The Power of Graphs Learning in Immersive Communications

Laura Toni

UCL - University College London

SocialXR - SPRING SCHOOL
March 4th-8th, 2024
CWI

A massive thanks to

UCL

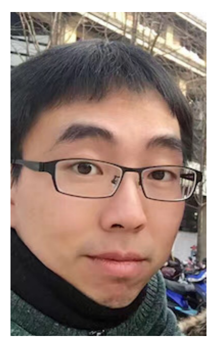
Silvia Rossi

Pedro Gomes

Alan Guedes

Cagri Ozcinar

Aljosa Smolic



Xiaowen Dong

Pablo Cesar

Irene Viola

Francesca De Simone

Pascal Frossard

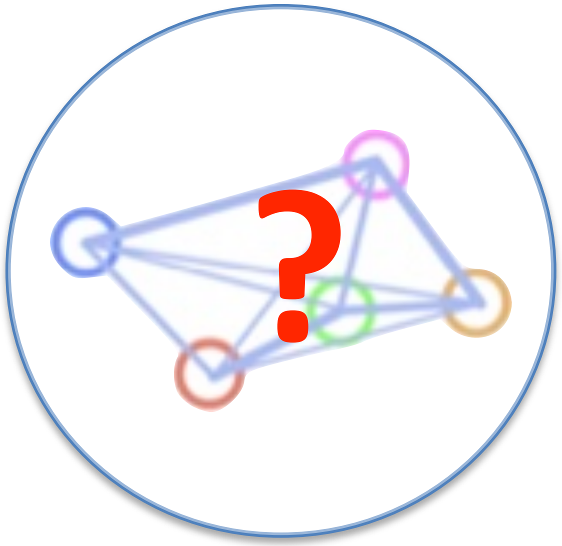
Trinity College Dublin
The University of Dublin
V-SENSE

UNIVERSITY OF OXFORD

CWI
Distributed and Interactive Systems

EPFL

This Talk



Immersive Systems

- New modality to be processed at high quality
- Users’ uncertainty to be well understood

ML-Based methodologies

- The importance of geometrical priors
- Refresh the basics

Immersive Communications: main challenges

How can we achieve a full sense of immersion?
What are the challenges?

- New **spherical/volumetric** content

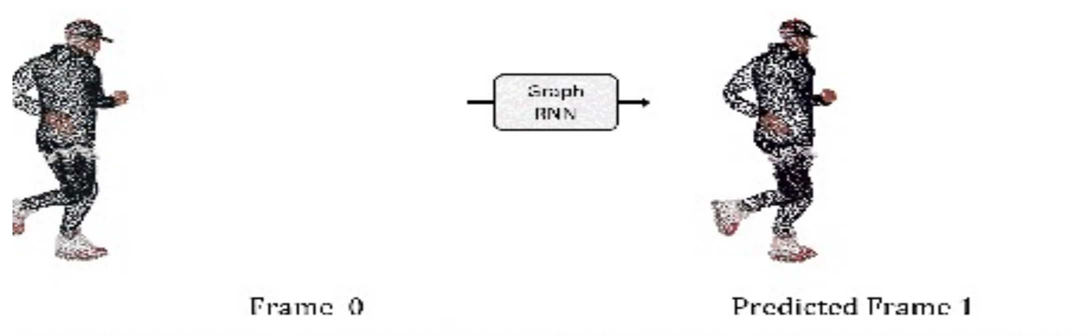
  Topic I: dynamic point cloud processing

- **Large volume** of data to store, deliver and display
- **Ultra-low-delay** constraints over **bandwidth-limited** resources
- **Uncertainty** on users' behaviour

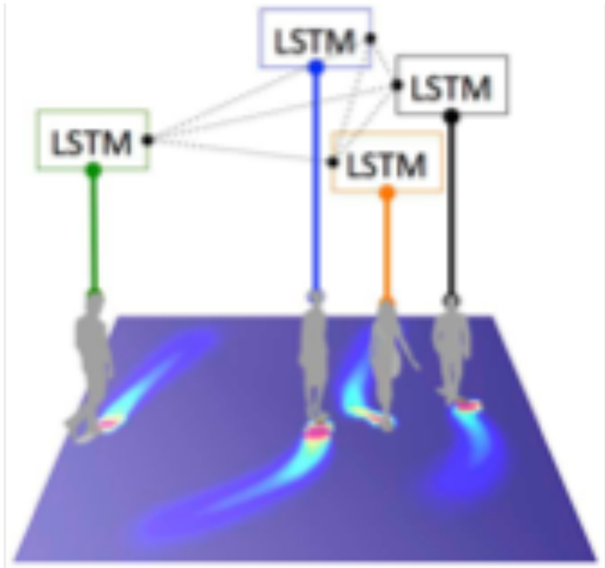
  Topic II: users' behavior in XR

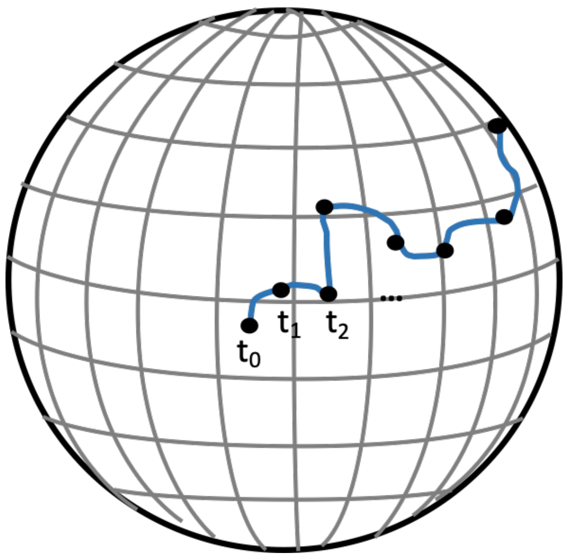
Immersive Communications: main challenges

Topic I: dynamic point cloud processing



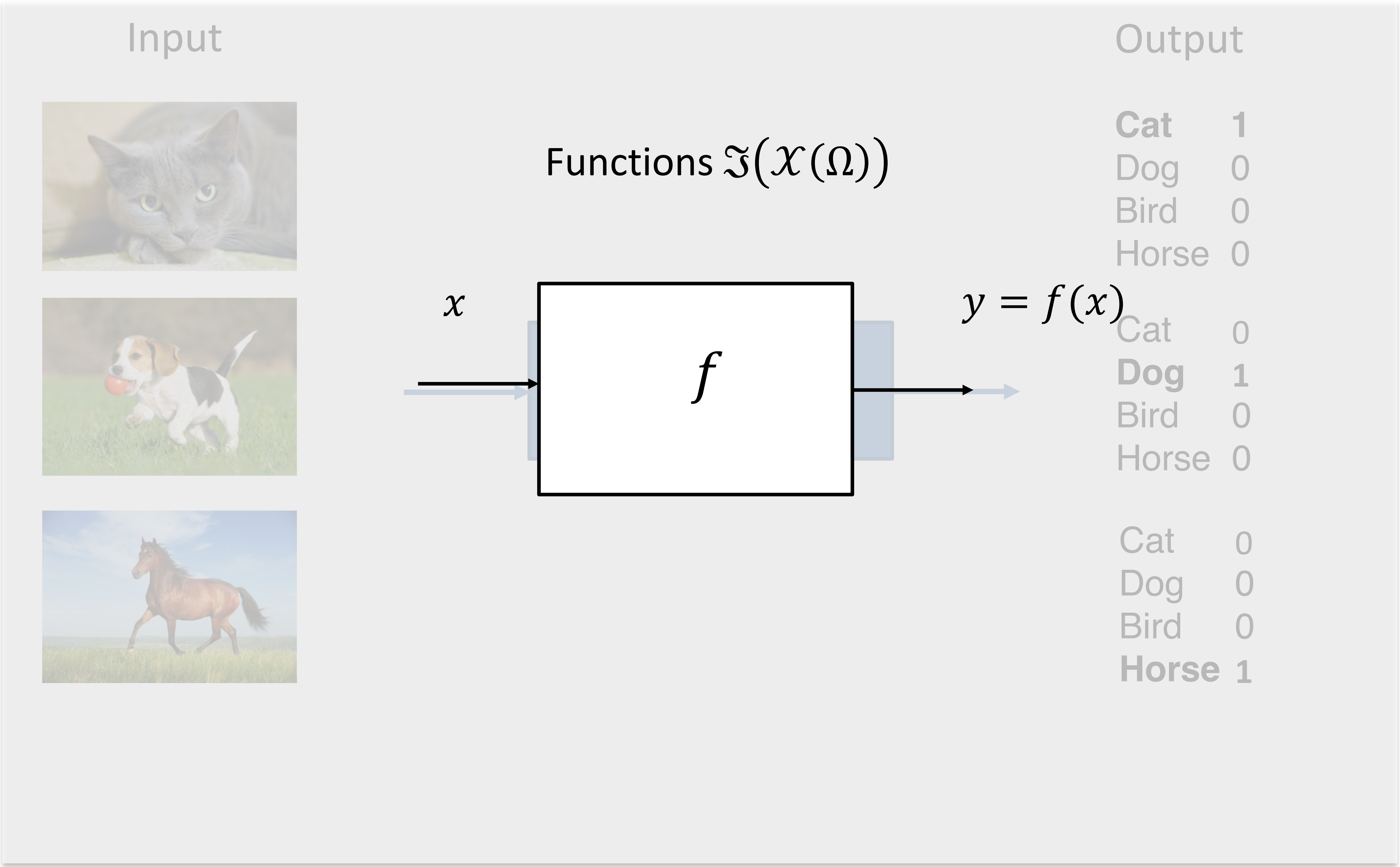
Topic II: users' behavior in XR





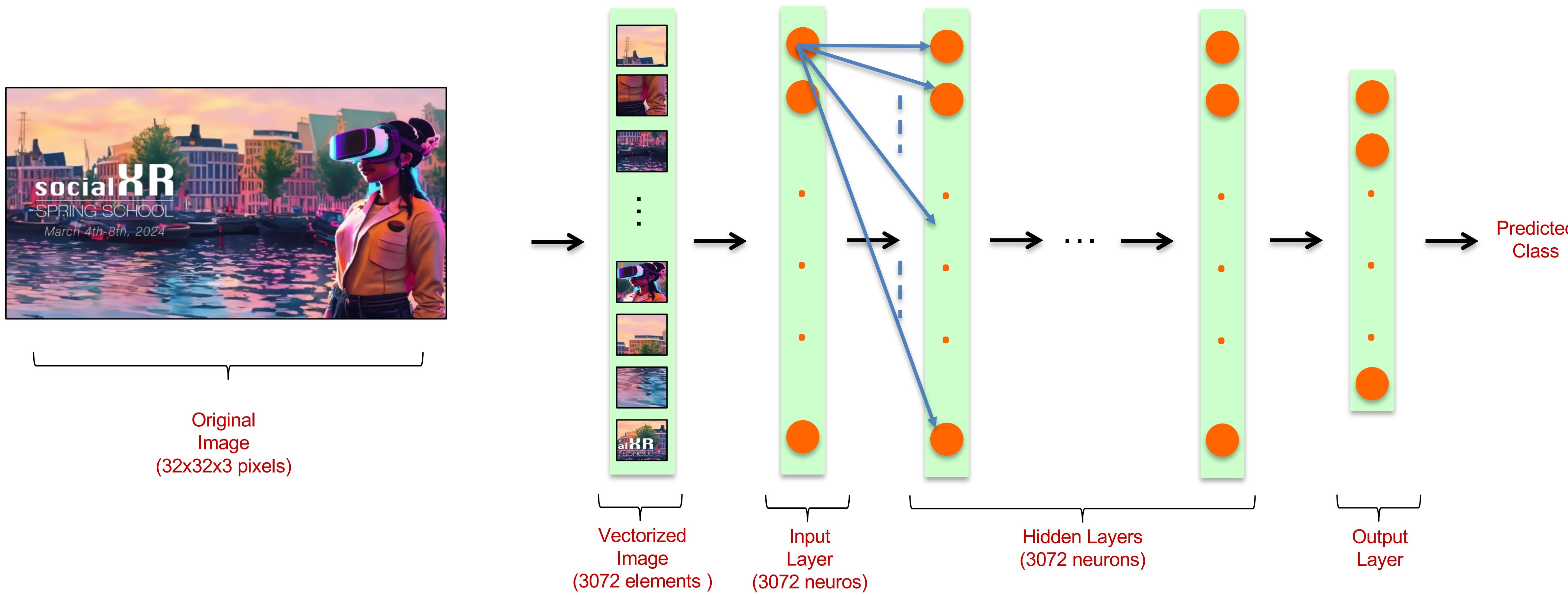
Graph-Based Machine Learning

Deep Learning



Deep Learning

Can we use a standard feed forward neural network
to solve computer vision tasks?

