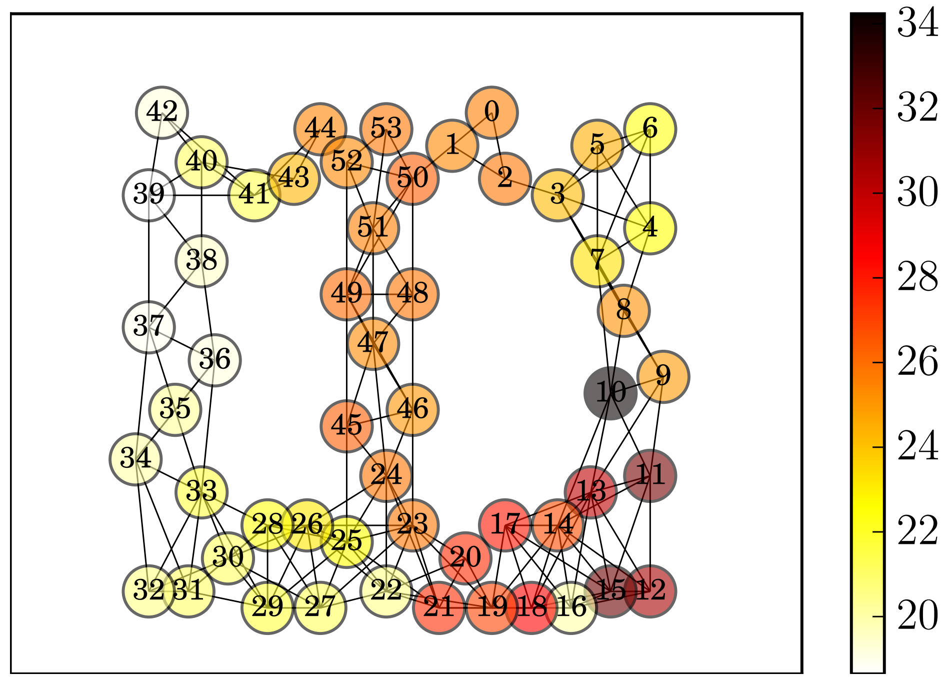


Mobile phones
[Blondel et al., 2008]

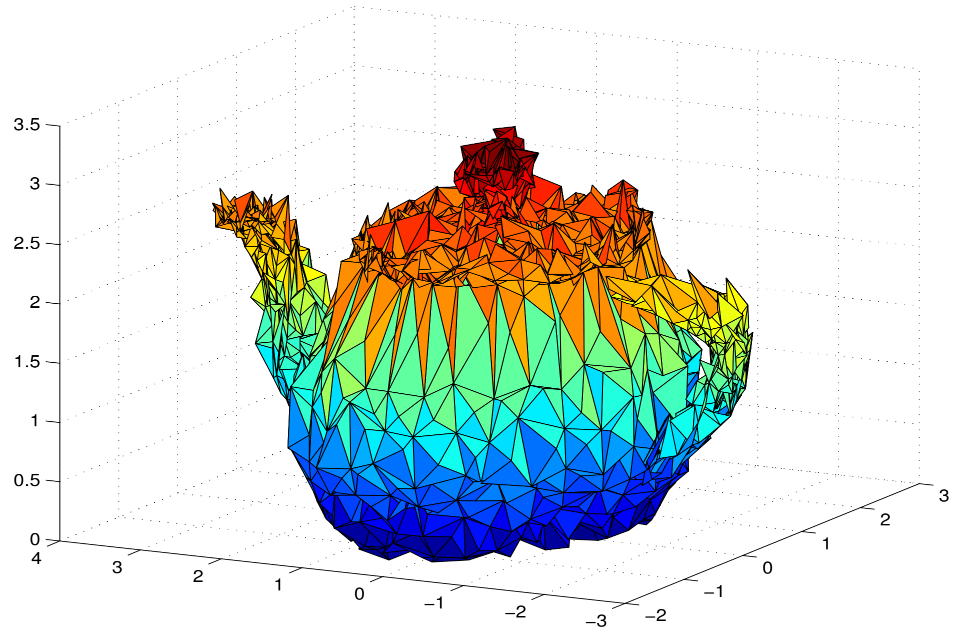


BSS Vélo'v in Lyon
[Borgnat et al., 2013]

- Good to code irregular shapes

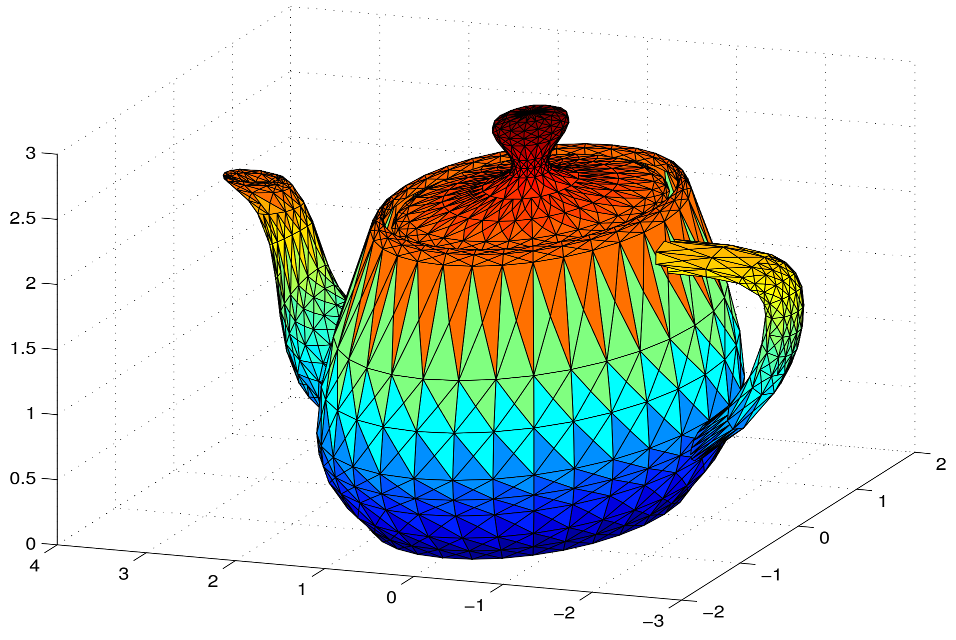


[R. Hamon et al., 2016]



