EPFL

Week 3: Graph theory and signal processing driven features

PRESENTERS
Thanos Charisoudis
Yuxuan Yang

DISCUSSION LEADERS
Leandro von Krannichfeldt
Keivan Faghih Niresi

A bit about me

Thanos Charisoudis

github.com/charisoudis
linkedin.com/in/charisoudis
charisoudis.com

CURRENT EDUCATION

PhD Student, LTS4 Lab, EDEE EPFL

AREAS OF INTEREST

Dynamic Human Representations and Reconstructions

Source: *Populating 3D Scenes by Learning Human-Scene Interaction*

PREVIOUS EDUCATION

MSc in Machine Learning, KTH

Dipl. in Electrical & Computer Engineering, AUTH

The Emerging Field of Signal Processing on Graphs

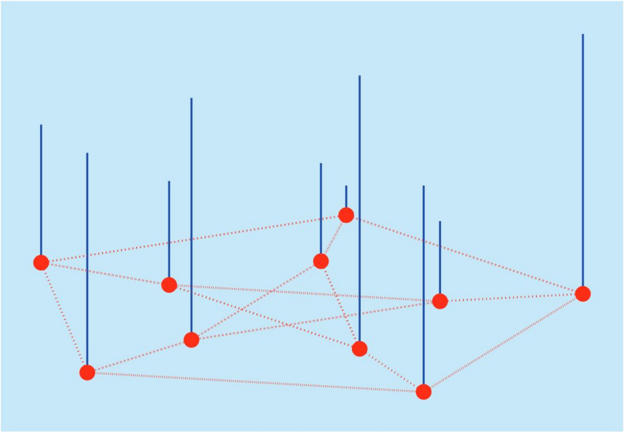
David I Shuman, Sunil K. Narang, **Pascal Frossard**,
Antonio Ortega, and Pierre Vandergheynst

Graph as a Data Structure $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathrm{W}\}$

- Edge weights correspond to vertex similarities
- Edges and weights dictated by physics of the problem or inferred from data

Graph Signals

- Defined on the vertices of the graph (discrete)
- More structured, less regular than analog signals
- Similar techniques to DSP
  - » tasks » »



[FIG1] A random positive graph signal on the vertices of the Petersen graph. The height of each blue bar represents the signal value at the vertex.

Graph Signal Processing

- **f: V → R**
- no SI, time/frequency duality properties
- employ localized operations
- use information from neighbouring vertices

1. construct graph
2. transform to spectral domain
3. solve downstream task

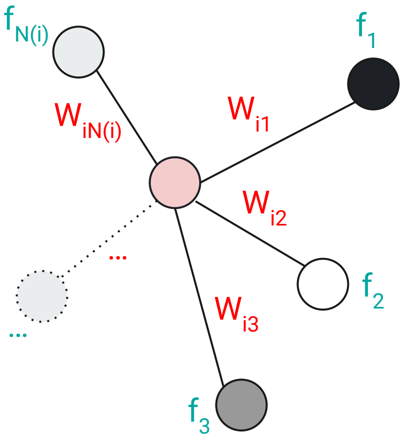
SPECTRAL GRAPH DOMAIN

The Laplacian Matrix

degree matrix (diag)

adjacency matrix (symmetric for underícted)

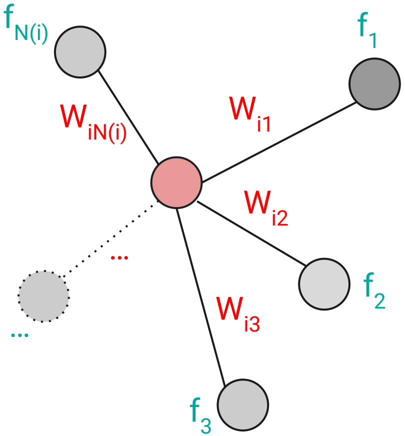
$$\mathcal{L} := \mathbf{D} - \mathbf{W}$$



1st order differences on graph:

$$(\mathcal{L}f)(i) = \sum_{j \in \mathcal{N}_i} W_{i,j}[f(i) - f(j)]$$

1st order smoothing operation



The Laplacian Matrix

For undirected graphs:

- real, symmetric
- **orthonormal eigenvectors**
- **real, non-negative eigenvalues**

Graph Spectrum: $\sigma(\mathcal{L}) := \{\lambda_0, \lambda_1, \ldots, \lambda_{N-1}\}$

SPECTRAL GRAPH DOMAIN

The Graph Fourier Transform

forward:

$$\hat{f}(\lambda_\ell) := \langle \mathbf{f}, \mathbf{u}_\ell \rangle = \sum_{i=1}^{N} f(i) u_\ell^*(i)$$

inverse:

$$f(i) = \sum_{\ell=0}^{N-1} \hat{f}(\lambda_\ell) u_\ell(i)$$

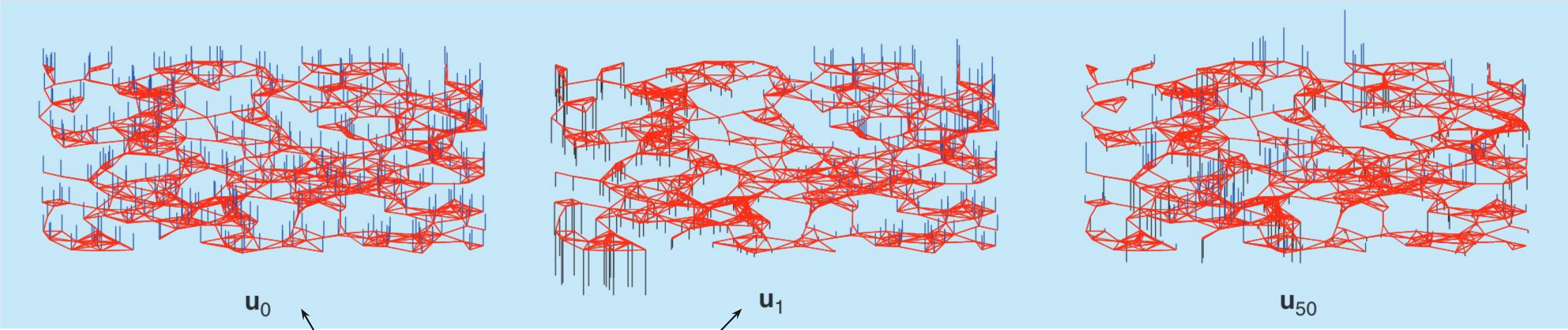
For connected graphs:

- lower l's ⇒ smoother eigenvectors (l=0, constant eigvec)
- higher l's ⇒ more oscillations with increasing edge weights
- **graph frequencies relate to zero crossing of a graph signals**