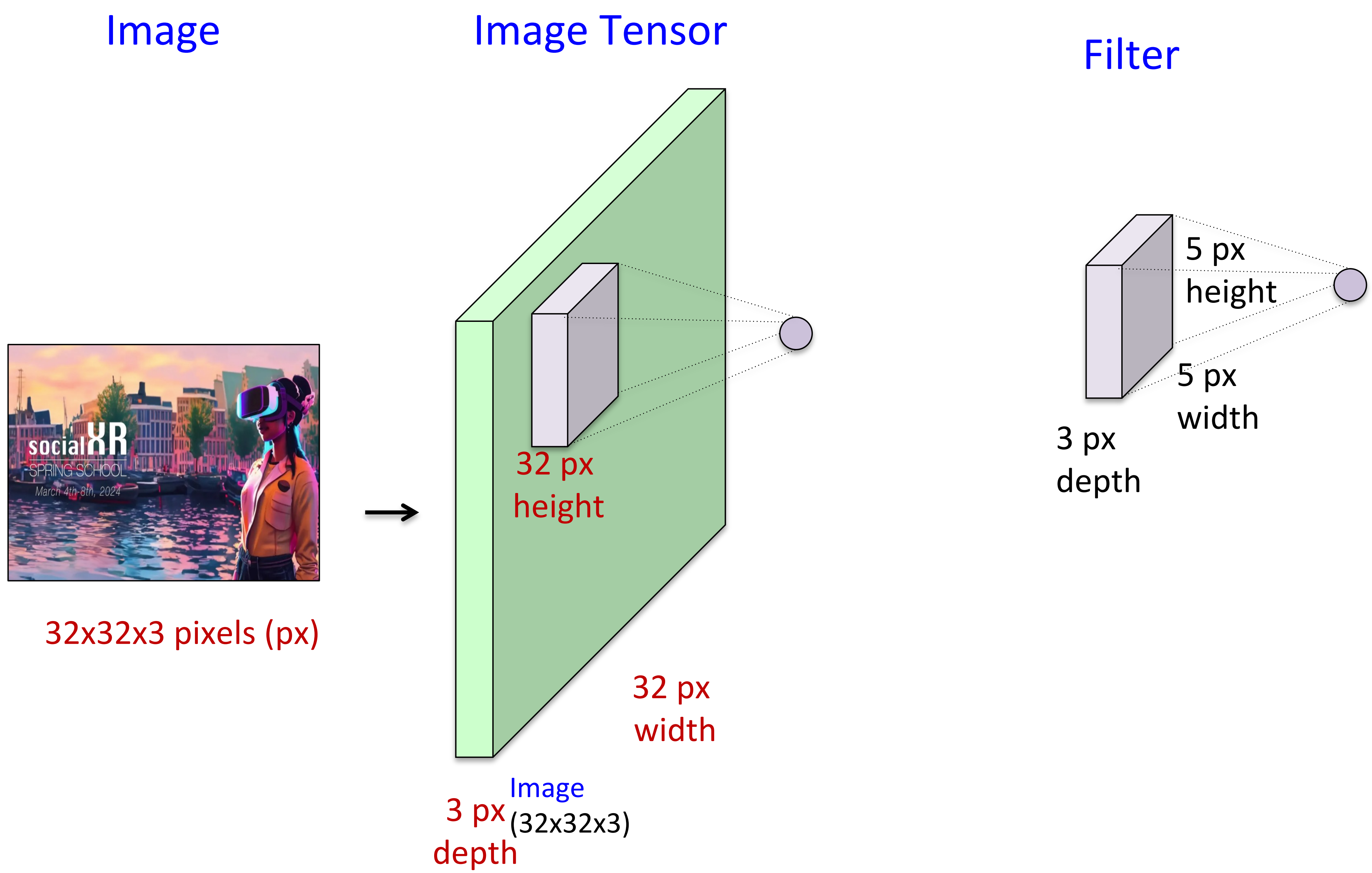


A single neuron is associated with 3072 weights, hence the 3072 neurons are associated with 3072*3072 = 9.437.184 weights for a 32x32x3 picture.

Curse of high dimensionality!

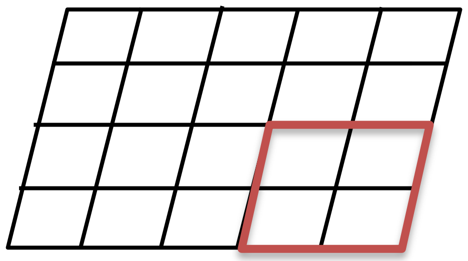
Convolutional Neural Networks

Convolutional neural networks emulate the visual cortex where neurons process information in their **receptive fields**



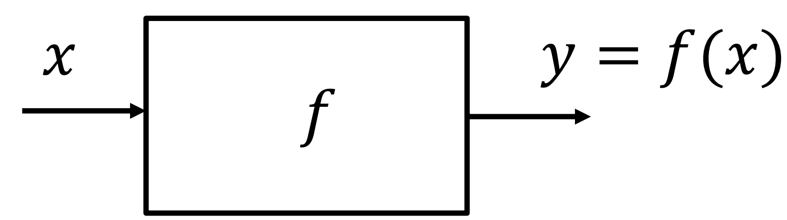
More In General

Domain Ω



Signal space $\mathcal{X}(\Omega)$

Functions $\mathfrak{F}(\mathcal{X}(\Omega))$



Symmetry group
T

Group representation
$\rho(T)$

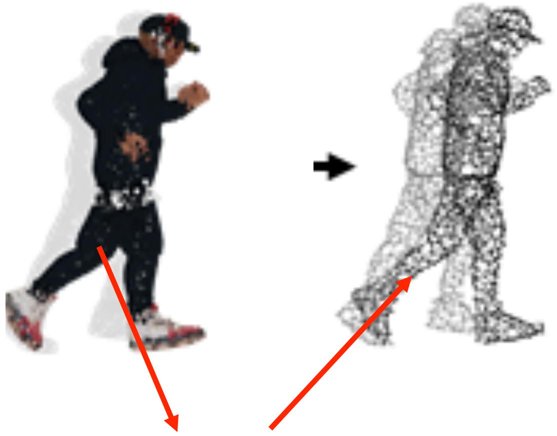
Invariance/Equivariance

The input is defined as a signal living on low-dimensional domain with associated a symmetry group

Can we extend such geometrical priors to less structured domains?

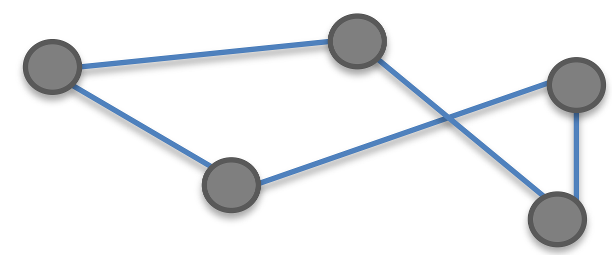
Graphs are appealing tools

Graphs provide mathematical representation for data defined on irregular domains

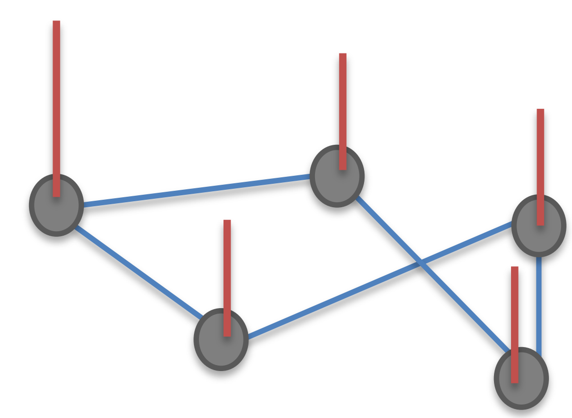


p_i: (x_i, y_i, z_i)

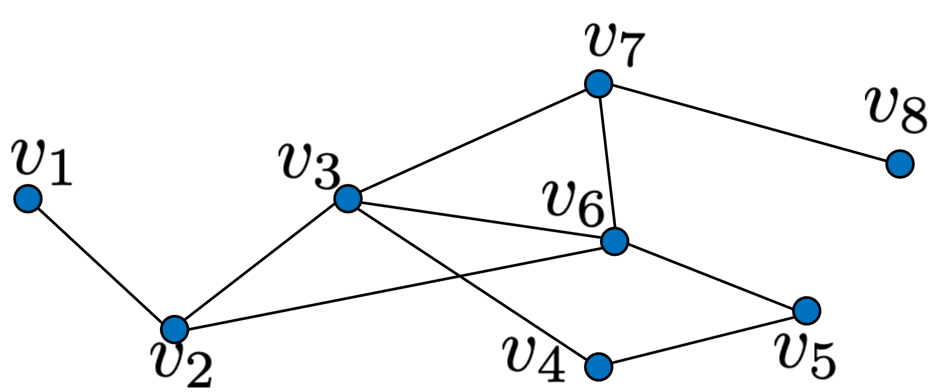
Graphs representation pairwise relations between entities



Graph signal processing allows us to capture both structure (edges) and data (values at vertices)



Graphs and Graph Signal



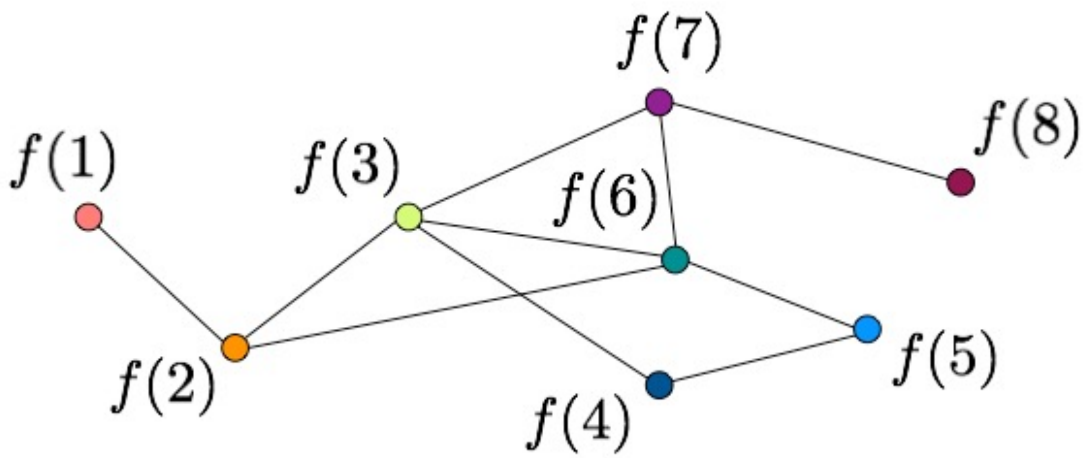
weighted and undirected graph:

$$\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$$

$$D = \mathrm{diag}(d(v_1), \cdots, d(v_N))$$

$$L = D - W$$ **equivalent to G!**

$$L_{\mathrm{norm}} = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}$$



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

Zhou and Schölkopf, “A regularization framework for learning from graph data,” ICML Workshop, 2004.