Noisy mesh y

$\overset{?}{\Rightarrow}$

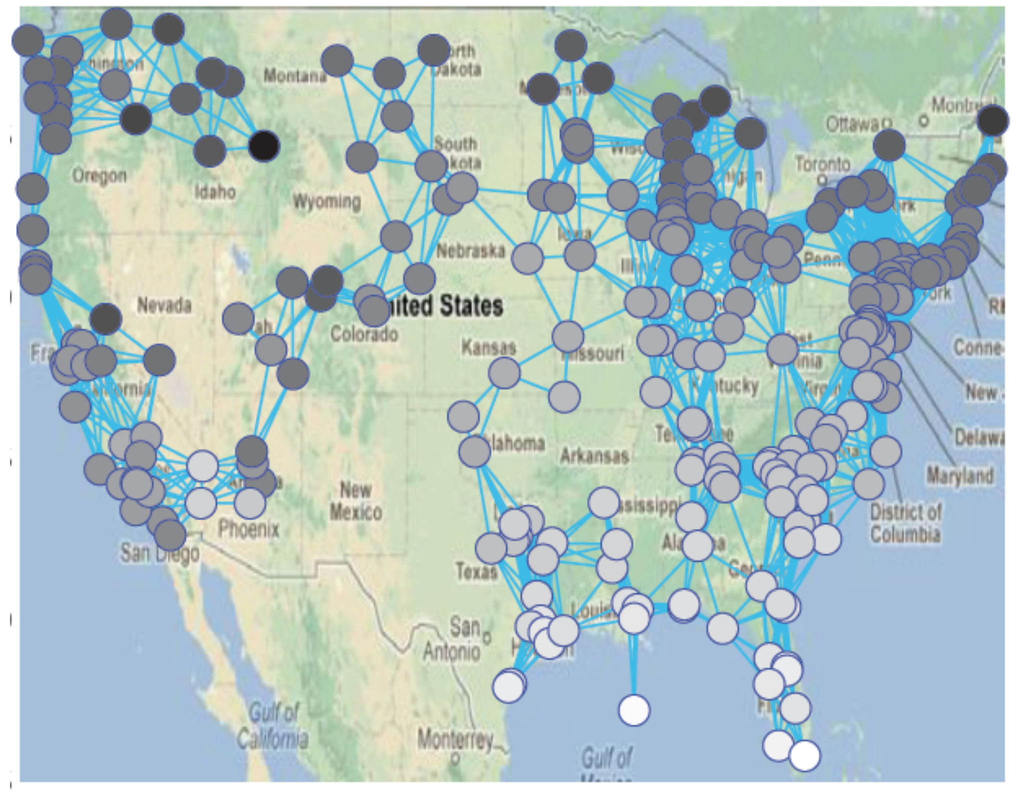


Original mesh $\overline{x}$

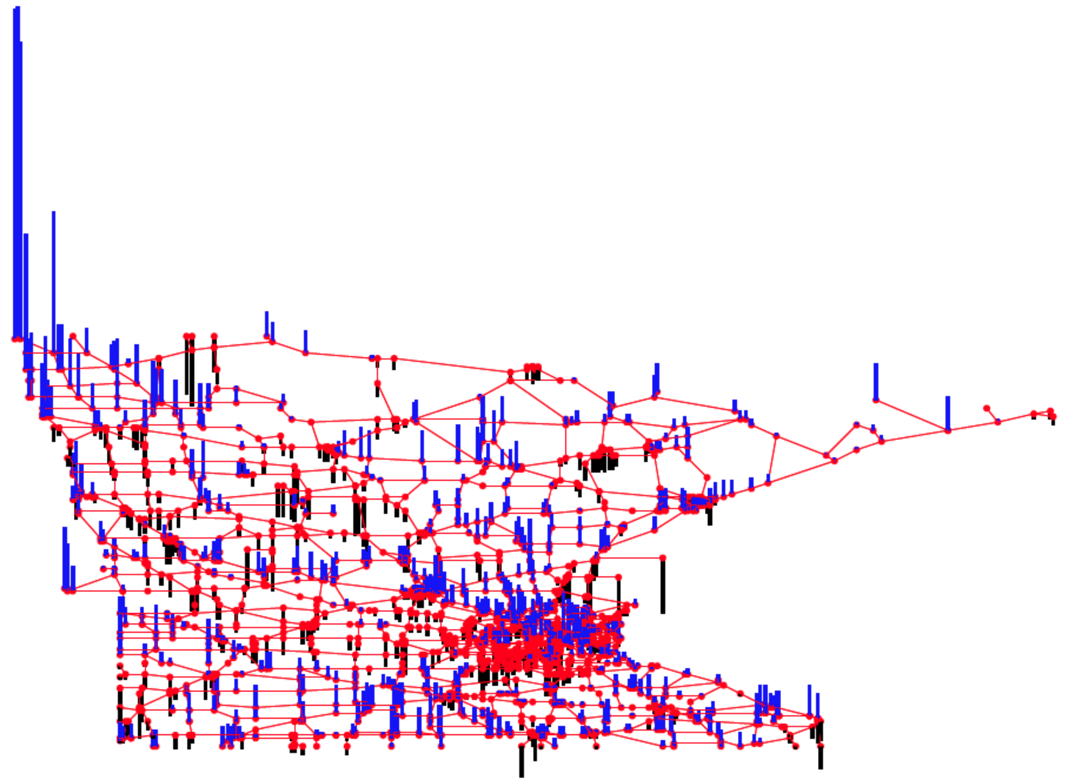
[Cours, N. Pustelnik & P.B., ENSL]

Examples of graph signals

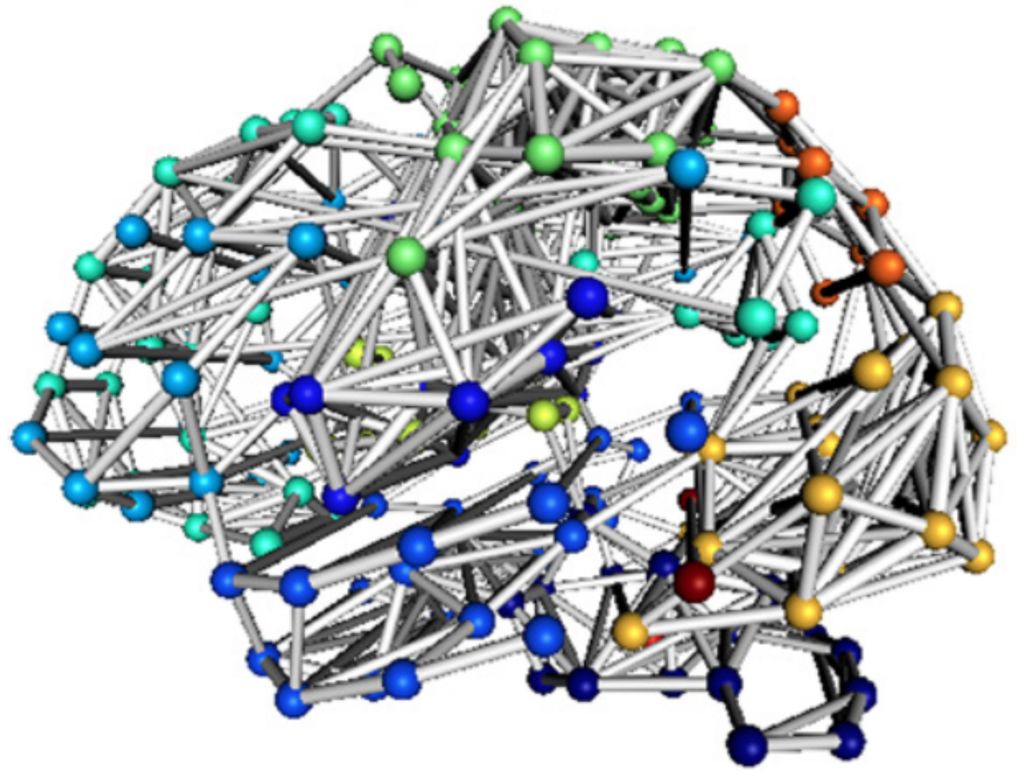
- Given a graph G, let's consider a signal x on the nodes V. If $N = |V|$, we have $x \in \mathbb{R}^N$ (could be in $\mathbb{C}^N$ or multivariate)