EPFL

SPECTRAL GRAPH DOMAIN

The Graph Fourier Transform



constant

Fiedler (smooth)

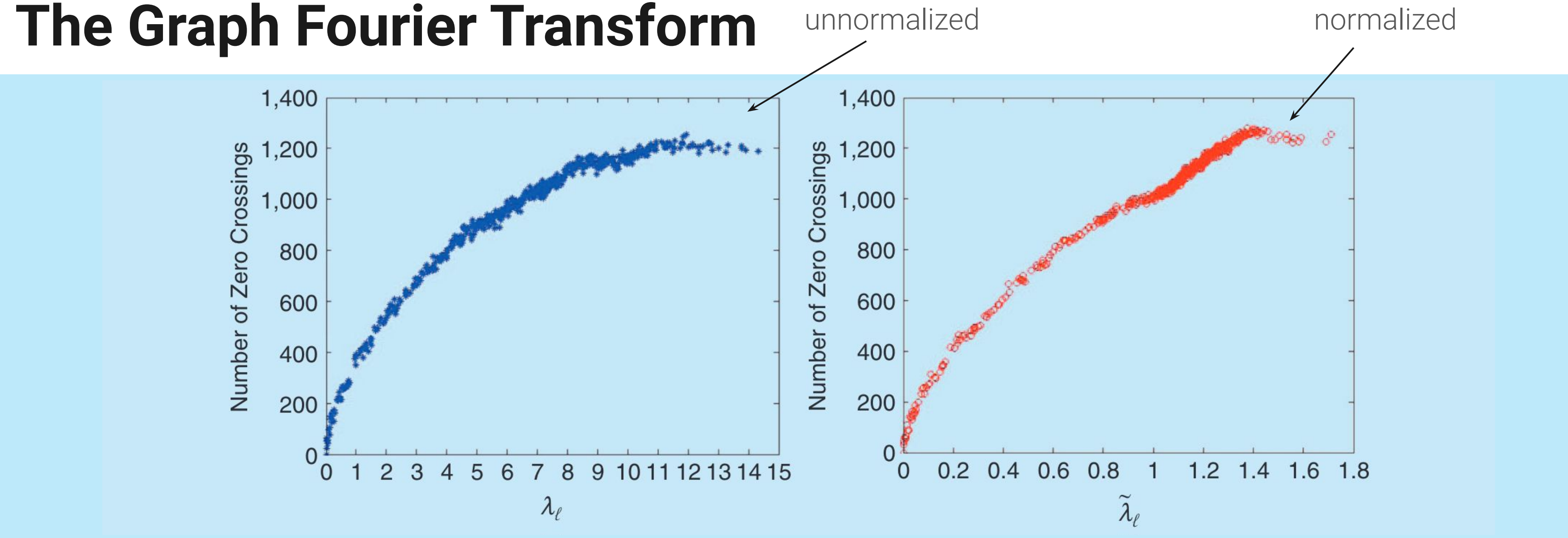
Visualizing Laplacian eigvec as signals.

For connected graphs:

- lower l's ⇒ smoother eigenvectors (l=0, constant eigvec)
- higher l's ⇒ more oscillations with increasing edge weights
- **graph frequencies relate to zero crossing of a graph signals**

SPECTRAL GRAPH DOMAIN

The Graph Fourier Transform

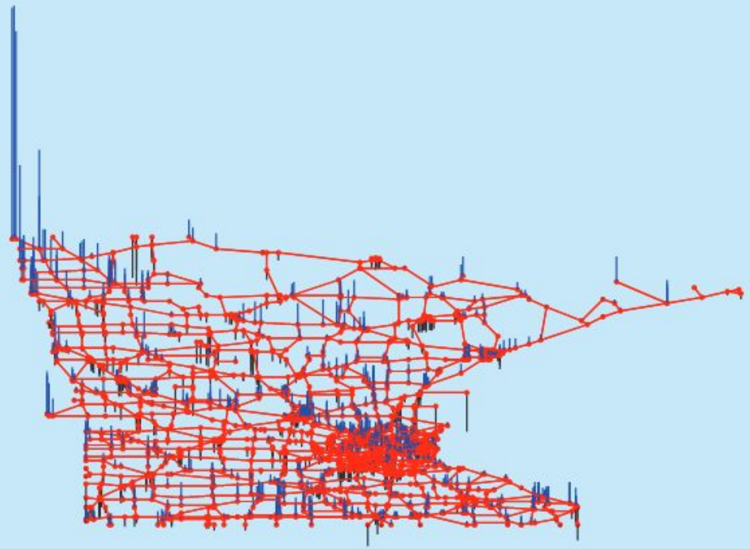


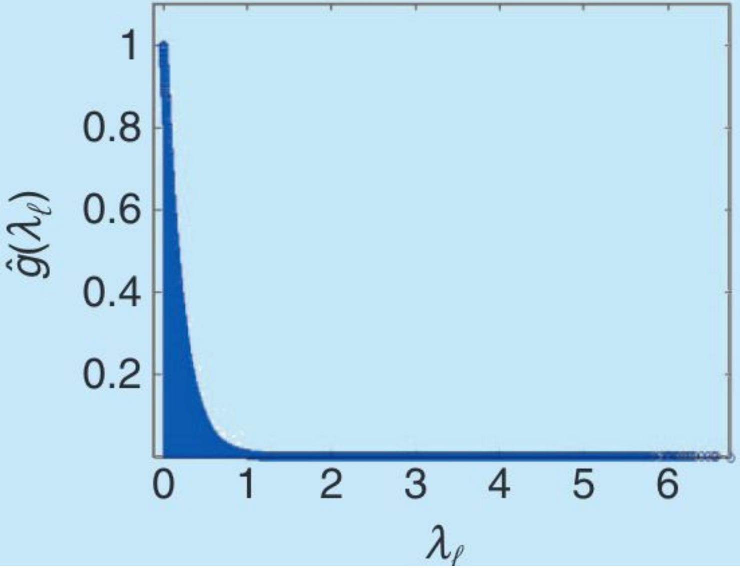
Zero-crossings as a notion of graph frequency.