Graph Laplacian and Filtering

A measure of Smoothness

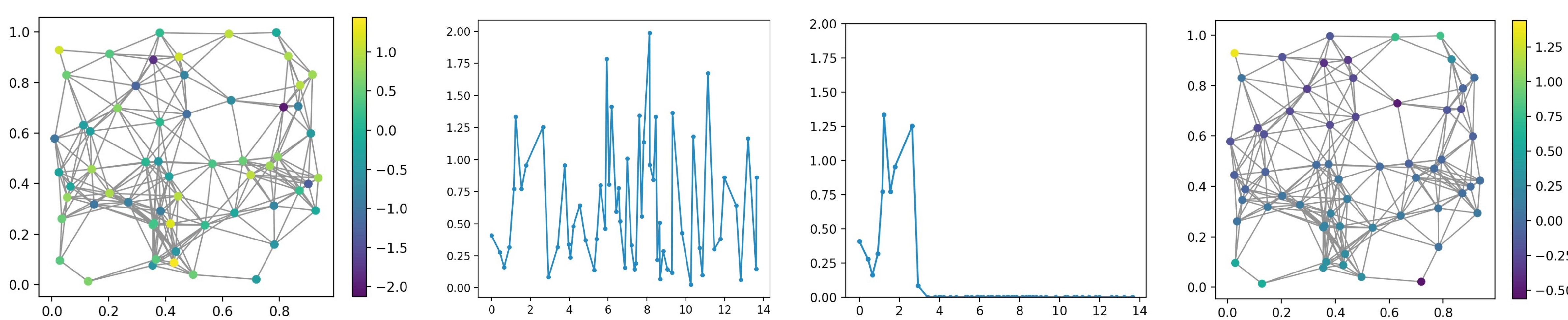
$$f^T L f = \frac{1}{2} \sum_{i,j=1}^{N} W_{ij} \left(f(i) - f(j) \right)^2$$

Graph Fourier Transform

$$\hat{f}(\ell) = \langle \chi_\ell, f \rangle = \sum_{i=1}^{N} \chi_\ell^*(i) f(i)$$

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

f $\xrightarrow{\text{GFT}}$ $\hat{f}(\ell)$ $\xrightarrow{\hat{g}(\lambda_\ell)}$ $\hat{g}(\lambda_\ell)\hat{f}(\ell)$ $\xrightarrow{\text{IGFT}}$ $f(i) = \sum_{\ell=0}^{N-1} \hat{g}(\lambda_\ell)\hat{f}(\ell)\chi_\ell(i)$



UCL

Convolution on graphs

$$f * g = \chi \hat{g}(\Lambda) \chi^T f = \hat{g}(L) f$$

parametric filter as polynomial of Laplacian

$$\hat{g}_\theta(L) = \sum_{j=0}^{K} \theta_j L^j$$

$$K = 1$$
$$\lambda_{N-1} \approx 2$$

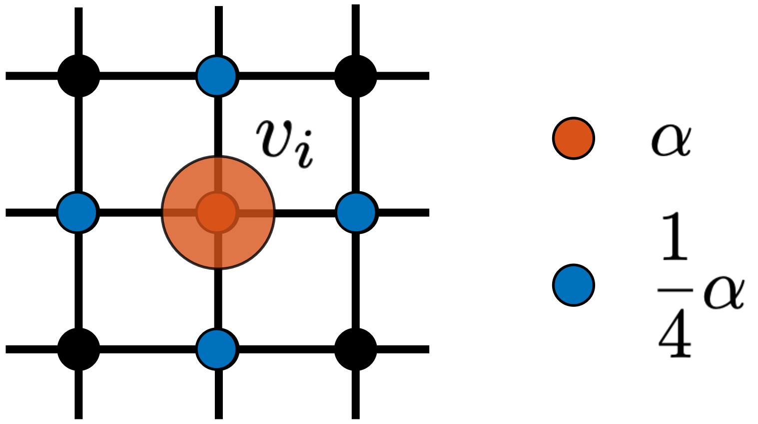
(localisation within1-hop neighbourhood)

$$= \theta_0 I - \theta_1 (D^{-\frac{1}{2}} W D^{-\frac{1}{2}})$$

$$\alpha = \theta_0 = -\theta_1$$

$$\boxed{\hat{g}_\theta(L) = \alpha(I + D^{-\frac{1}{2}} W D^{-\frac{1}{2}})}$$

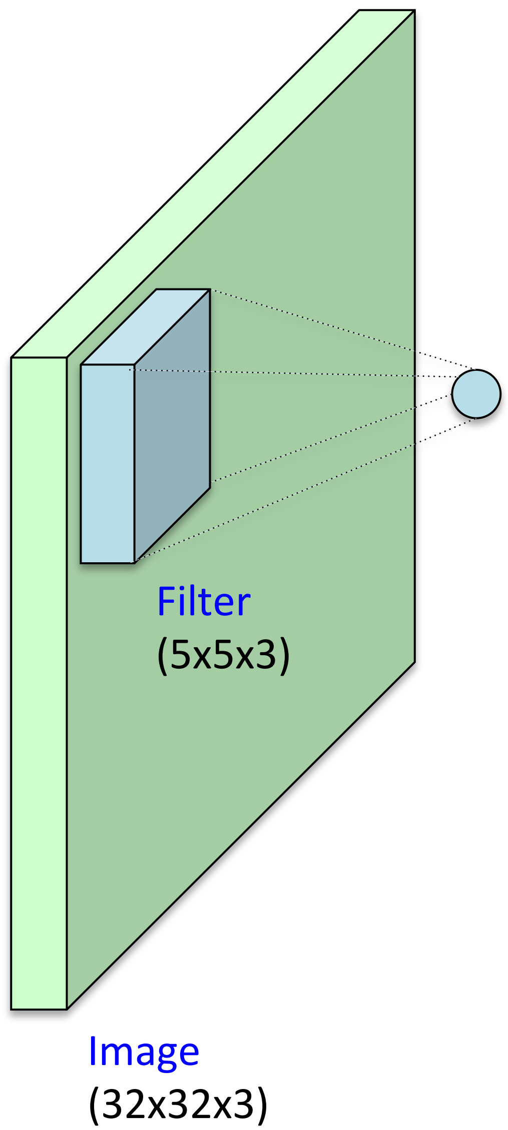
Spectral filtering can be seen as operation on the vertex domain!

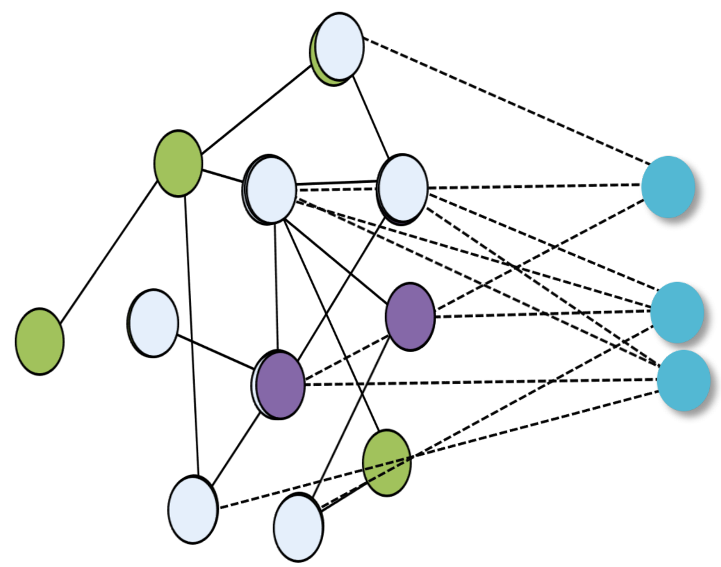


$$y_i = \alpha f_i + \frac{1}{4} \alpha \sum_{j:(i,j) \in \mathcal{E}} f_j$$

Defferrard et al., "Convolutional neural networks on graphs with fast localized spectral filtering," NIPS, 2016.

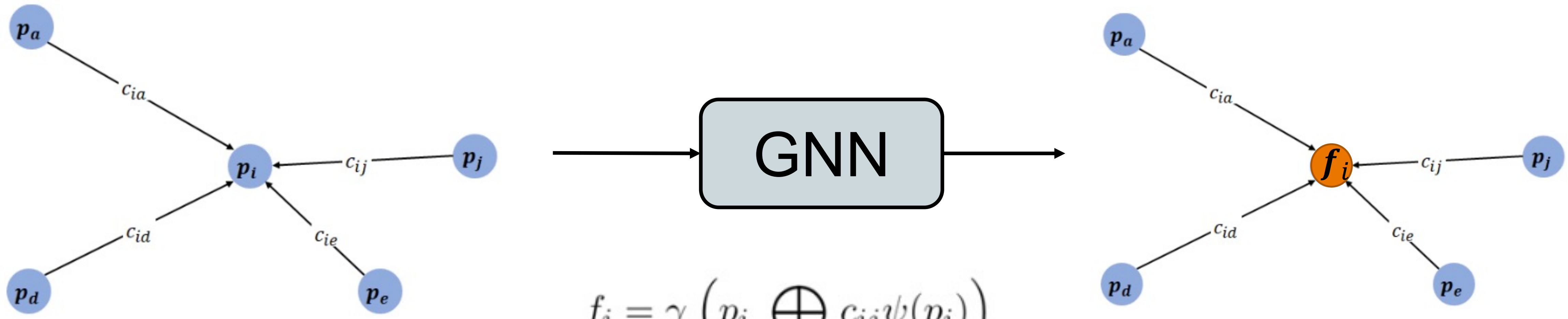
From Grids to Graphs





The Rise of Geometric Deep Learning

Graph Neural Networks



$$f_i = \gamma\left(p_i \bigoplus_{j \in N_i} c_{ij}\psi(p_i)\right)$$

- Similar to a normal convolution a graph convolution aggregates a neighborhood.
- Each node is processed independently, and the outputs aggregated to generate node feature

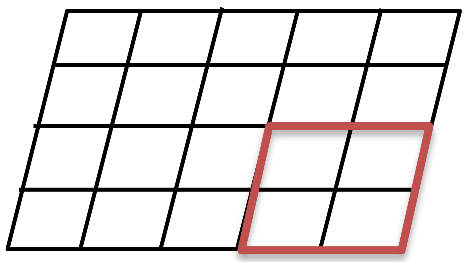
Representation Learning:

- The GNN learn a point representation (feature) $\boldsymbol{f}_i$ describing the node $\boldsymbol{p}_i$ local neighborhood

- By using edges, the GNN takes structure into account while learning new representations.