USA Temperature



Minnesota Roads



fMRI Brain Network

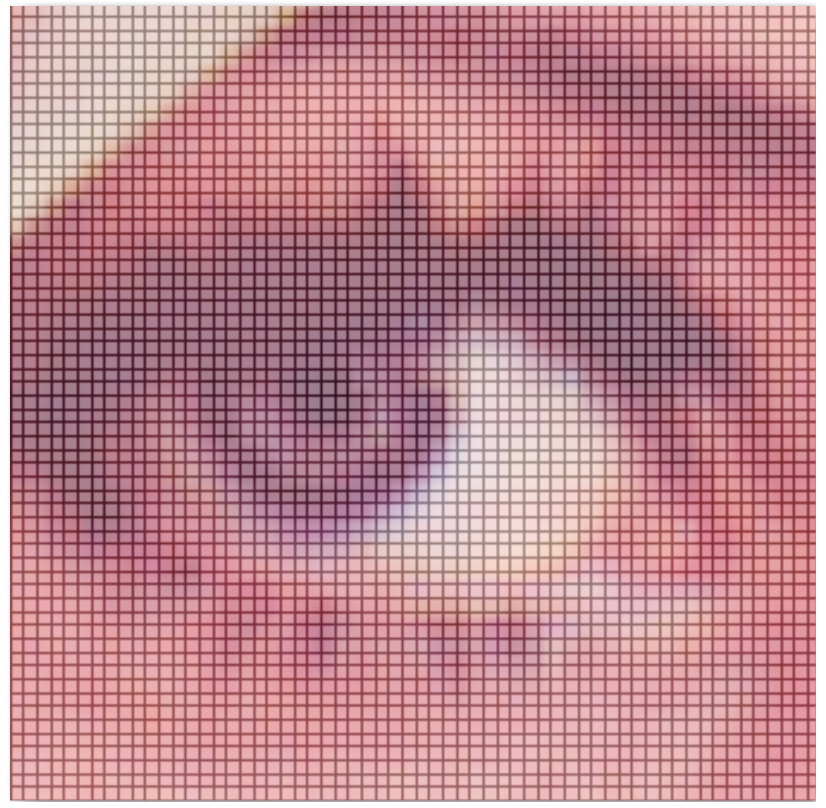


Image Grid

Color Point Cloud

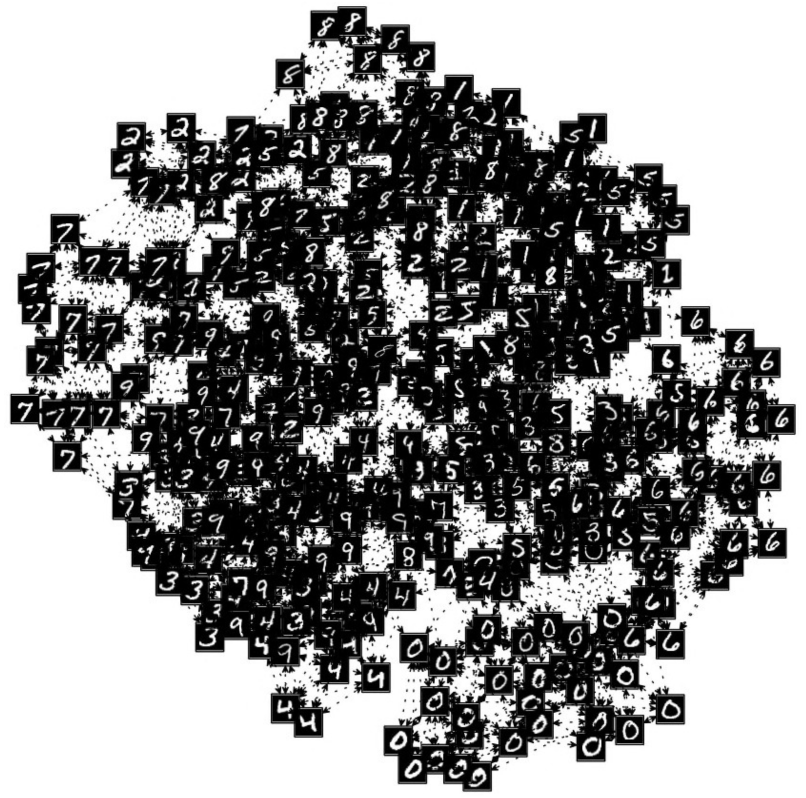
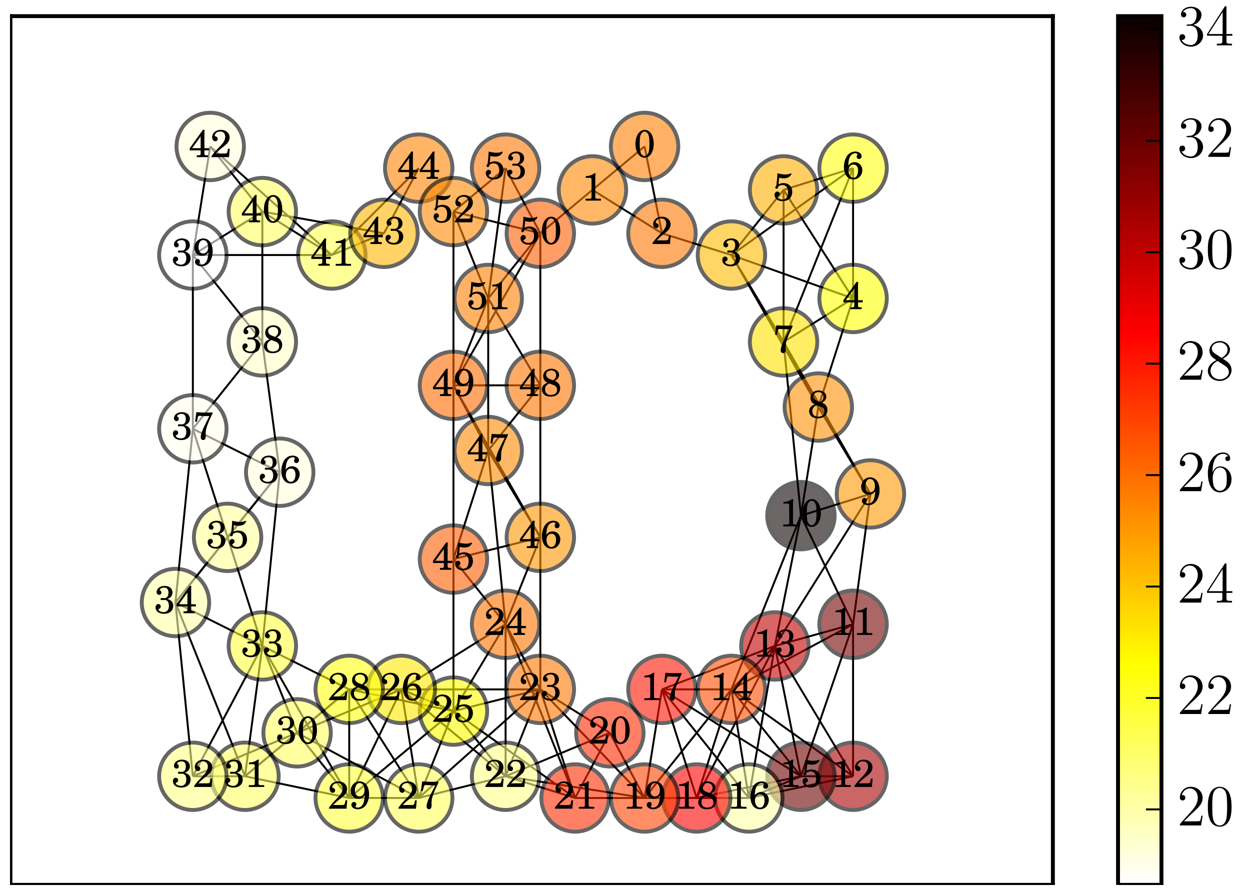
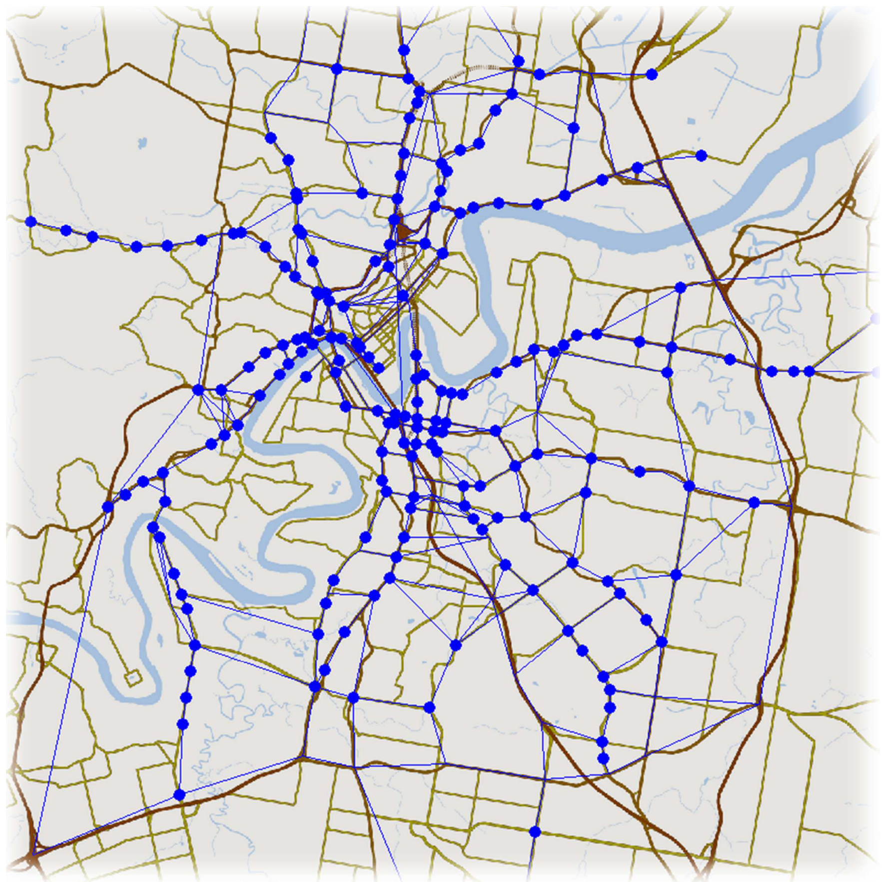


Image Database

Typical problems for graph signal processing

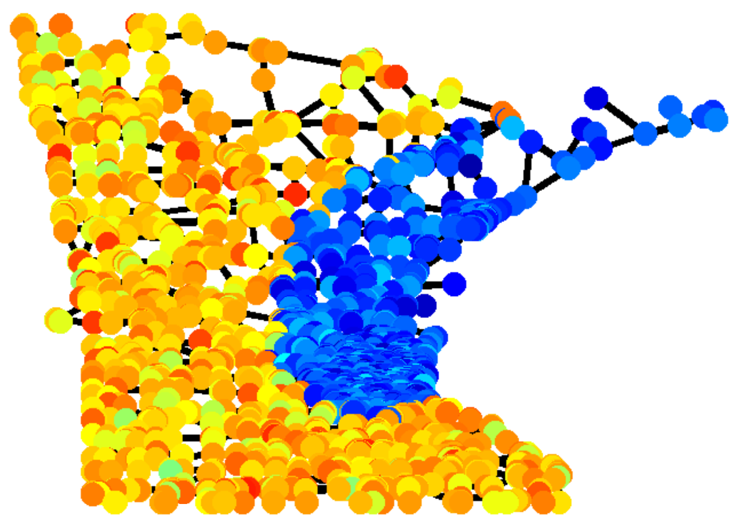
- Often, the graph is not a regular (yet it could be)