For connected graphs:

- lower l's ⇒ smoother eigenvectors (l=0, constant eigvec)
- higher l's ⇒ more oscillations with increasing edge weights
- **graph frequencies relate to zero crossing of a graph signals**

The Graph Fourier Transform

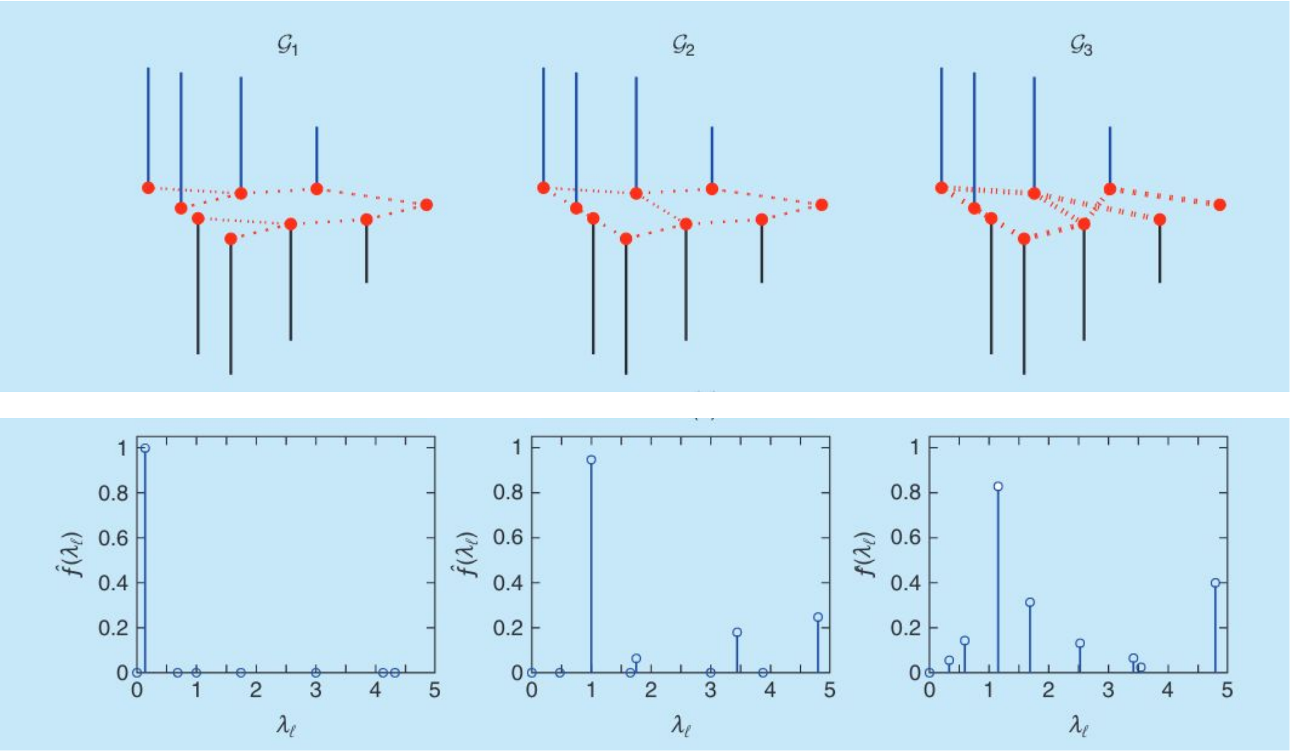




Equivalent representations of a graph signal in the vertex and graph spectral domains. Signal with gaussian edge weights (can be defined directly on the spectral domain).

SPECTRAL GRAPH DOMAIN

The Graph Fourier Transform



Graph structure plays a key role in spectral analysis.

Measures of Graph Smoothness

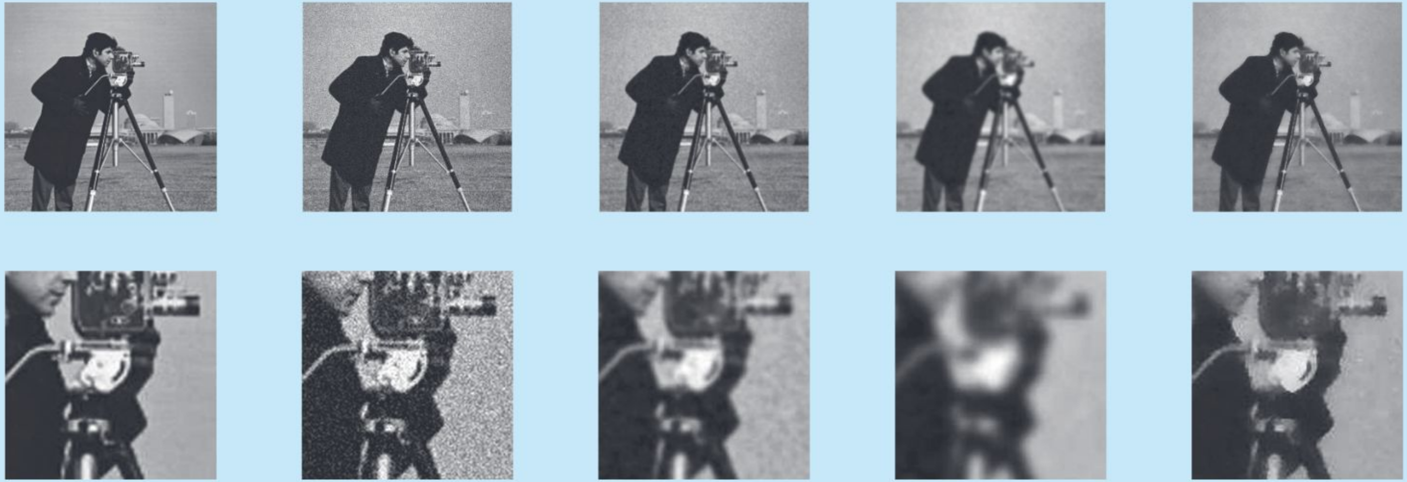
- derivative wrt edge: $\left.\frac{\partial f}{\partial e}\right|_i := \sqrt{W_{i,j}}[f(j) - f(i)]$
- gradient at v_i: $\nabla_i f := \left[\left\{\left.\frac{\partial f}{\partial e}\right|_i\right\}_{e \in \mathcal{E} \text{ s.t. } e=(i,j) \text{ for some } j \in \mathcal{V}}\right]$
- local smoothness: $\|\nabla_i f\|_2 := \left[\sum_{j \in \mathcal{N}_i} W_{i,j}[f(j) - f(i)]^2\right]^{\frac{1}{2}}$
- global smoothness: $S_p(f) := \frac{1}{p}\sum_{i \in V} \|\nabla_i f\|_2^p = \frac{1}{p}\sum_{i \in V}\left[\sum_{j \in \mathcal{N}_i} W_{i,j}[f(j) - f(i)]^2\right]^{\frac{p}{2}}$