Deep Learning … on graphs

Domain Ω

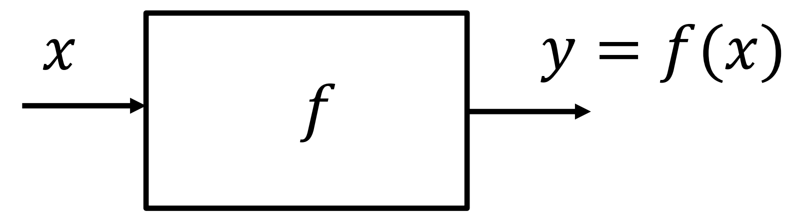


Symmetry group
T

Signal space $\mathcal{X}(\Omega)$

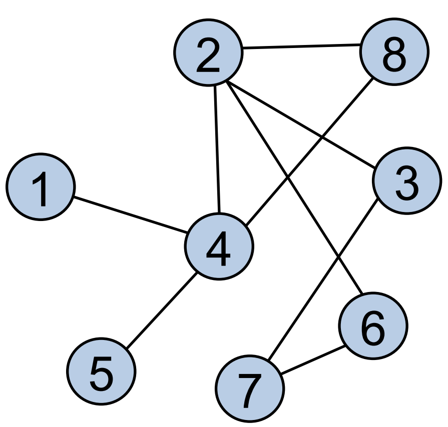
Group representation
$\rho(T)$

Functions $\mathfrak{I}(\mathcal{X}(\Omega))$

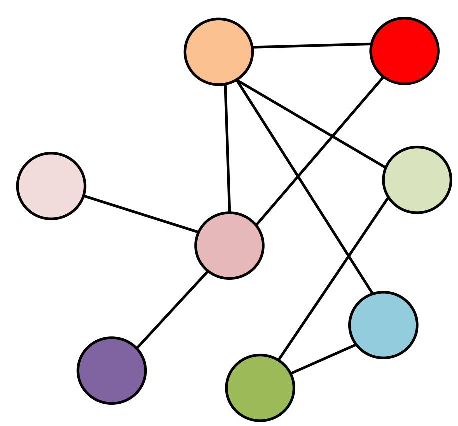


Invariance/Equivariance

Graph G: (V, E)

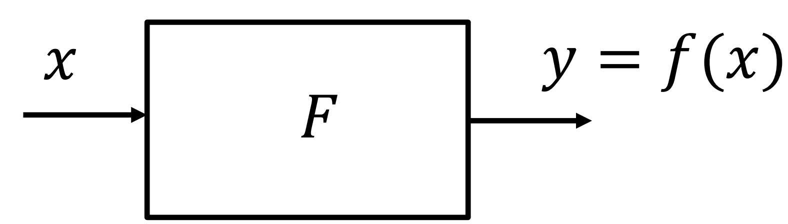


Node Features $\mathcal{X}(G)$



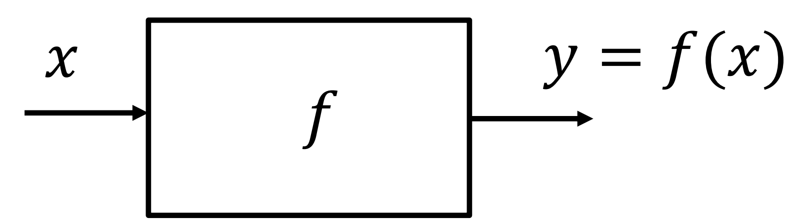
Permutation Matrix **P**

Functions $\mathfrak{I}(\mathcal{X}(G))$

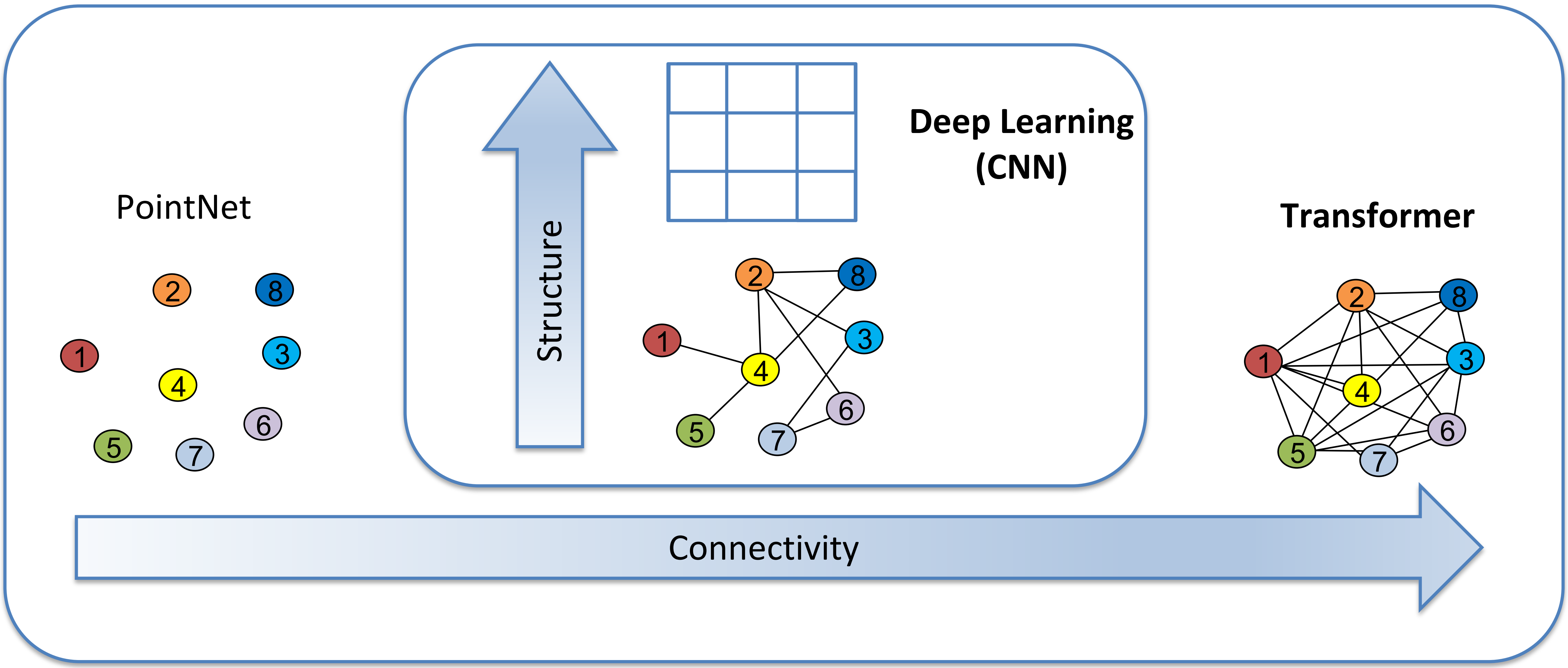


Message Passing
F(**PX**,**PA**$\boldsymbol{P^T}$) = **P**F(**X**,**A**)

Deep Learning …on graphs



Input x

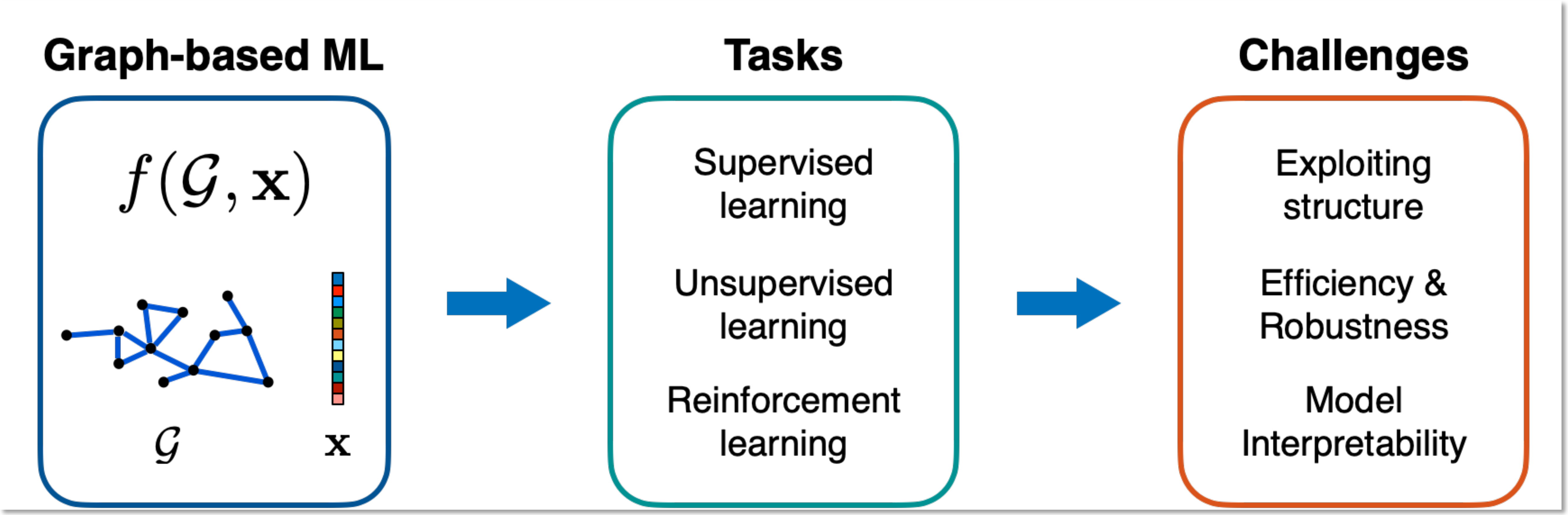


Graph Regression / Classification

Node Classification

Link Prediction

Taxonomy



Dong, X., Thanou, D., **Toni, L.,** Bronstein, M., & Frossard, P., “Graph signal processing for machine learning”, IEEE SPM 2020

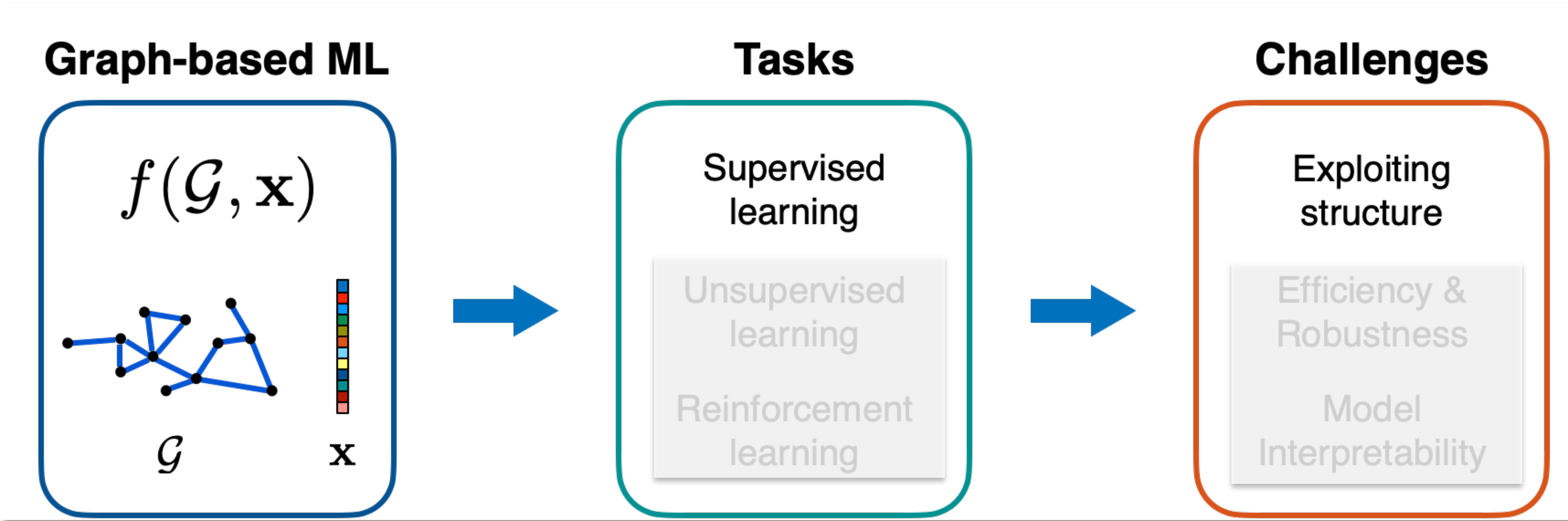
Take Home Message

- Graphs as versatile tools to model data on irregular domains
- Frequency analysis led to key insights of today GNNs
- It is essential to exploit geometrical priors when processing data
- Graph-based ML is very wide – generative models, regression models, etc.