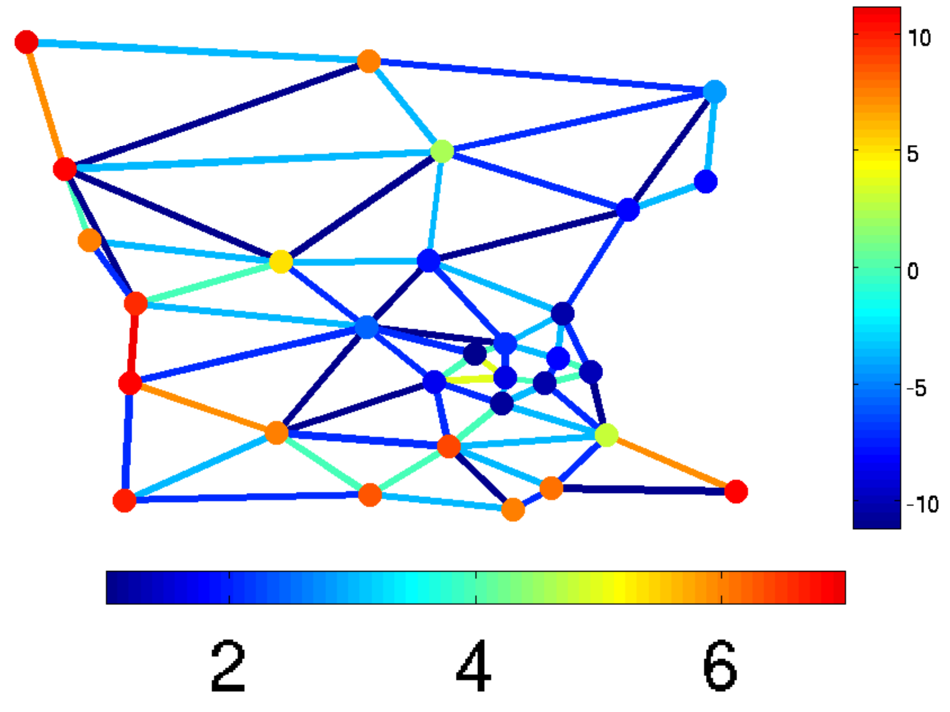


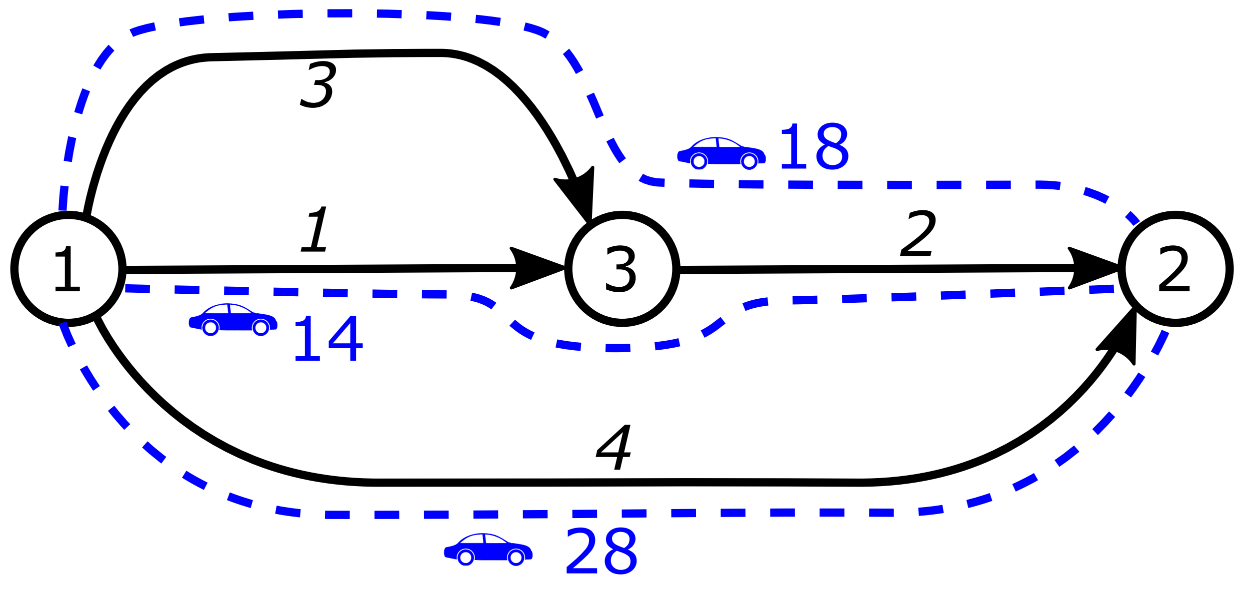


- How to answer typical signal/image processing questions?

Denoising? Compression + Coarsening ? Estimation ?



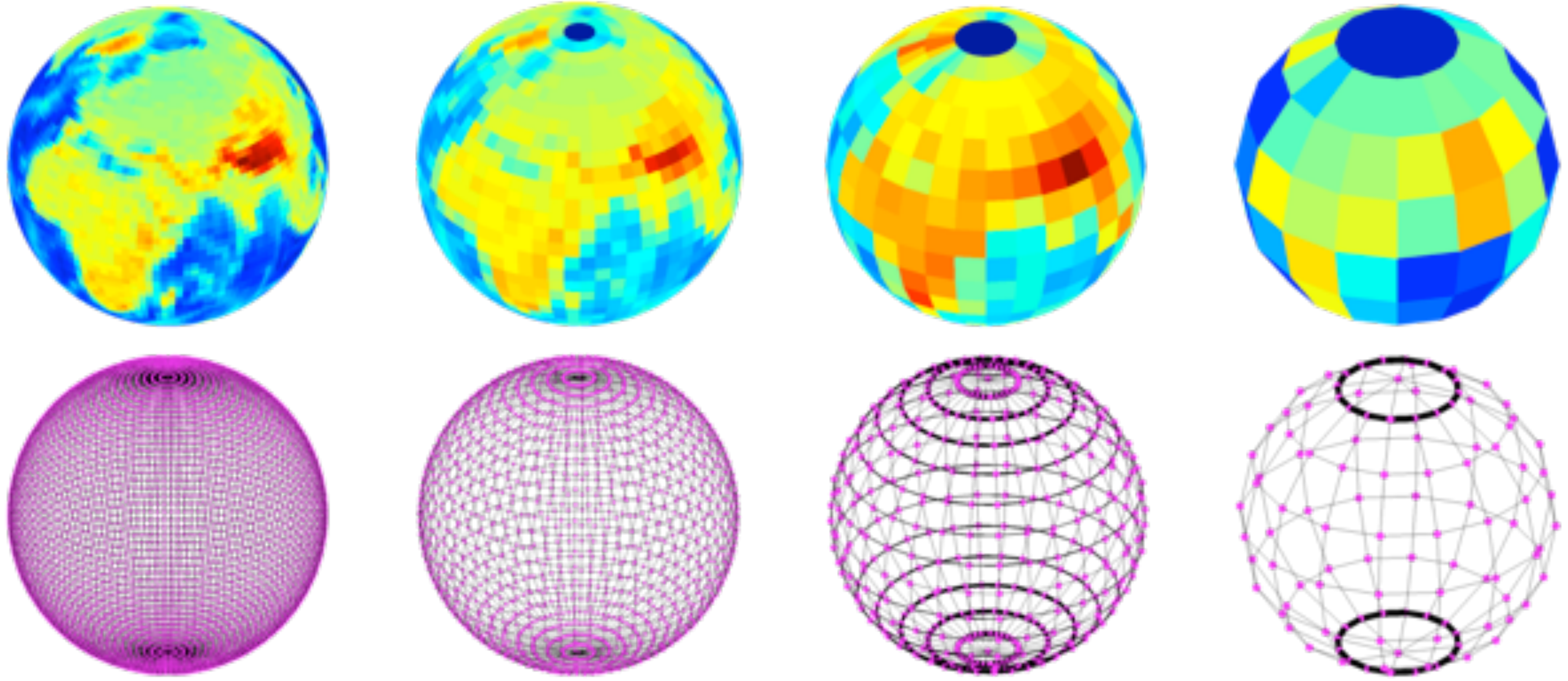




Typical problems

[P. Vandergheynst, EPFL, 2013]