p=1: total variation

p=2: laplacian quadratic form

can be used as regularizers in graph objective functions

Measures of Graph Smoothness



$$\underset{\mathbf{f}}{\operatorname{argmin}}\{\|\mathbf{f}-\mathbf{y}\|_2^2+\gamma\mathbf{f}^\top\mathcal{L}\mathbf{f}\}$$

$$f_*(i)=\sum_{\ell=0}^{N-1}\left[\frac{1}{1+\gamma\lambda_\ell}\right]\hat{y}(\lambda_\ell)\,u_\ell(i),$$

can be used as regularizers in graph objective functions

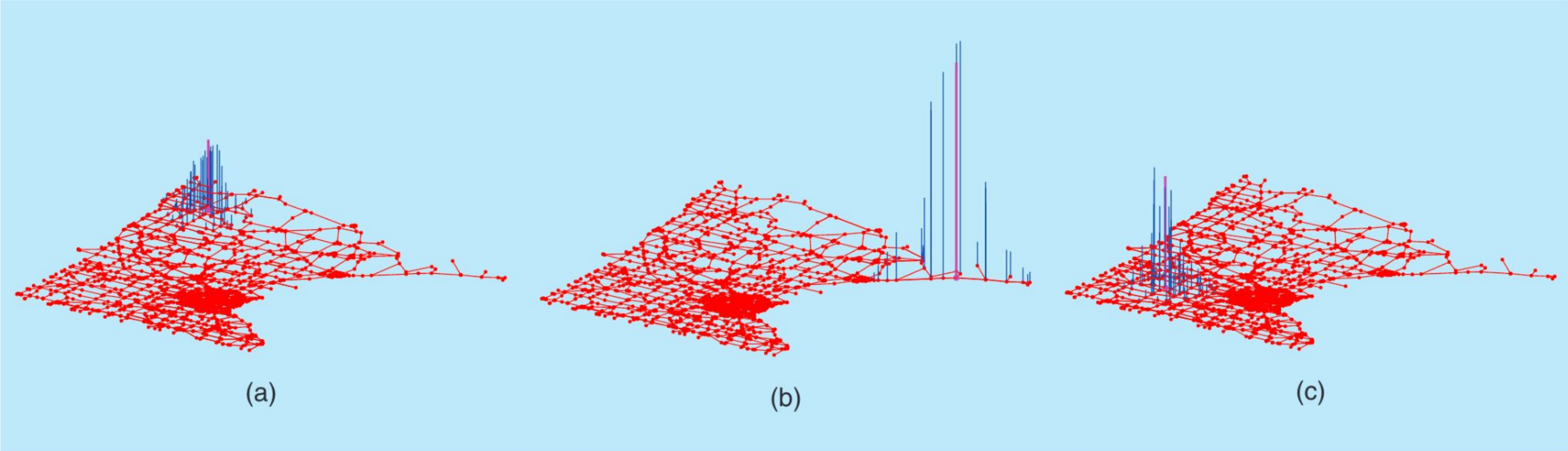
Operations For Signals On Graphs

CLASSICAL	GRAPHS
FILTERING / CONVOLUTION	
freq: $\hat{f}_{\text{out}}(\xi) = \hat{f}_{\text{in}}(\xi)\hat{h}(\xi)$ time: $f_{\text{out}}(t) = \int_{\mathbb{R}} \hat{f}_{\text{in}}(\xi)\hat{h}(\xi)e^{2\pi i \xi t} d\xi = \int_{\mathbb{R}} f_{\text{in}}(\tau)h(t-\tau)d\tau =: (f_{\text{in}} * h)(t)$	$f_{\text{out}}(i) = \sum_{\ell=0}^{N-1} \hat{f}_{\text{in}}(\lambda_\ell)\hat{h}(\lambda_\ell)u_\ell(i).$ $\hat{f}_{\text{out}}(\lambda_\ell) = \hat{f}_{\text{in}}(\lambda_\ell)\hat{h}(\lambda_\ell)$
TRANSLATION	
$(T_\upsilon f)(t) := f(t-\upsilon)$ $(T_\upsilon f)(t) = (f * \delta_\upsilon)(t)$ viewed as convolution	$(T_n g)(i) := \sqrt{N}(g * \delta_n)(i) \overset{(20)}{=} \sqrt{N} \sum_{\ell=0}^{N-1} \hat{g}(\lambda_\ell)u_\ell^*(n)u_\ell(i)$ analogous but changes the signal

Operations For Signals On Graphs

CLASSICAL	GRAPHS
FILTERING / CONVOLUTION	
freq: $\hat{f}_{\text{out}}(\xi) = \hat{f}_{\text{in}}(\xi)\hat{h}(\xi)$ time: $f_{\text{out}}(t) = \int_{\mathbb{R}} \hat{f}_{\text{in}}(\xi)\hat{h}(\xi)e^{2\pi i\xi t}d\xi = \int_{\mathbb{R}} f_{\text{in}}(\tau)h(t-\tau)d\tau =: (f_{\text{in}} * h)(t)$	$f_{\text{out}}(i) = \sum_{\ell=0}^{N-1} \hat{f}_{\text{in}}(\lambda_\ell)\hat{h}(\lambda_\ell)u_\ell(i).$ $\hat{f}_{\text{out}}(\lambda_\ell) = \hat{f}_{\text{in}}(\lambda_\ell)\hat{h}(\lambda_\ell)$
TRANSLATION	
$(T_v f)(t) := f(t-v)$ $(T_v f)(t) = (f * \delta_v)(t)$ viewed as convolution	$(T_n g)(i) := \sqrt{N}(g * \delta_n)(i) \overset{(20)}{=} \sqrt{N}\sum_{\ell=0}^{N-1} \hat{g}(\lambda_\ell)u_\ell^*(n)u_\ell(i)$ analogous but changes the signal

Operations For Signals On Graphs



[FIG5] The translated signals (a) $T_{100}\,g$, (b) $T_{200}\,g$, and (c) $T_{2,000}\,g$, where g is the heat kernel shown in Figure 4(a) and (b).

EPFL

Graph Wavelet Transform

Why?