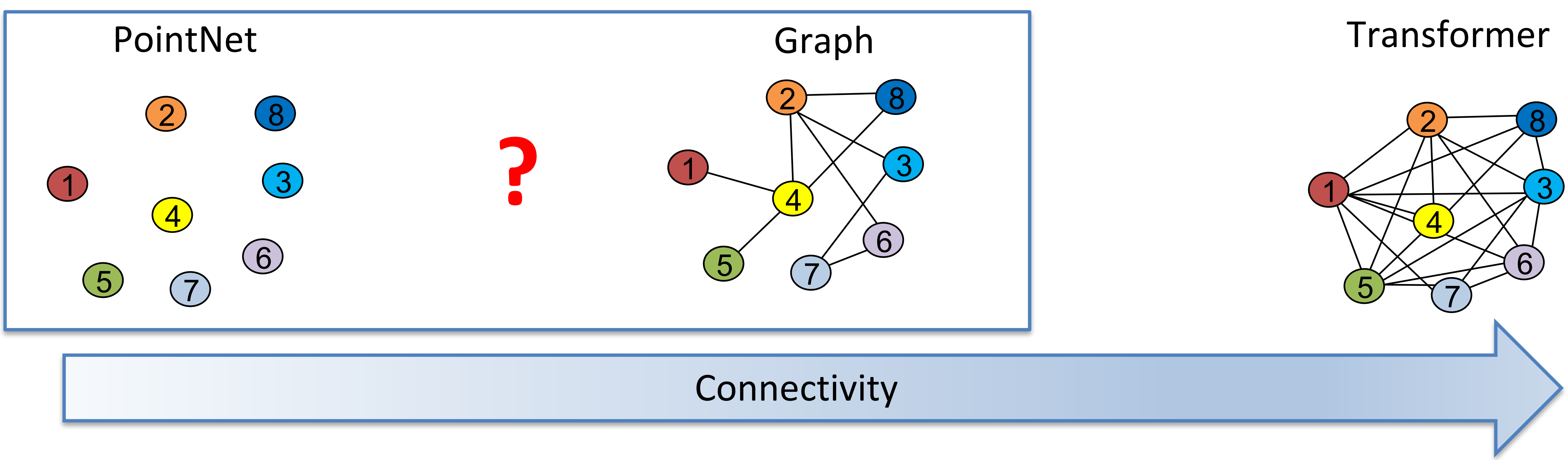
What Graph-Based ML in Social-XR?

Graph-Based Models for **Dynamic** Point Cloud **Prediction**

Taxonomy

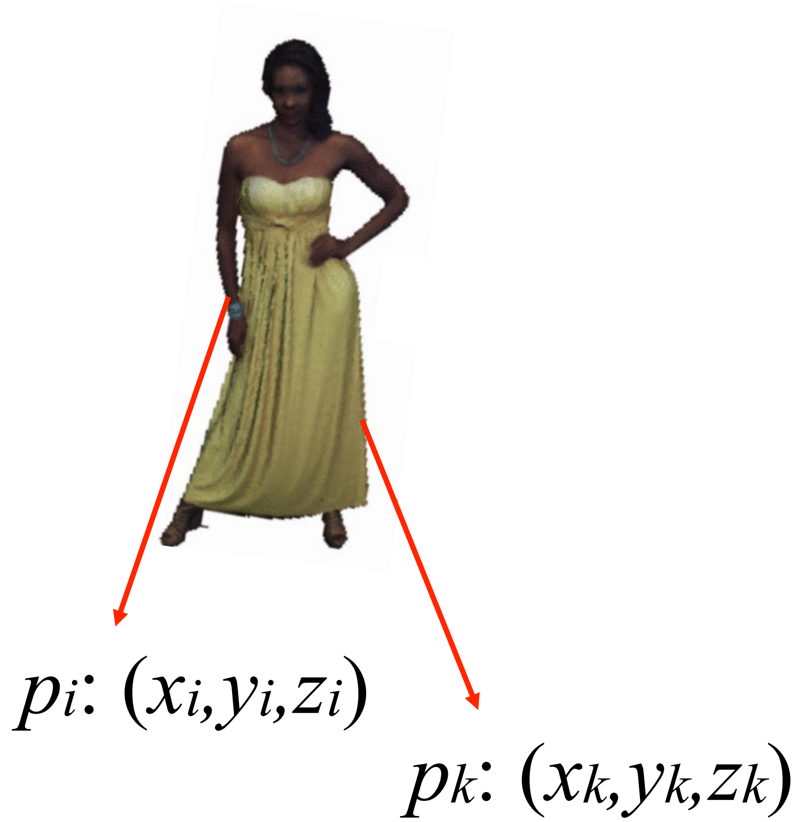


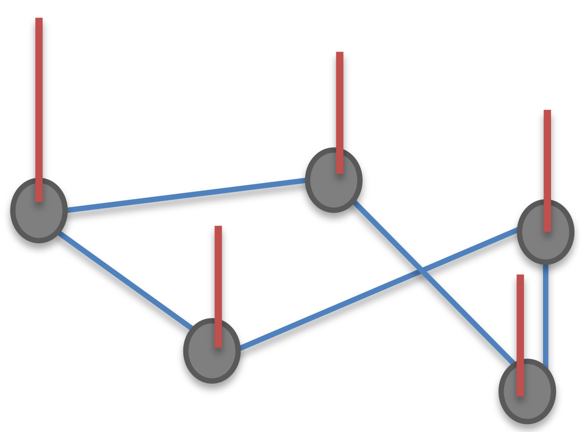
Graph-based ML
Tasks
Challenges
$f(\mathcal{G}, \mathbf{x})$
$\mathcal{G}$
$\mathbf{x}$
Supervised learning
Unsupervised learning
Reinforcement learning
Exploiting structure
Efficiency & Robustness
Model Interpretability



PointNet
?
Graph
Transformer
Connectivity

Point Cloud Signal



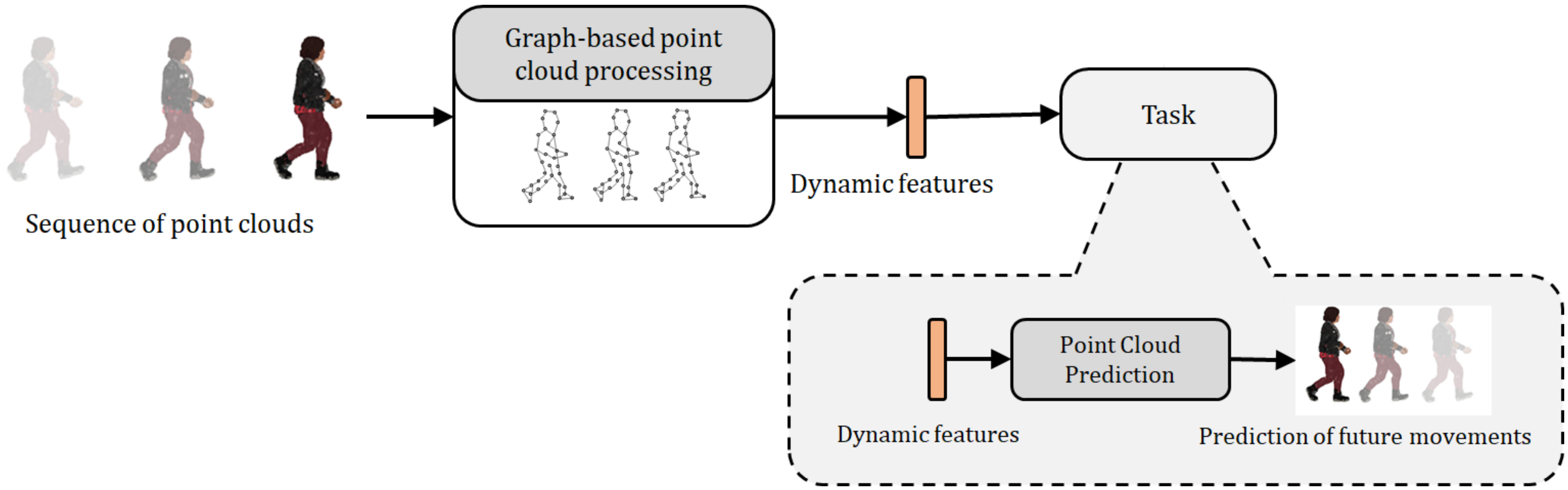


A point cloud is a set of points in space representing a 3D scene

- Direct output from Lidar sensors
- Flexible
- High-spatial resolution without discretization
- Un-ordered set of points
- No point to point correspondence across point clouds

[Image Credit] D Thanou, PA Chou, P Frossard,"Graph-based compression of dynamic 3D point cloud sequences", IEEE Transactions on Image Processing, 2016

Point Cloud Prediction: Goal

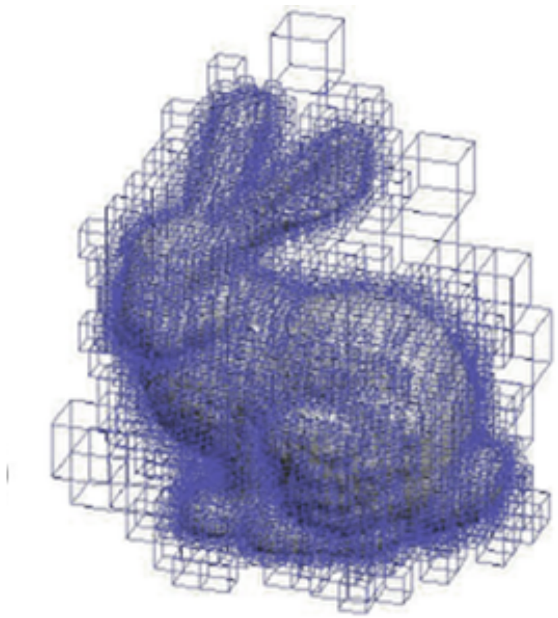


Given a set of sequential point clouds, predict (short/long) term future point clouds

- How do we represent **dynamic** point clouds?
- Why it is important?
- Can graphs help us?