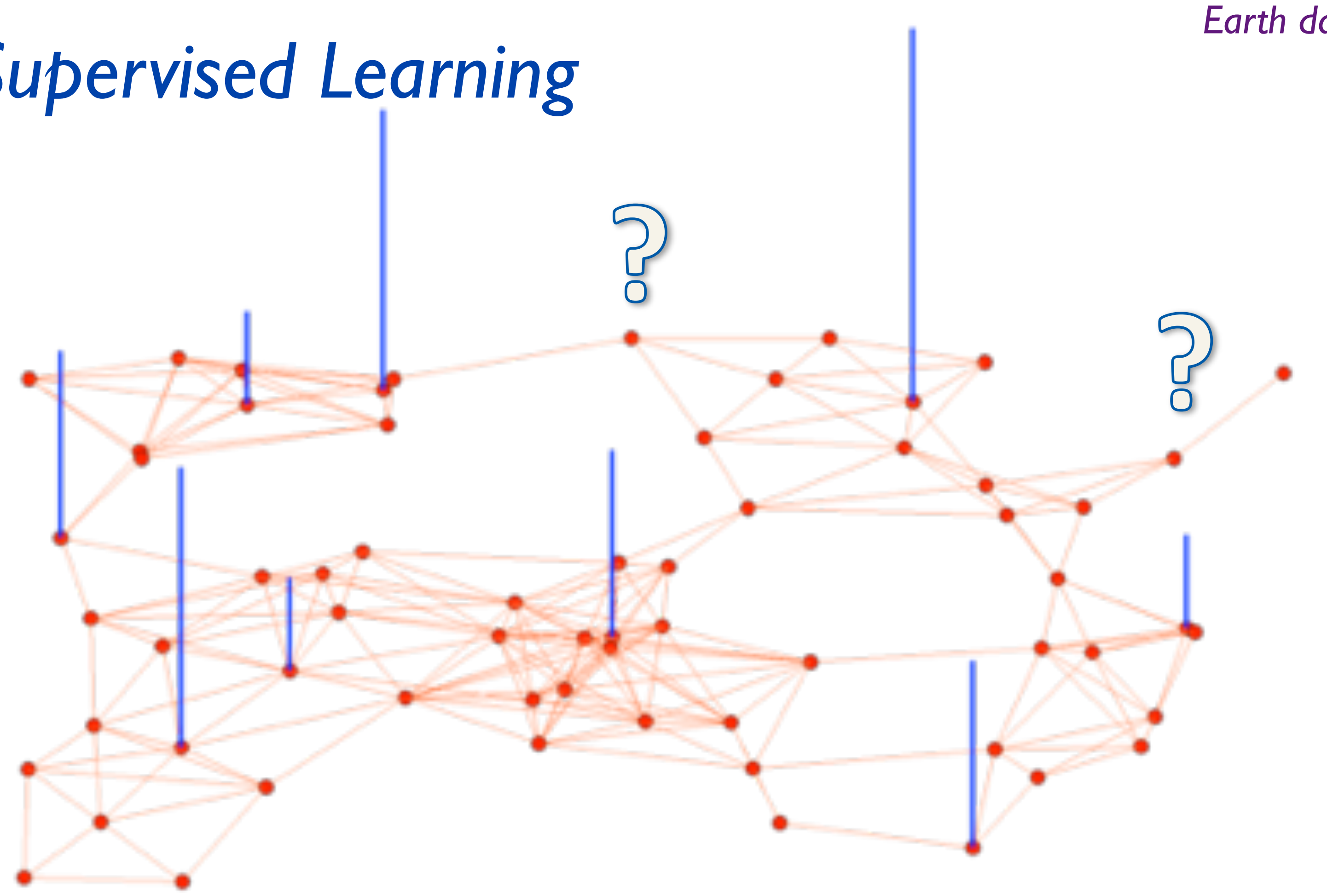
Compression / Visualization

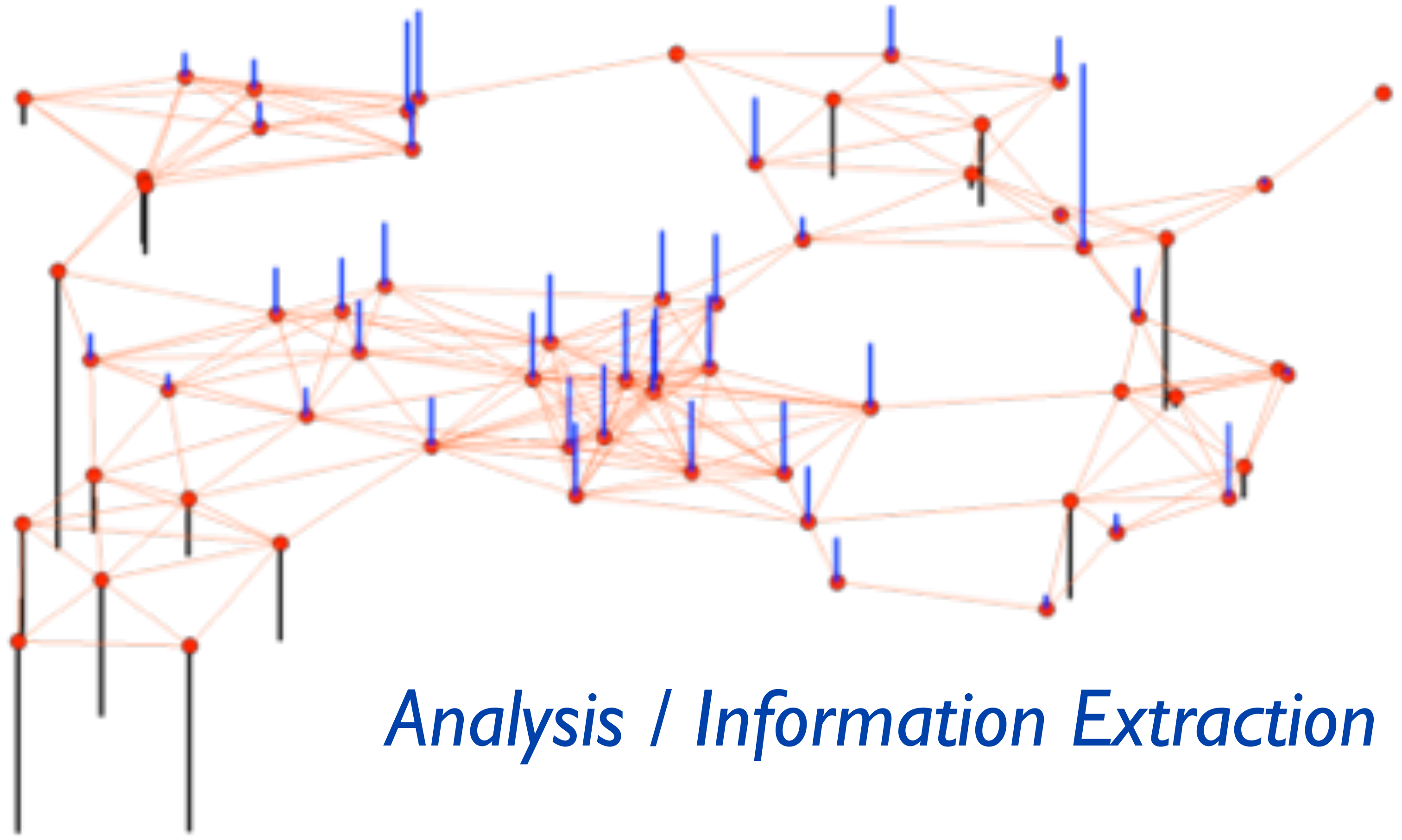


Earth data source: Frederik Simons

Denoising

Semi-Supervised Learning

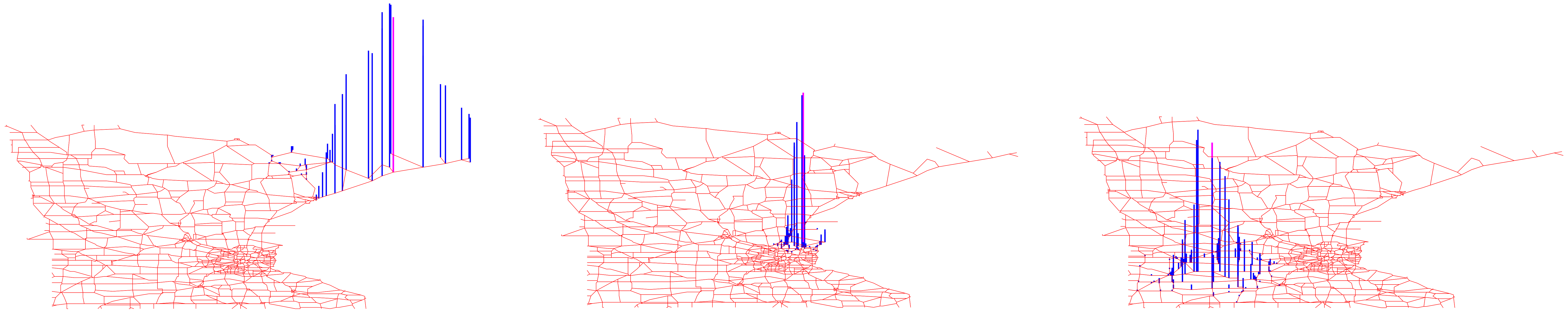




Analysis / Information Extraction

Examples of solutions in signal/data processing for graph signals

- Translations on graphs [Shuman et al., 2013]

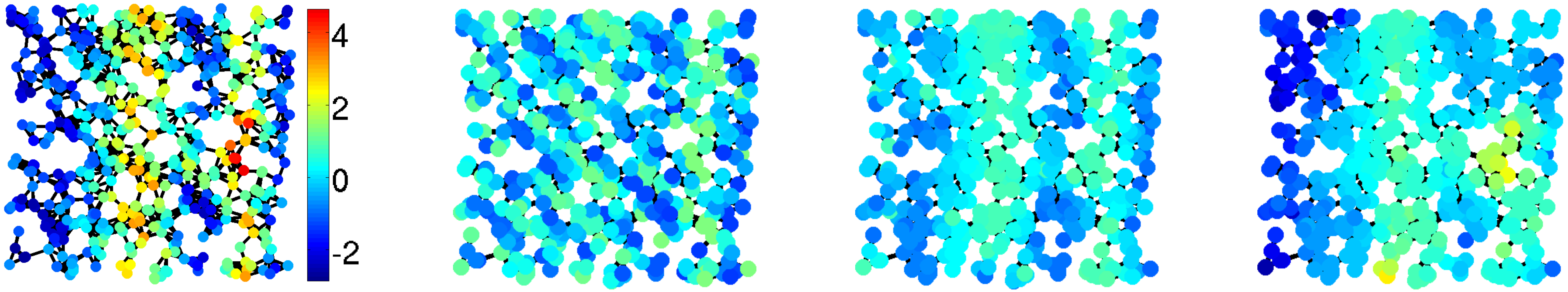


- Denoising on arbitrary graph [Tremblay, Borgnat, 2016]