- Simultaneously localize signal on the time and frequency domain
- Exploit the time-frequency resolution tradeoff

Spread of graph:

vertex:

$$\Delta^2_{\mathcal{G},i}(\mathrm{f}) := \frac{1}{\|\mathrm{f}\|_2^2} \sum_{j \in \mathcal{V}} [d_{\mathcal{G}}(i,j)]^2 [f(j)]^2.$$

spectral:

$$\Delta^2_{\sigma}(\mathrm{f}) := \min_{\mu \in \mathbb{R}_+} \left\{ \frac{1}{\|\mathrm{f}\|_2^2} \sum_{\lambda \in \sigma(\mathcal{L})} [\sqrt{\lambda} - \sqrt{\mu}]^2 [\hat{f}(\lambda)]^2 \right\}.$$

design bases that are localized in both domains

LOCALIZED MULTISCALE TRANSFORMS

Graph Wavelet Transform

VERTEX

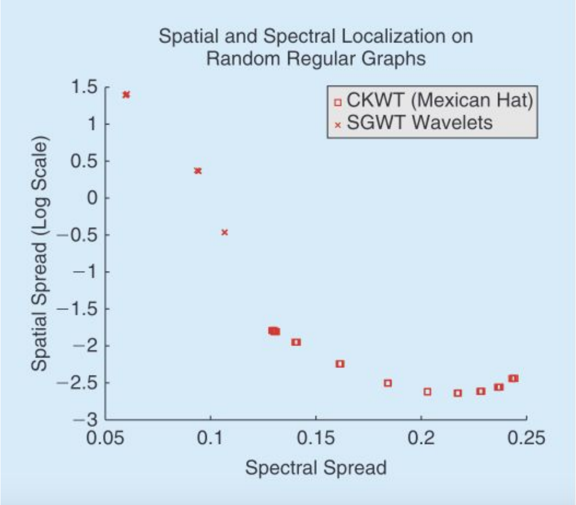
$$\psi^{\text{CKWT}}_{k,i}(j) = \frac{a_{k,\tau}}{|\partial\mathcal{N}(i,\tau)|}, \quad \forall j \in \partial\mathcal{N}(i,\tau)$$

support only on K-hop neighbourhood of v_i

SPECTRAL

$$\psi^{\text{SGWT}}_{\text{scal},i} := T_i \text{h} = \hat{h}(\mathcal{L})\delta_i,$$

translated low pass filters



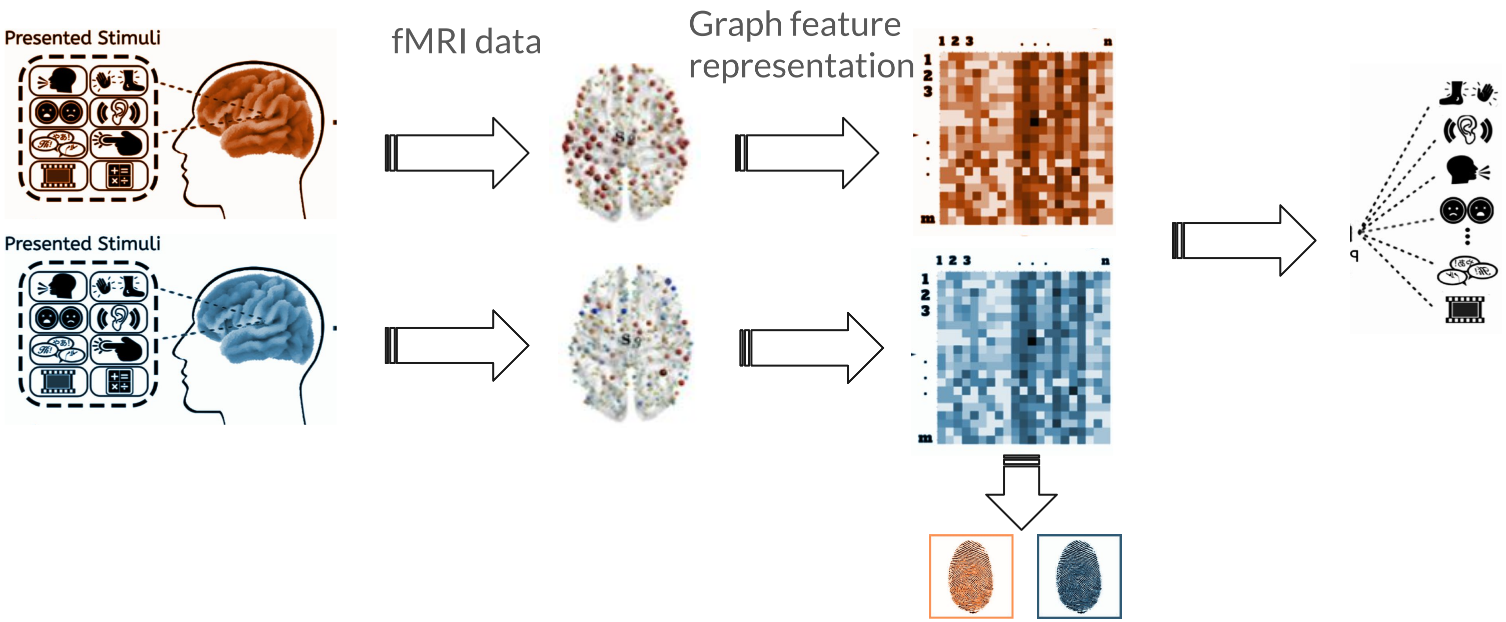
Brain structure-function coupling provides signatures for task decoding and individual fingerprinting

Alessandra Griffa [a,b], Enrico Amico [b,c], Raphaël Liégeois [b,c],
Dimitri Van De Ville [b,c,d], Maria Giulia Preti [b,c,d]

a Department of Clinical Neurosciences, Division of Neurology, Geneva University Hospitals and Faculty of Medicine, University of Geneva, Geneva, Switzerland
b Center of Neuroprosthetics, Ecole Polytechnique Fédérale De Lausanne (EPFL), Institute of Bioengineering, Geneva, Switzerland
c Department of Radiology and Medical Informatics, University of Geneva (UNIGE), Geneva, Switzerland
d CIBM Center for Biomedical Imaging, Switzerland

Introduction

Task decoding and individual fingerprinting



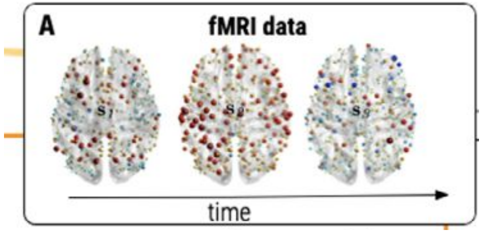
EPFL

Background

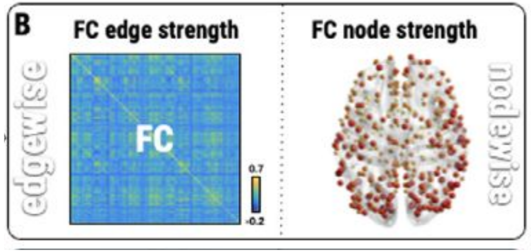
- Analysis on the data during resting state and seven different tasks are through structure-function signatures quantification of
  - SDI (Structural-Decoupling Index)
    - Quantifies the degree of local (dis)alignment between structure and function.
  - c(coupled)-FC & d(decoupled)-FC(functional connectivity)