A Point Based Approach

Voxel based

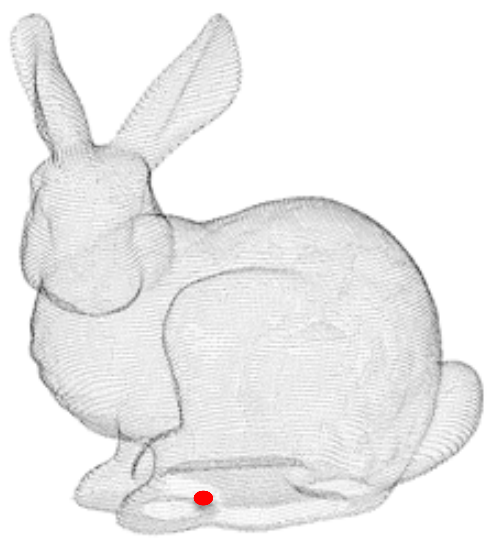


CNN

Graph based

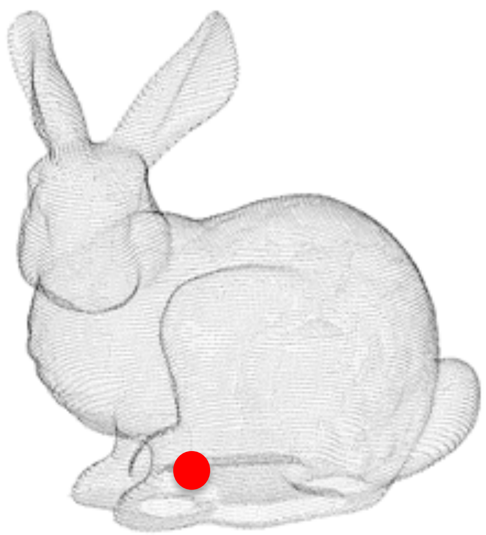
GNN/GCN

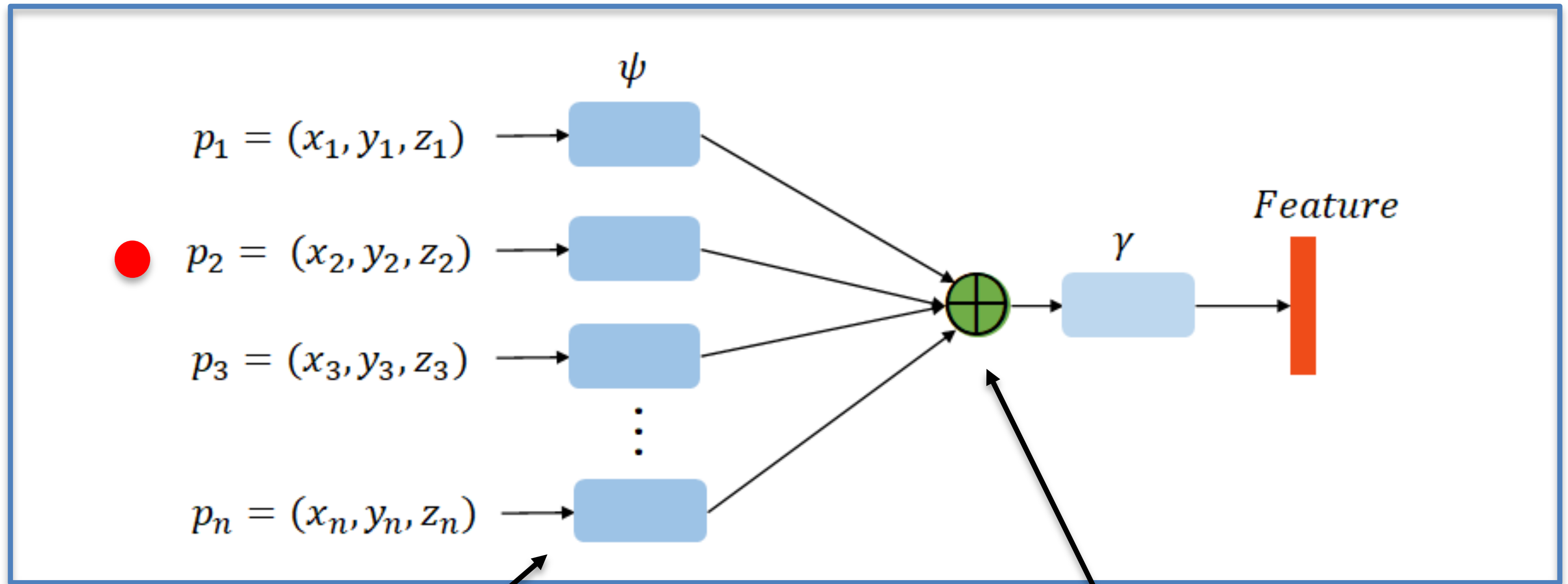
Point based



PointNet

Point-Based Learning





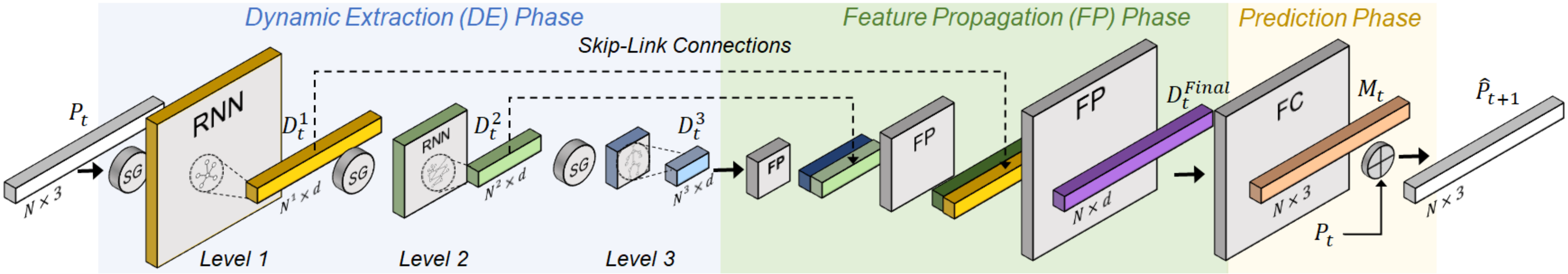
Each point processed **independently**

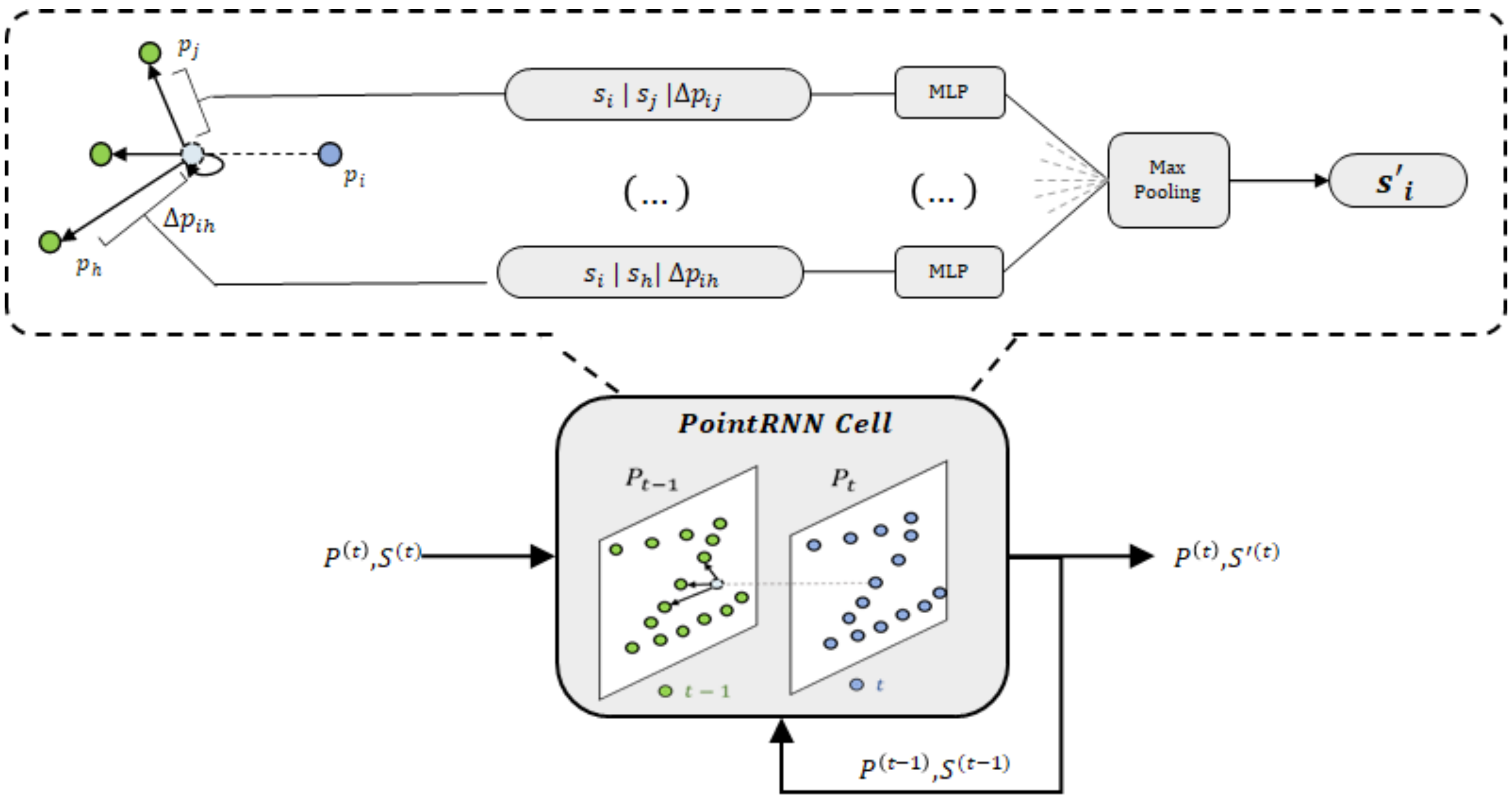
Aggregation function (max pooling) invariant to the order of the input

$$f(P) = \gamma\left(\bigoplus_{i \in P} \psi(p_i)\right)$$

Qi, Charles R., et al. "Pointnet: Deep learning on point sets for 3d classification and segmentation." CVPR 2017

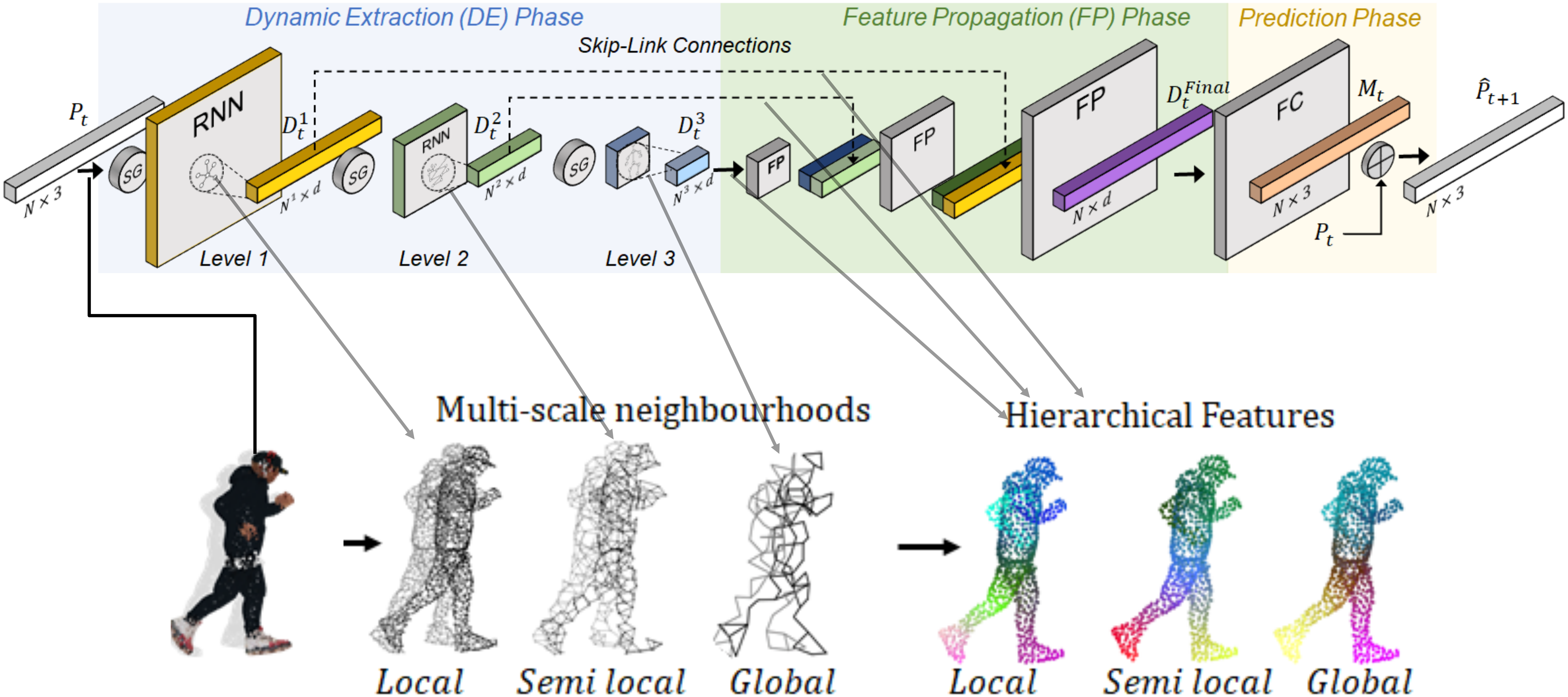
What Does Exist in the Literature?





Guangming Wang, Xinrui Wu, Zhe Liu, and Hesheng Wang. 2021. Hierarchical Attention Learning of Scene Flow in 3d Point Clouds. IEEE Transactions on Image Processing (2021).
Hehe Fan, Yi Yang, and Mohan Kankanhalli. 2022. Point Spatio-temporal Transformer Networks for Point Cloud Video Modelling. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022).
Hehe Fan and Yi Yang. 2019. PointRNN: Point Recurrent Neural Network for Moving Point Cloud Processing. arXiv preprint arXiv:1910.08287 (2019).