Noisy graph signal (SNR = 12)   Denoised with filterbanks (SNR = 23.3)

Empirical mode decomposition on graphs

- Objective: decompose a graph signal in various “elementary” modes in a data-driven and non stationary approach



[N. Tremblay, P. Flandrin, P. Borgnat, 2014]

Some bibliography

Books

- Antonio Ortega, *Introduction to Graph Signal Processing*, CUP 2022
- *Cooperative and Graph Signal Processing*, Ed. P. Djuric and C. Richard, Academic Press, 2018
- William H. Hamilton, *Graph Representation Learning*, M&C Publishers, 2020
- Leo Grady, Jonathan Polimeni, *Discrete Calculus (Applied Analysis on Graphs for Computational Science)*, Springer, 2010
- Fan R.K. Chung, *Spectral Graph Theory*, AMS, 197

Some bibliography

Articles - surveys or historical landmarks

- “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains” D.I. Shuman ; S.K. Narang ; P. Frossard ; A. Ortega ; P. Vandergheynst, IEEE Signal Processing Mag., May 2013

- “Discrete Signal Processing on Graphs” A. Sandryhaila, J.M.F. Moura IEEE Transactions on Signal Processing, April 2013

- “Graph signal processing: Overview, challenges, and applications.” A. Ortega, P. Frossard, J. Kovacevic, J.M.F. Moura, and P. Vandergheynst. Proceedings of the IEEE, 106(5):808–828, 2018.

- “Fourier could be a data scientist: From graph Fourier transform to signal processing on graphs” B Ricaud, P Borgnat, N Tremblay, P Gonçalves, P Vandergheynst, Comptes Rendus Physique 20 (5), 474-488, 2019

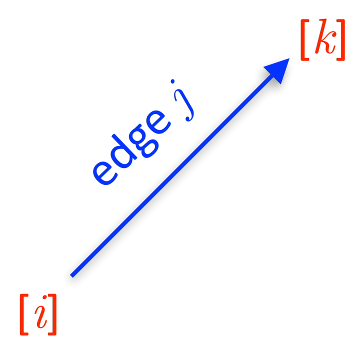
- "Graph Signal Processing: History, development, impact, and outlook", G Leus, AG Marques, JMF Moura, A Ortega, DI Shuman, IEEE Signal Processing Magazine 40 (4), 49-60, 2023

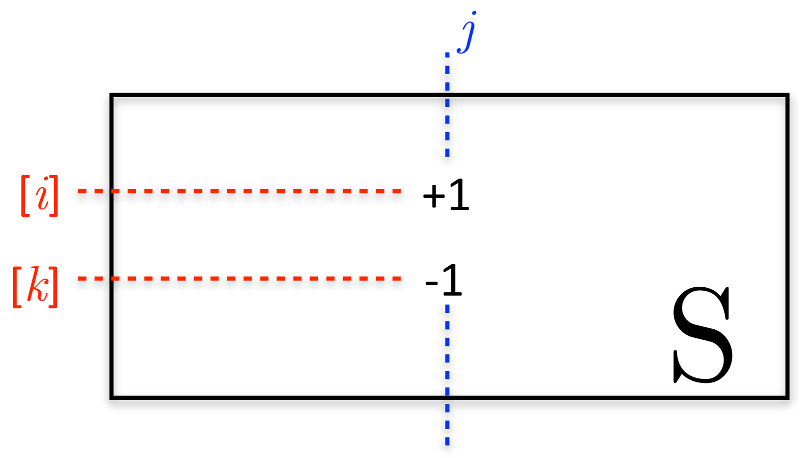
On Graphs, signals, matrices and spectrum

Graphs: Notations and some useful definitions

- Formal def. of a graph : $\mathcal{G} = (V, E)$; *V* set of *N* nodes and *E* of *M* edges
- Adjacency matrix $\mathbf{A}$ s.t. $\mathbf{A}_{ij} = 1$ if $(j, i) \in E$ (warning: convention for the direction). Note that $\mathbf{A} \in \mathbb{R}^{N \times N}$
- Degree of a node : $d_i = \sum_{i \sim j} \mathbf{A}_{ij}$; matrix of degrees $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_N)$
- **Incidence matrix** :

$$\mathbf{S}(i,j) = \begin{cases} +1 & \text{if } e_j = (v_i, v_k) \text{ for some } k \\ -1 & \text{if } e_j = (v_k, v_i) \text{ for some } k \\ 0 & \text{otherwise} \end{cases}$$





Graphs: Some useful definitions

- Walk / Trail / Path   ; (un-)directed edges ; multiple edges or simple
- Diameter: diam $(\mathscr{G})$ is the length of the longest path
- Volume of a subset S of nodes : $\mathrm{vol}(S) = \sum_{i \in S} d(i)$
- Symmetric graphs : $\mathbf{A}_{ij} = \mathbf{A}_{ji}$
- Weighted graphs $\mathbf{W}$ (weight) remplaces $\mathbf{A}$ (sometimes : $\mathbf{K}$ (strength) remplaces $\mathbf{D}$)

Graph Laplacian
with orientation agnostic definitions

With these definitions we have:

$$\mathbf{S}\mathbf{S}^T = \mathbf{D} - \mathbf{A}$$

$\mathbf{L} = \mathbf{D}$ - $\mathbf{A}$ is called unnormalized Laplacian of G

$\mathbf{L}$ does not depend on the orientation (so OK for undirected)

For a weighted graph we have $\mathbf{L} = \mathbf{D}$ **-** $\mathbf{W}$ (attention to degrees)

$\mathbf{L}$ is a symmetric, positive semi-definite matrix

Graph Laplacian

is positive semi-definite for undirected graphs

Proposition: $\mathbf{L}$ is positive semi-definite

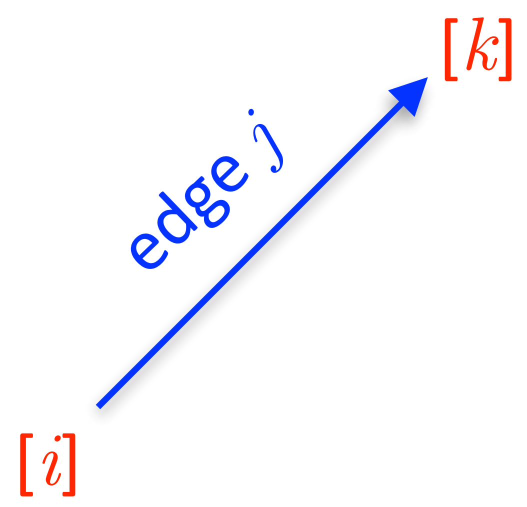
For any N-by-N weight matrix $\mathbf{W}$, if $\mathbf{L} = \mathbf{D}\text{-}\mathbf{W}$ where $\mathbf{D}$ is the degree matrix of $\mathbf{W}$, then

$$x^T \mathbf{L} x = \frac{1}{2} \sum_{i,j} \mathbf{W}(i,j)(x[i] - x[j])^2 \geqslant 0 \quad \forall x \in \mathbb{R}^N$$

Graph Laplacian

is positive semi-definite for undirected graphs

Attribute, signal (function) f defined on the vertices $f \in \mathbb{R}^N$



$(\mathbf{S}^T f)[j] = f[i] - f[k]$ derivative of f along edge j

$F = \mathbf{S}^T f \in \mathbb{R}^M$ gradient of f (F = edge-based signal)

$g = \mathbf{S}G \in \mathbb{R}^N$ divergence of g (G = edge-based signal)

$$\mathbf{L} = \mathbf{S}\mathbf{S}^T \qquad \begin{aligned} f^T \mathbf{L} f &= f^T \mathbf{S}\mathbf{S}^T f \\ &= \|\mathbf{S}^T f\|_2^2 \\ &= \sum_{i \sim k} (f[i] - f[k])^2 \end{aligned}$$

In general for a weighted graph: $f^T \mathbf{L} f = \sum_{i \sim k} \mathbf{W}(i,k)(f[i] - f[k])^2$

This quadratic (Dirichlet) form is a measure of how smooth the signal is

Graph Laplacian: Properties

Since $\mathbf{L}$ is real, symmetric and PSD:

- It has an eigen decomposition into real eigenvalues and eigenvectors λ_i, u_i
- The eigenvalues are non-negative $0 = \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_N$

$$\mathbf{L}\mathbf{1} = 0$$

What can be learned from eigenvectors and eigenvalues ?