Methodology - preprocessing

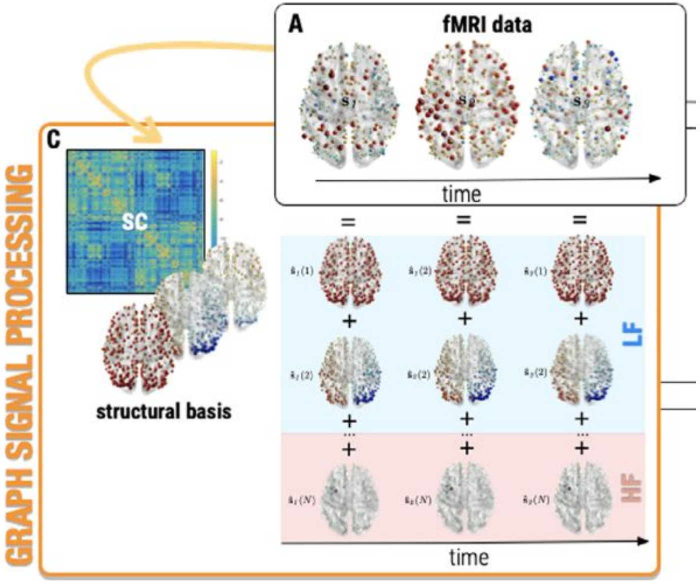
Data recording
fMRI acquired with 8 different task conditions on 100 unrelated healthy subjects. (resting state and 7 tasks: emotion, gambling, language, motor, relation, social, working memory)
Sequences were pre-processed with state-of-the-art pipelines to obtain regional functional time courses and structural connections



Feature extraction
FC edge strength:
Pearson's correlation between pairwise time courses
FC nodal strength:
Sum of absolute values of all the connections



Methodology - feature representation



SC Laplacian

$L = I\text{-}A_{symm}$

Structural harmonics u_k by eigendecomposition:

$LU=L\Lambda$

Where eigenvalue $[\Lambda]_{k,k}=\lambda_k$ interpreted as spatial frequency of the corresponding eigenvector u_k.

Frequency filter is applied.

LF: c first eigenvectors complemented by zero

HF: last $(N\text{-}c)$ first eigenvectors complemented by zero

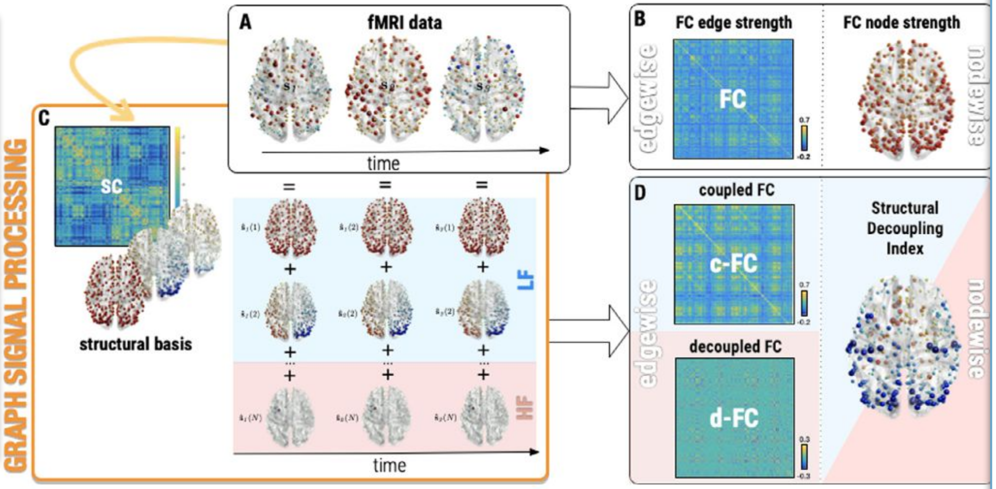
Methodology - feature representation

SC Laplacian:
$L = I\text{-}A_{symm}$

Eigendecomposition:
$LU = L\Lambda$

Low-pass filter:
$U^{(low)}$

High-pass filter:
$U^{(high)}$



Function data is projected:
$\hat{s}_t = U^T s_t$

Low-frequency functional activity comp:
$s^C_t = U^{(low)} U^T s_t$

High-frequency functional activity comp:
$s^D_t = U^{(high)} U^T s_t$

c-FC and d-FC are computed with Pairwise Pearson's correlations of s^C_t, s^D_t.