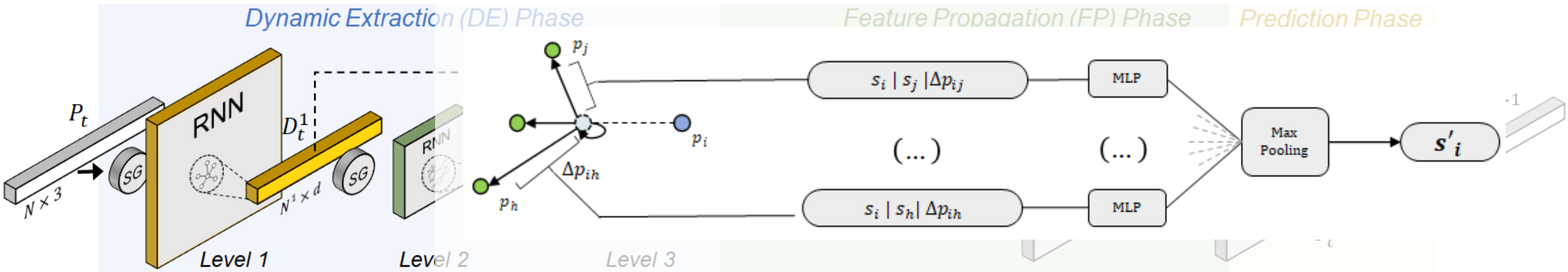
Dynamic Extraction Phase



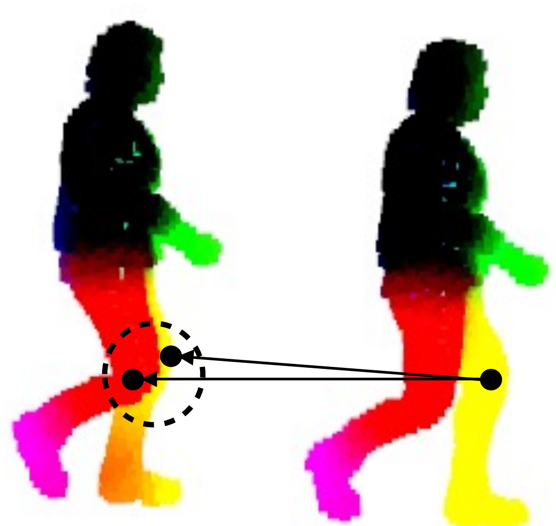
Current works learn hierarchical features, what are the main limitations?

Challenges in Processing Deformable Shapes

Limitation 1: Lack of structural relationship between points in point cloud prediction



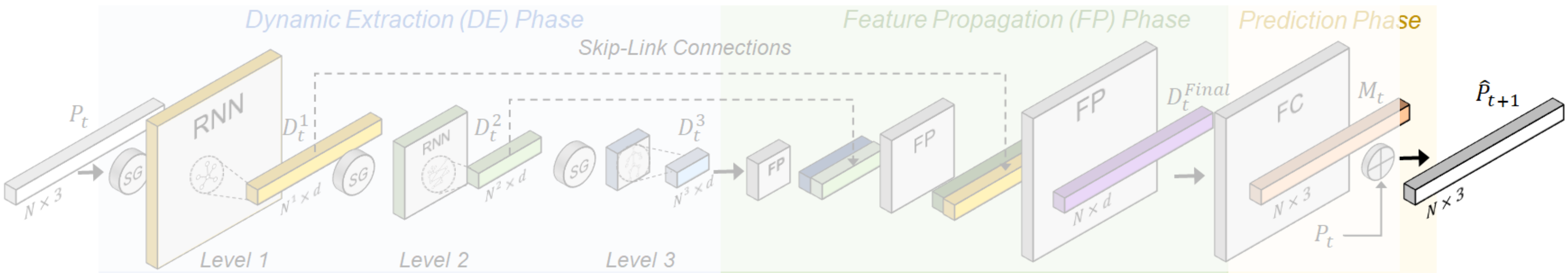
- Point to point correspondence



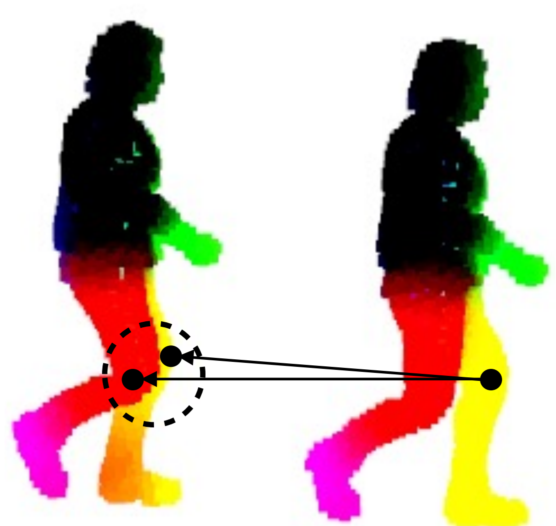
Geometrical proximity is misleading

Challenges in Processing Deformable Shapes

Limitation 1: Lack of structural relationship between points in point cloud prediction



- Point to point correspondence

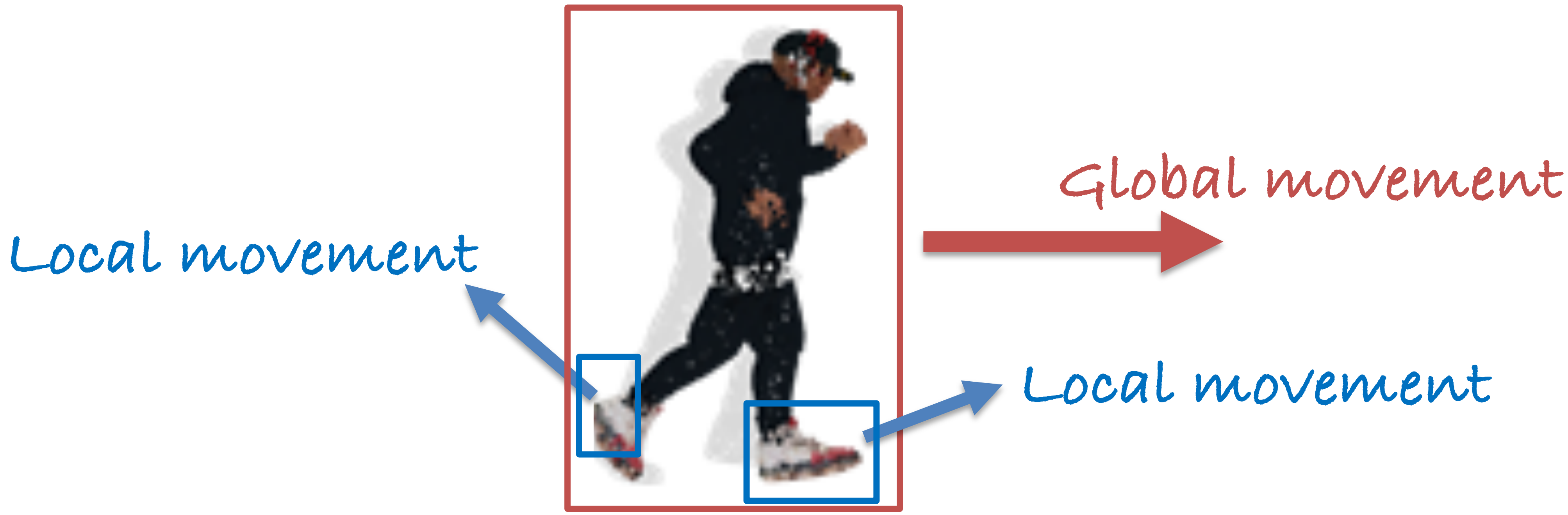


Geometrical proximity is misleading

- Shape Deformation

UCL

Challenges in Processing Deformable Shapes

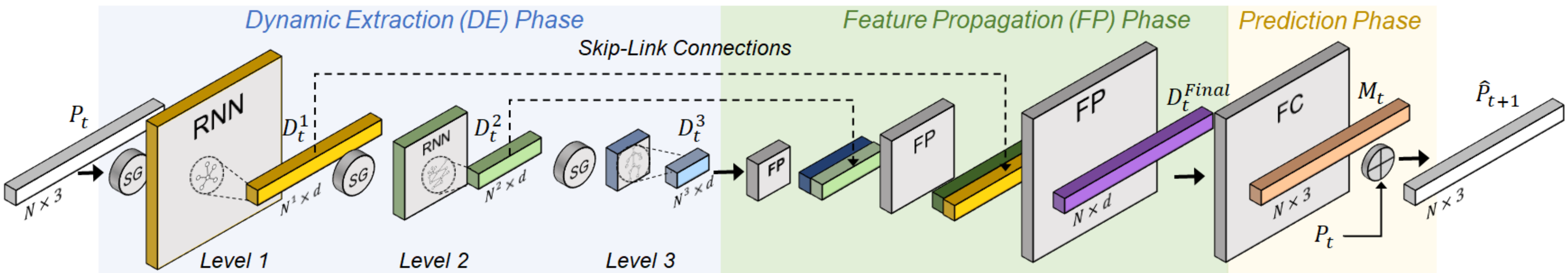


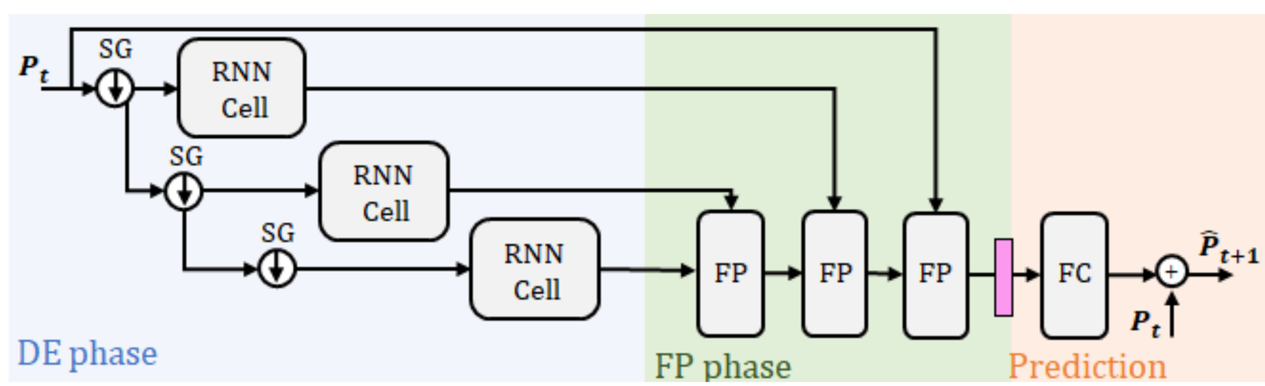
Challenges in Processing **Complex Motion**:
??

<u>Opening the black-box:</u>

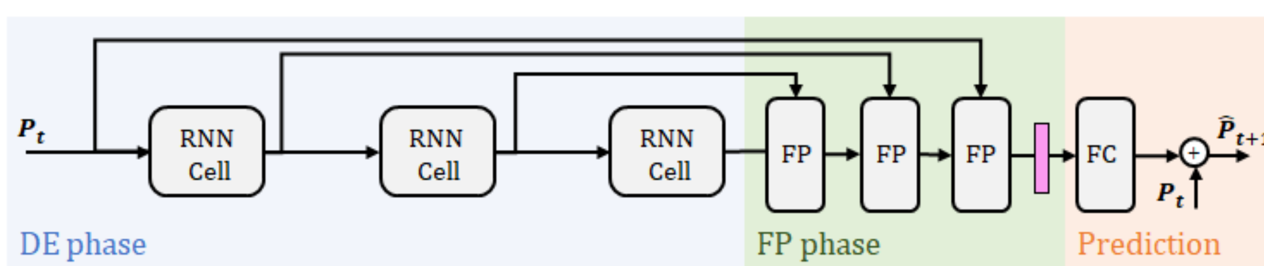
- Disentangling effects
- MVs interpretation

Understanding Learning Hierarchical Features

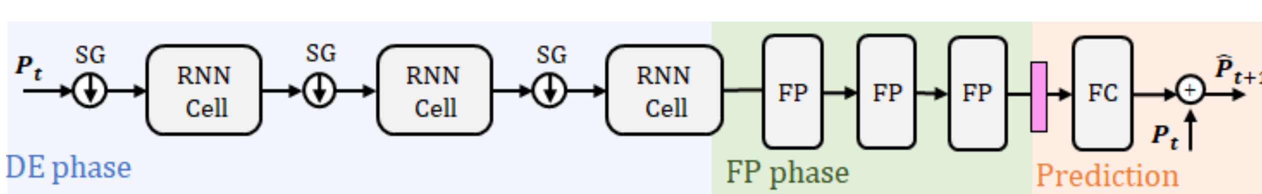




(a) *Shallow* hierarchical architecture



(b) *Single-scale* hierarchical architecture



(c) *Without-combination* hierarchical architecture

Is **Stacked** effect the key?