Graph Laplacian: Some examples

Path graph

$$\begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & & & \\ & 2 & & & & \\ & & 2 & & & \\ & & & \ddots & & \\ & & & & 2 & \\ & & & & & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & & & & \\ 1 & 0 & 1 & & & \\ & 1 & 0 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & 0 & 1 \\ & & & & 1 & 0 \end{bmatrix}$$

$$\lambda_k = 2 - 2\cos\frac{\pi k}{N} = 4\sin^2\frac{\pi k}{2N},\ k = 0, ..., N-1$$

$$u_k[\ell] = \cos\left(\pi k(\ell + \frac{1}{2})/N\right),\ \ell = 0, ..., N-1$$

DCT II transform

Graph Laplacian: Some examples

Ring graph

$$\begin{pmatrix} 2 & -1 & & & -1 \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ -1 & & & -1 & 2 \end{pmatrix}$$

$$\lambda_k = 2 - 2\cos\frac{\pi k}{N} = 4\sin^2\frac{\pi k}{2N},\ k = 0, ..., N-1$$

$$u_k^c[\ell] = \cos\left(2\pi k\ell/N\right),\ \ell = 0, ..., N-1$$

DCT transform

$$u_k^s[\ell] = \sin\left(2\pi k\ell/N\right),\ \ell = 0, ..., N-1$$

Graph Laplacian: An Analogy

Fourier transform of signals

“Signal processing 101”

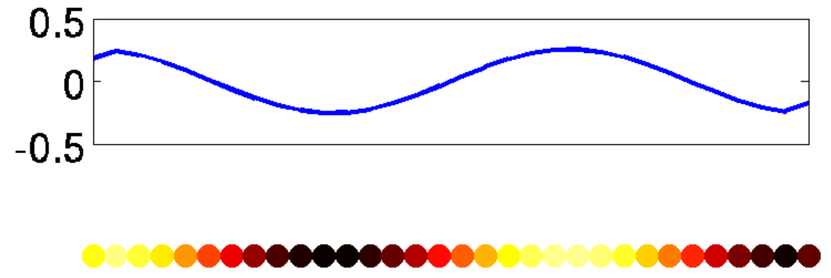
The Fourier transform is of paramount importance:
Given a times series $x_n,\ n = 1, 2, ..., N$, let its Discrete Fourier Transform (DFT) be

$$\forall k \in \mathbb{Z} \quad \hat{x}_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N}$$

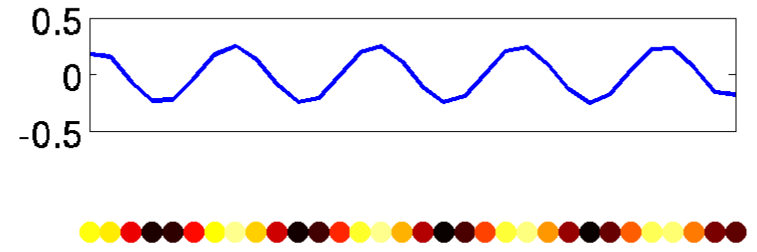
Why ?

- Inversion: $x_n = \frac{1}{N} \sum_{k=0}^{N-1} \hat{x}_k e^{-i2\pi kn/N}$
- Best domain to define **Filtering** (operator is diagonal)
- Definition of the **Spectral analysis** (FT of the autocorrelation)
- Alternate representation domains of signals are useful: Fourier domain, DCT, time-frequency representations, wavelets, chirplets,...

LOW FREQUENCY:



HIGH FREQUENCY:



Graph Laplacian: An Analogy

A fundamental analogy

On *any* graph, the **eigenvectors** χ_i **of the Laplacian matrix L will be considered as the Fourier modes**, and its eigenvalues λ_i the associated (squared) frequencies.

Hence, a Graph Fourier Transform is defined as:

$$\boxed{\hat{x} = \boldsymbol{\chi}^\top x}$$

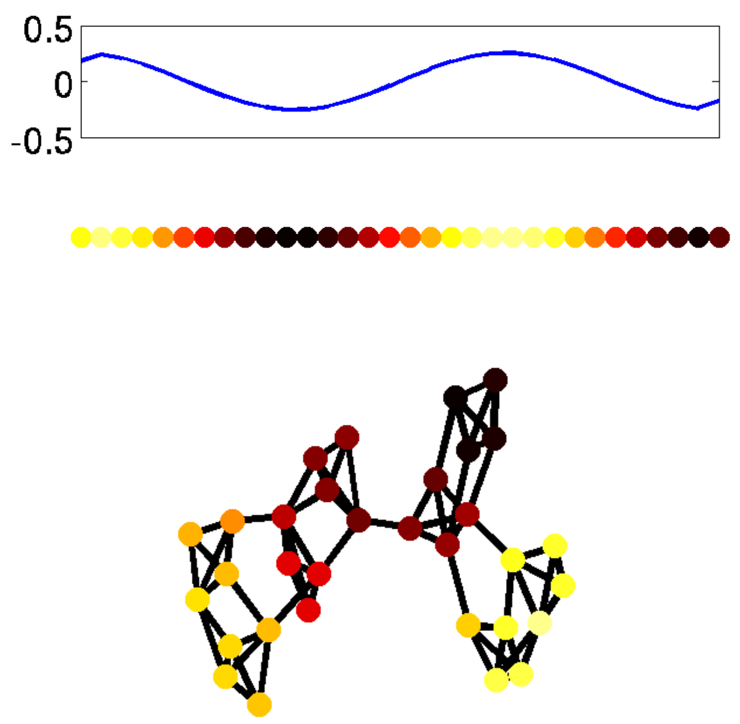
where $\boldsymbol{\chi} = (\boldsymbol{\chi}_0 | \boldsymbol{\chi}_1 | \dots | \boldsymbol{\chi}_{N-1})$.

- Two ingredients:
  - **Fourier modes** = Eigenvectors $\boldsymbol{\chi}_i$ (with increasing oscillations)
  - **Frequencies** = The measures of variations of an eigenvector is linked to its eigenvalue:

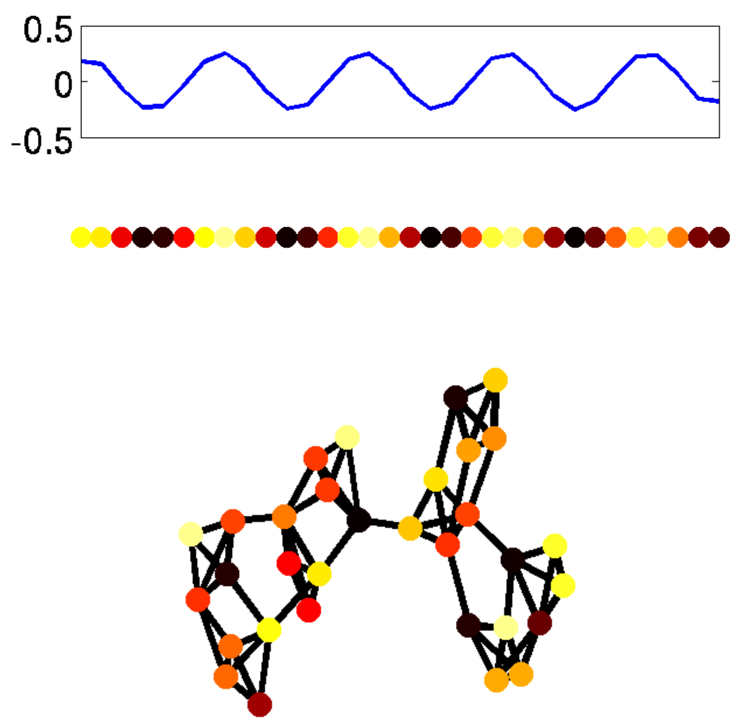
$$\frac{||\nabla \boldsymbol{\chi}_i||^2}{||\boldsymbol{\chi}_i||^2} = \lambda_i$$

because: $\forall \mathbf{x} \in \mathbb{R}^N \quad \sum_{e=(i,j)\in E} A_{ij}(\mathbf{x}_i - \mathbf{x}_j)^2 = \mathbf{x}^\top \mathbf{L}\mathbf{x}$ is the Dirichlet norm

LOW FREQUENCY:



HIGH FREQUENCY:



Graph Laplacian: An Analogy

[Tremblay, Gonçalves, PB, 2017]