L2 norm across time measures (de)coupling:
$SDI = ||s^C||_2 / ||s^D||_2$

Methodology - overview

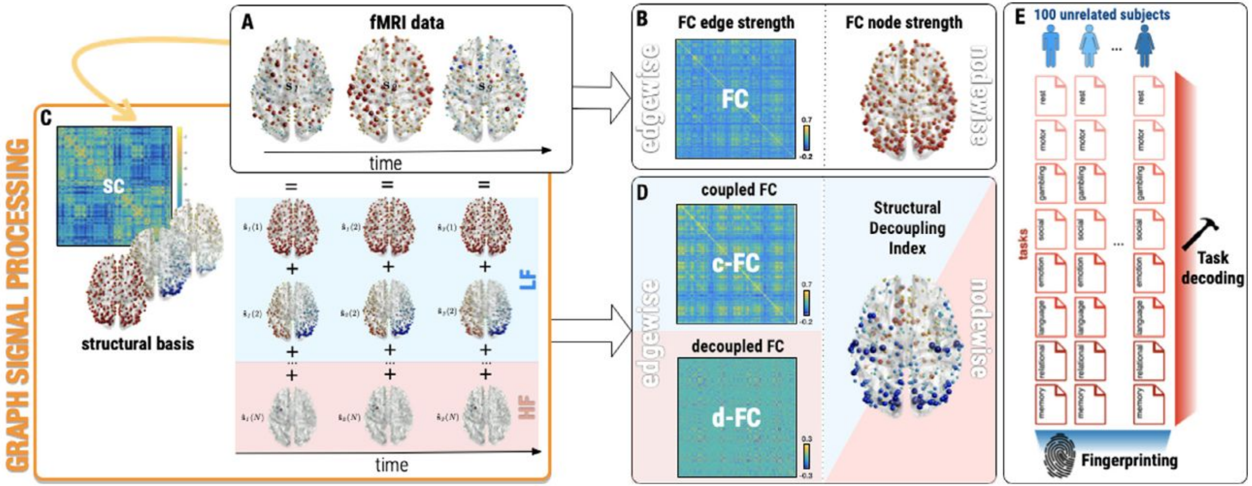


Fig. 1. Method workflow. From fMRI nodal signals at each timepoint (A), functional connectivity (FC) is evaluated through conventional edgewise (FC matrix) and nodewise (FC node strength) measures (B). The graph signal processing (GSP) pipeline is applied to decompose functional signals into the structural harmonics obtained from the eigendecomposition of the structural connectome (SC) Laplacian (C). Functional signals are then filtered into two components; i.e., one coupled and one decoupled from structure, by applying ideal low pass (light blue) and high pass (pink) filters in the graph spectral domain (C). Edgewise and nodewise metrics evaluating structure-function coupling are obtained by computing FC matrices from coupled and decoupled signals (coupled and decoupled FC (c-FC and d-FC), respectively), and the structural decoupling index (SDI). Edgewise and nodal measures of both FC (B) and structure-function coupling (D) enter separate support vector machine (SVM) classifications with various cross validation settings to test their task decoding and fingerprinting value, quantified by task and subject identification accuracies (E).

Results - accuracy

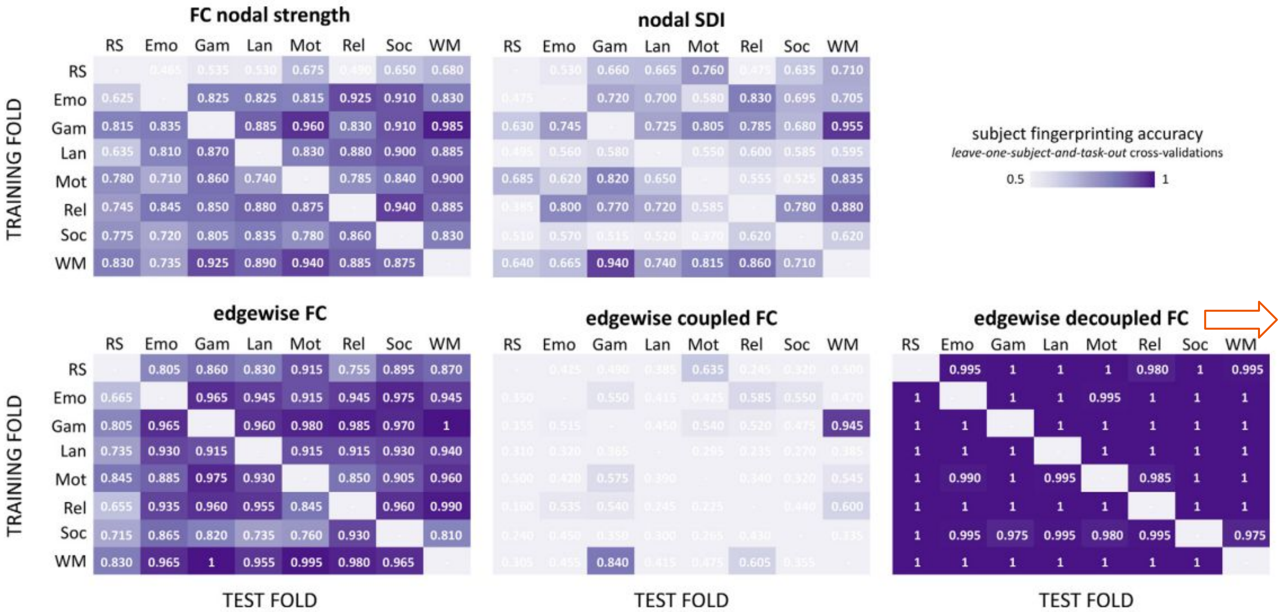
- SVM is used to classify different task-related states based on the extracted features.

Table 1
Task decoding, subject fingerprinting, and brain-cognition relationships. First column: task decoding accuracies for nodewise (FC nodal strength; SDI) and edgewise (FC, c-FC, d-FC) functional and structure-function coupling measures estimated with 100-fold leave-one-subject-out cross-validation and once-versus-one multiclass linear SVM classifiers. Second column: subject fingerprinting accuracies estimated with 800-fold leave-one-subject's-task-out cross-validation and one-versus-all multiclass SVM classifiers. Third column: brain-cognition r-squared (r^2) computed as the squared Pearson's correlation coefficient between the brain and cognition latent scores obtained from significant partial least squares correlation (PLSC) components. The brain-cognition r^2 quantifies the amount of inter-individual cognitive traits' variance explained by the five different brain features, respectively.

	Task Decoding *accuracy*	Subject Fingerprinting *accuracy*	Brain-Cognition r^2
FC nodal strength	0.544	0.984	0.211
nodal SDI	0.756	0.997	0.180
FC	0.919	0.964	0.224
c-FC	0.893	0.972	0.209
d-FC	0.873	1.000	0.654

Edge based features well characterizing task decoding patterns.

Results - cross-task fingerprinting accuracies



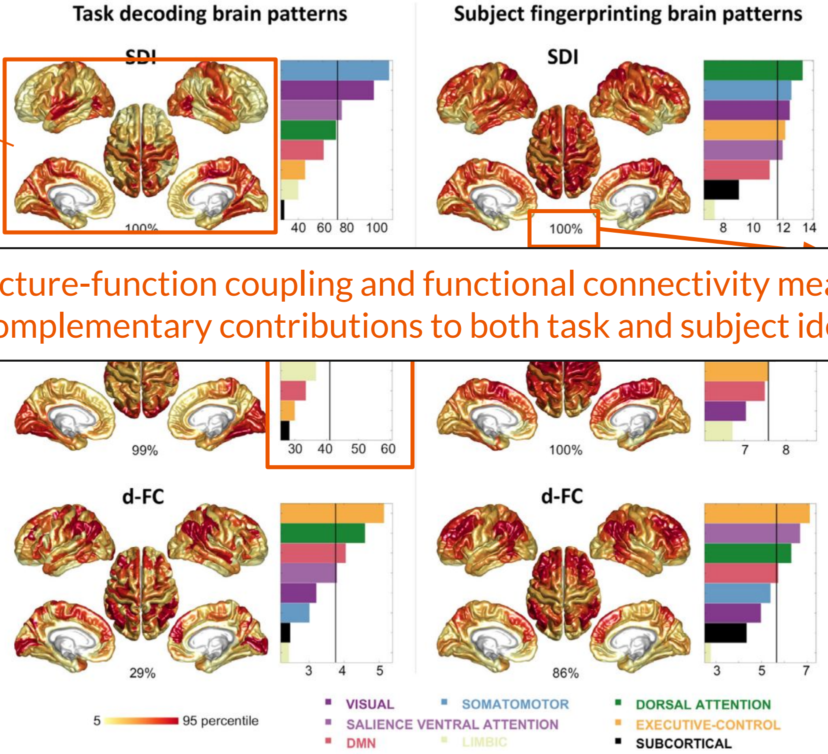
Largely outperform than other metrics!