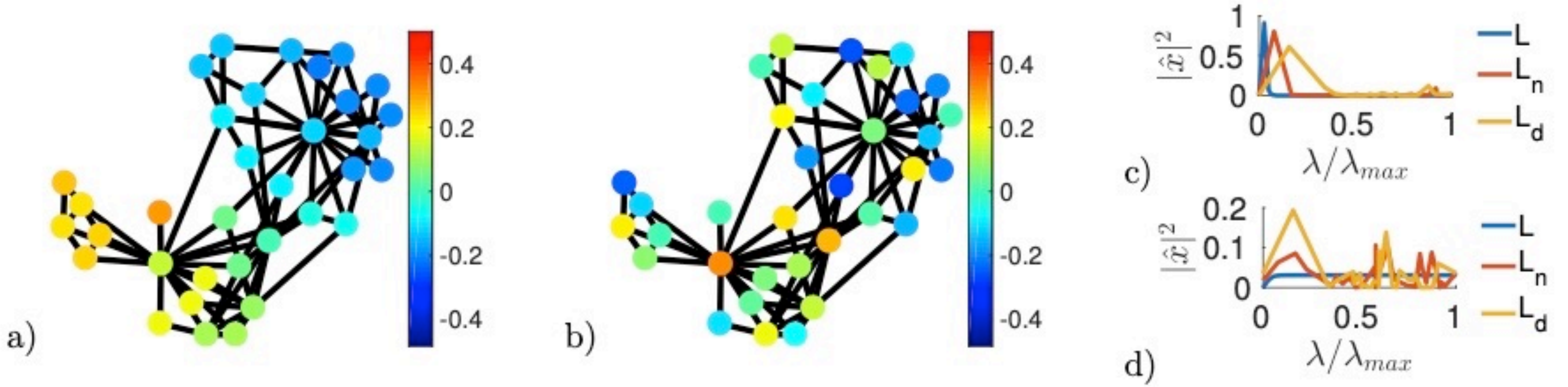
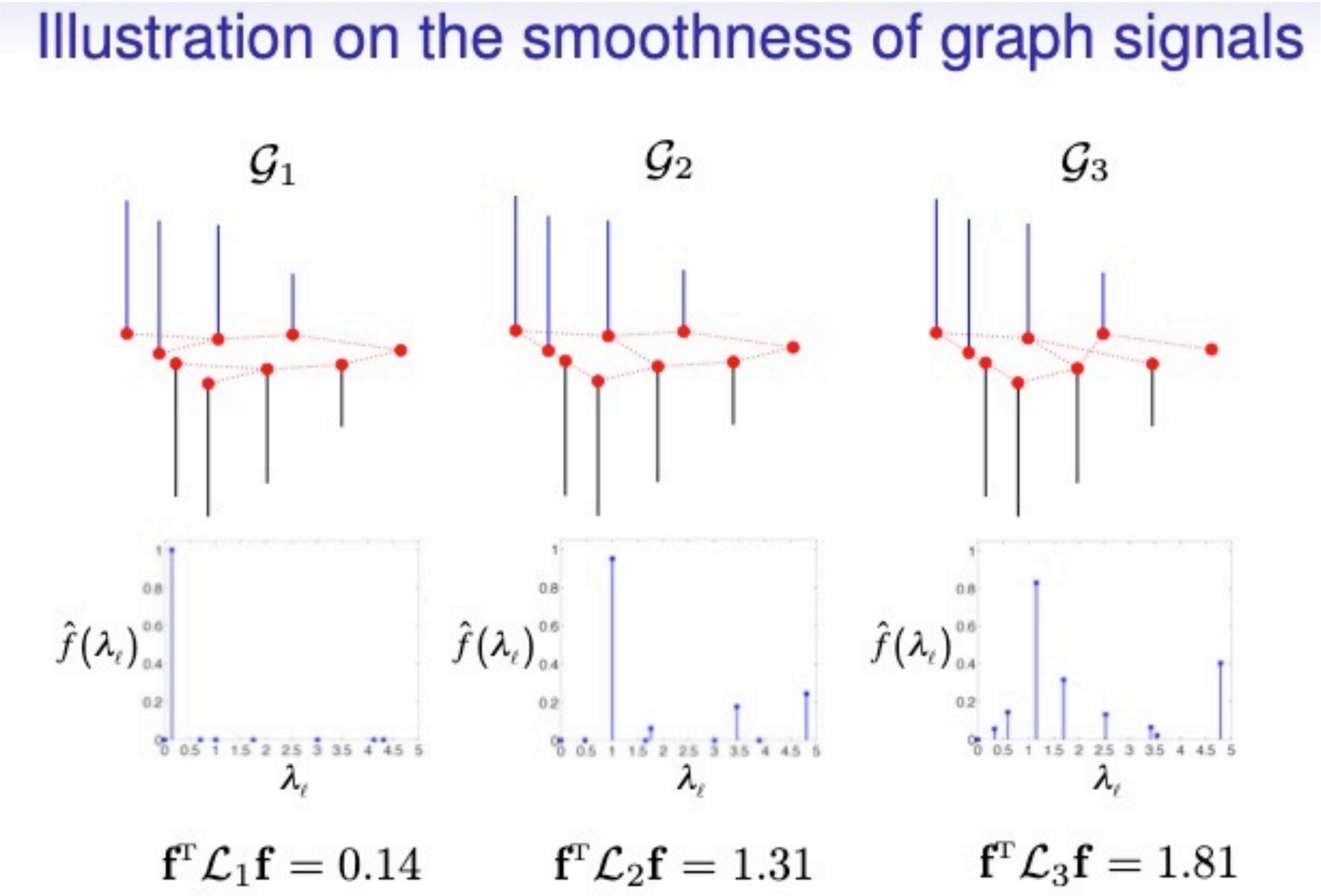


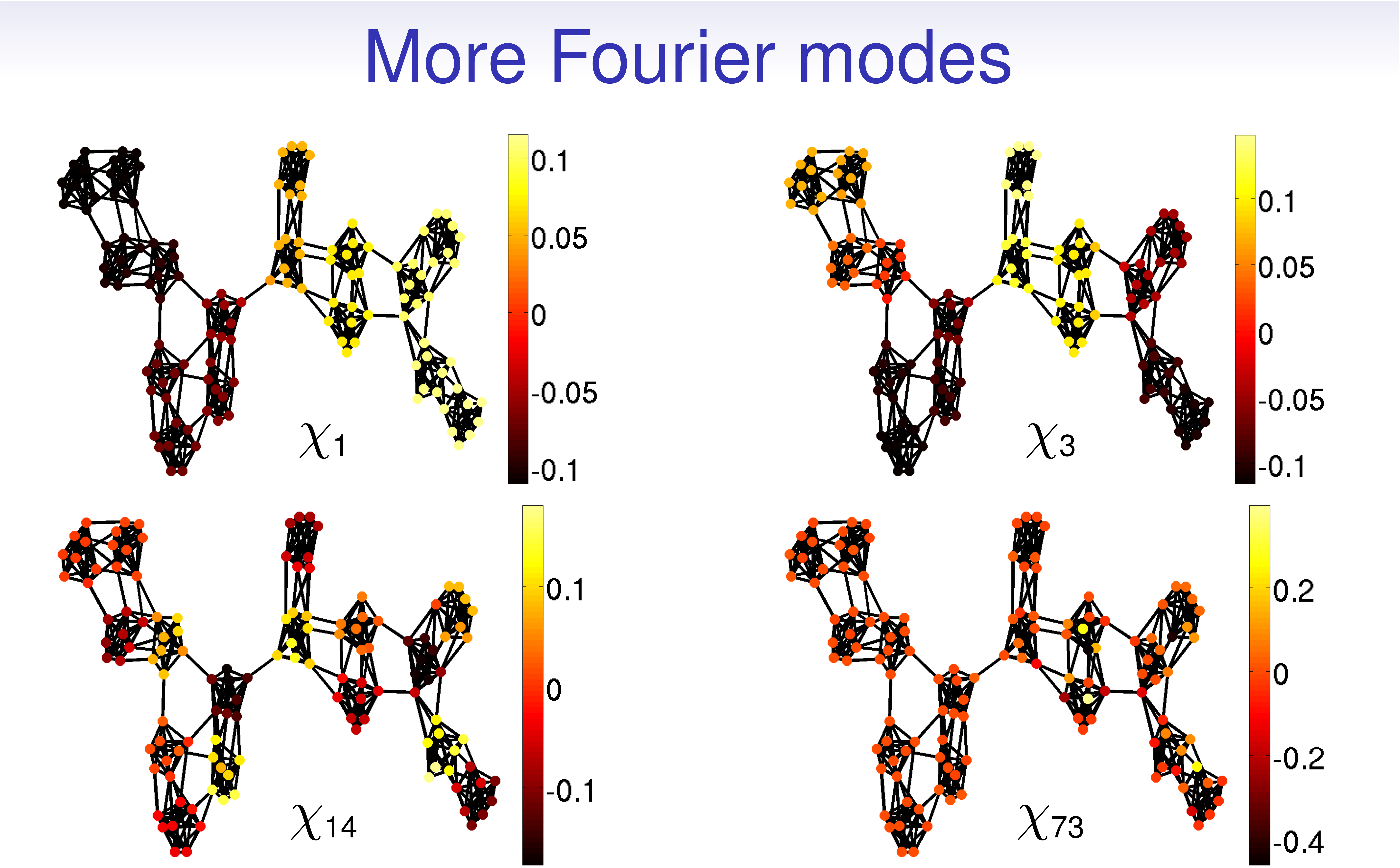
Figure 1: ***Two graph signals and their GFTs.*** *Plots a) and b) represent respectively, a low-frequency and a high-frequency graph signal on the binary Karate club graph [21]. Plots c) and d) are their corresponding GFTs computed for three reference operators:* $\mathbf{L}$, $\mathbf{L_n}$ *and* $\mathbf{L_d}$ *(equivalent to the GFT defined via the adjacency matrix).*

Graph Laplacian: An Analogy

[Vandergheynst & Shuman, 2013]



Graph Laplacian: An Analogy



Graph Laplacian: Properties

- **Multiplicity of eigenvalue λ_0** is equal to the number of connected components

- **Oscillation of the Laplacian eigenvalues:**
  - Property: $u_k = \arg \min_{s \in \mathrm{Span}(u_0, \ldots, u_{k-1}} \frac{s^\dagger L s}{s^\dagger s}$
  - hence u_k is always the functions = oscillations of the smallest global variation a.k.a. frenquency

Graph Laplacian: Properties
(maybe)

- **Prop:** No non-negative local minimum nor non-positive local maximum

- **The (Discrete Local Theorem)**