Fig. 3. Cross-task fingerprinting accuracies for functional and structure-function coupling measures. Subject classification accuracies when using only one condition –task or resting state– for training (matrices' rows) and one for testing (matrices' columns), with all pairwise task combinations explored and for all nodewise (FC nodal strength; SDI) and edgewise (FC, c-FC, d-FC) measures. Classification accuracies were assessed with *leave-one-subject-and-task-out* cross-validation and *one-versus-all* multiclass SVM classifiers. RS=resting state; Emo=emotion; Gam=gambling; Lan=language; Mot=motor; Rel=relational; Soc=social; WM=working memory.

Results - effect

Distribution of F-values from two-factor ANOVA analyses

Average of F-values in different regions.



statistically significant brain
...ctions
...cted($p<0.05$)

Structure-function coupling and functional connectivity measures provide complementary contributions to both task and subject identification.

Task decoding pattern involved more prominently regions such as visual, somatomotor, auditory.

Subject fingerprinting is more spatially distributed.

Recovering Gene Interactions from Single-Cell Data Using Data Diffusion

David van Dijk, Roshan Sharma, Juozas Nainys, Kristina Yim, Pooja Kathail, Ambrose J. Carr , et. al.

EPFL

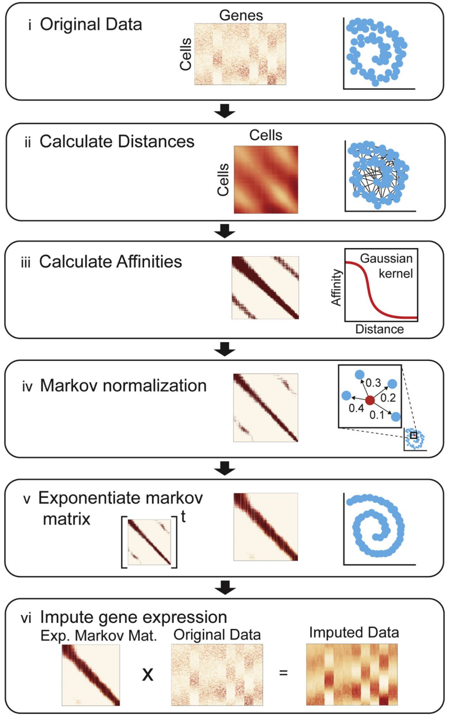
Background

Dropout: Single-cell RNA sequencing suffer from under-sampling of mRNA molecules, which can obscure gene-gene relationships.

MAGIC: Markov affinity-based graph imputation of cells.
A method that shares information across similar cells via data diffusion, to denoise the cell count matrix and fill in missing transcripts.

Methodology

Embed cell phenotypes in a lower dimensional manifold using a nearest neighbor graph.



Matrix of cells by genes (middle) of the data (right).

Identifies the cells that are most similar and aggregates gene expression across these highly similar cells to impute gene expression that corrects for dropout and other sources of noise.

Nearest neighbors in the raw data do not necessarily represent the most biologically similar cells due to sparsity.

kernel.

Similarity between two cells decreases **exponentially** with their distance.

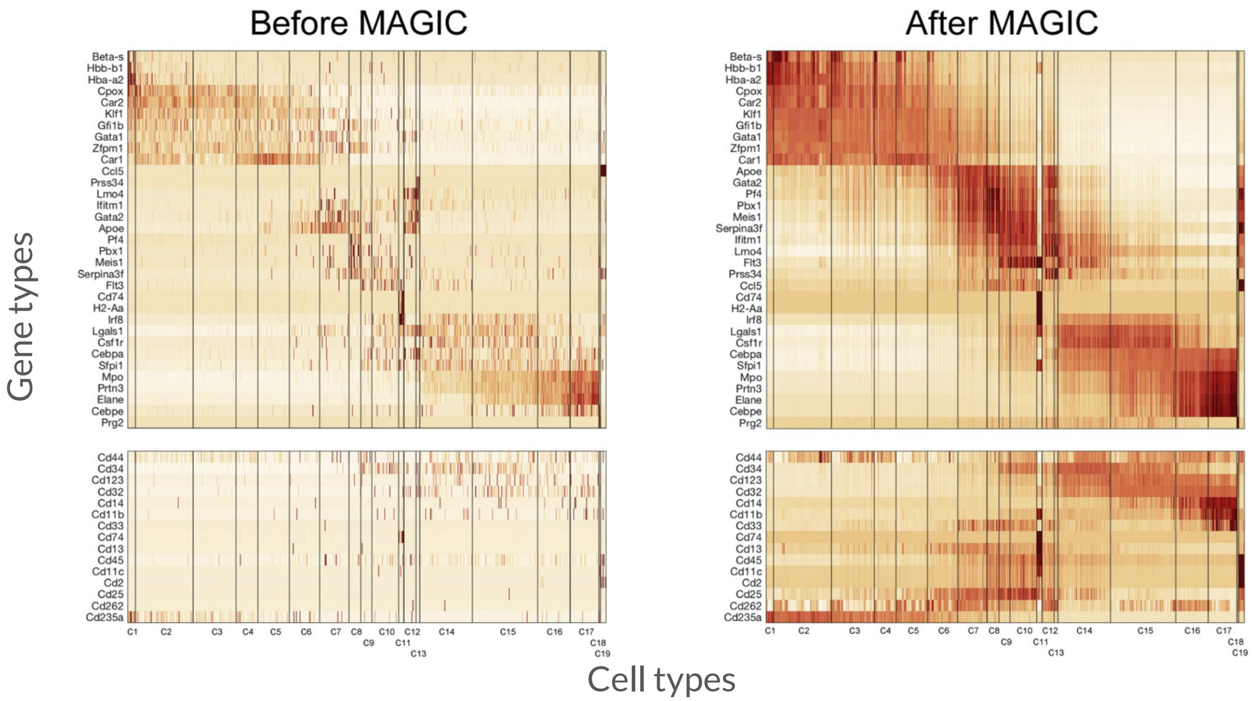
Normalized affinities are shown for a single point as transition probabilities, resulting in a Markov matrix.

M representing the probability distribution of transitioning from one cell to another in a single step.

Multiplication of the exponentiated Markov matrix by the original data matrix to obtain a denoised and imputed data matrix.

Case study

MAGIC Enhances Structures in Bone Marrow, data evaluated in mouse bone marrow dataset.



Discussion