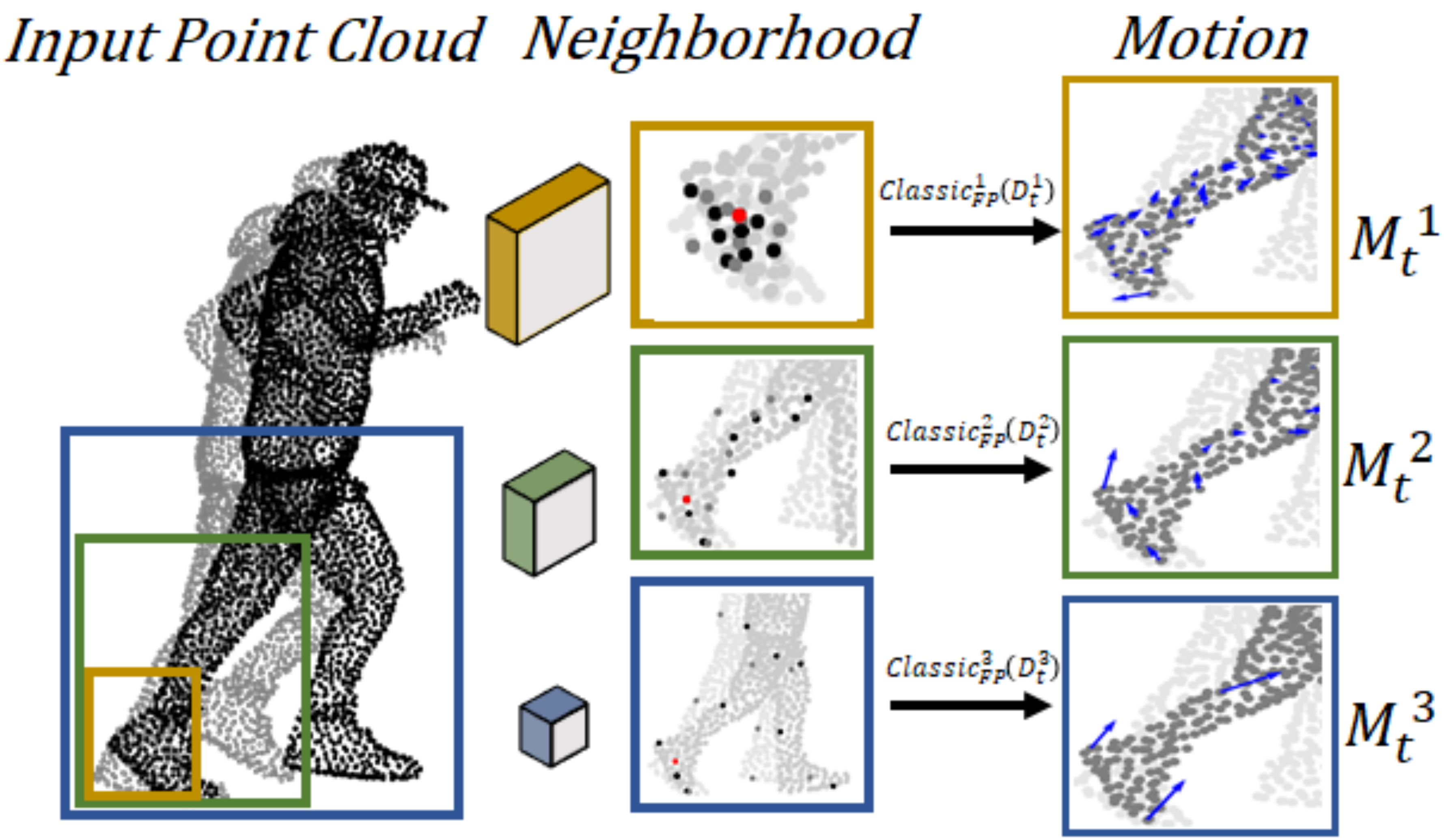
Is **Multi-Scale** effect the key?

Is **Skip-Link connection** the key?

Pedro Gomes, Silvia Rossi, and **Laura Toni,** "Explaining Hierarchical Features in Dynamic Point Cloud Processing", IEEE PCS 2022.

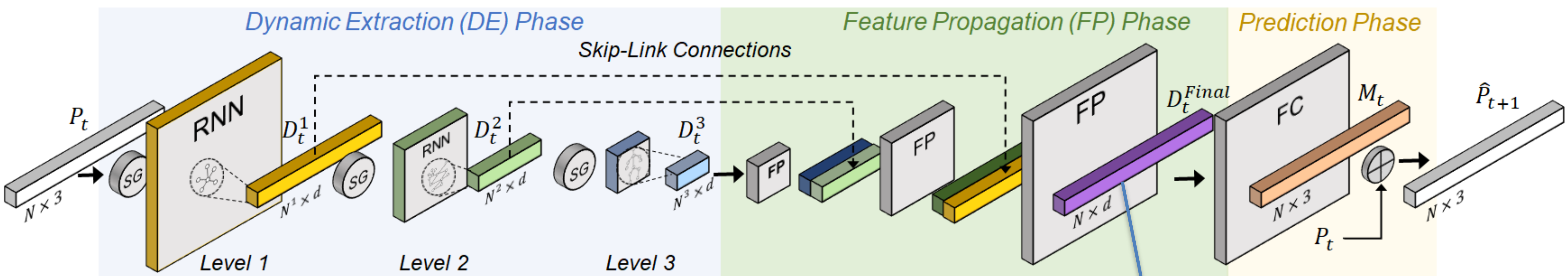
An "Interpretability" Perspective

Hierarchical features learning is essential, but is it good enough?



Challenges in Processing Deformable Shapes

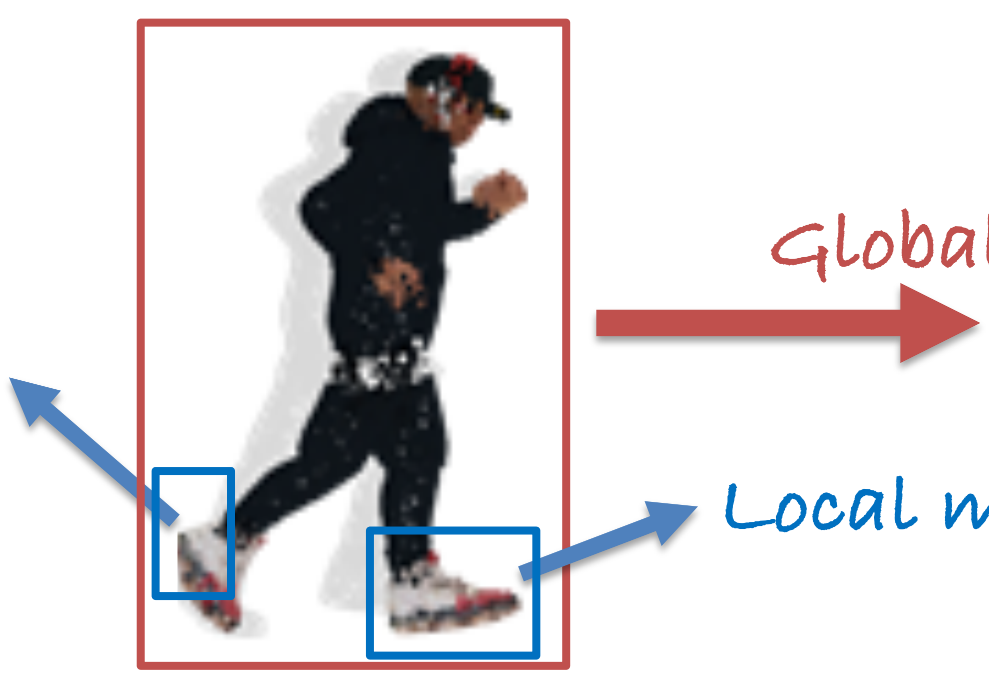
Limitation 2: the fixed combination of hierarchical features in the prediction phase



$$D_t^{l^{\mathrm{FP}}} = \mathrm{ReLU}\left(\Theta_{FP}^{l}\left\{D_t^l; \tilde{D}_t^l\right\}\right)$$

Interpolated features

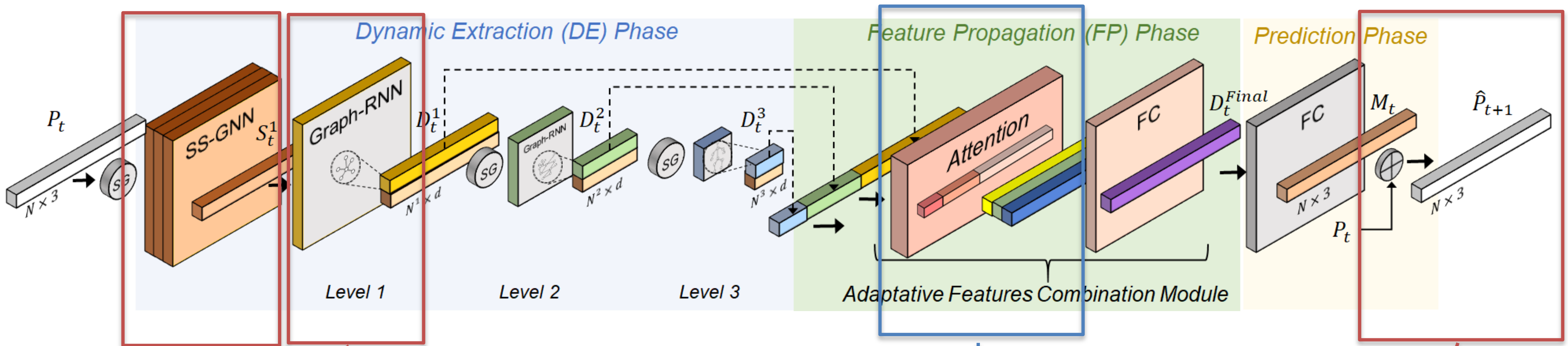
Shared (among points) learnable weights



Global movement

Local movement

AGAR: Attention Graph-RNN for Adaptative Motion Prediction of Point Clouds of Deformable Objects



To learn **spatial features** that carry local **topological** information

Attention module to assign an attention value to each level based on the learned features

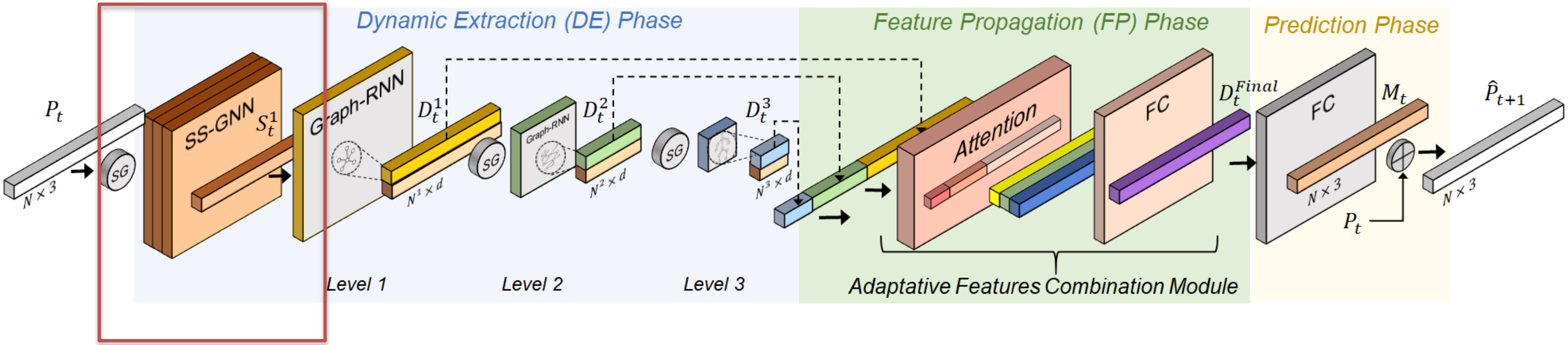
spatio-temporal graph constructed to incorporate the point structural/semantic information and build **representative neighbourhoods** of points

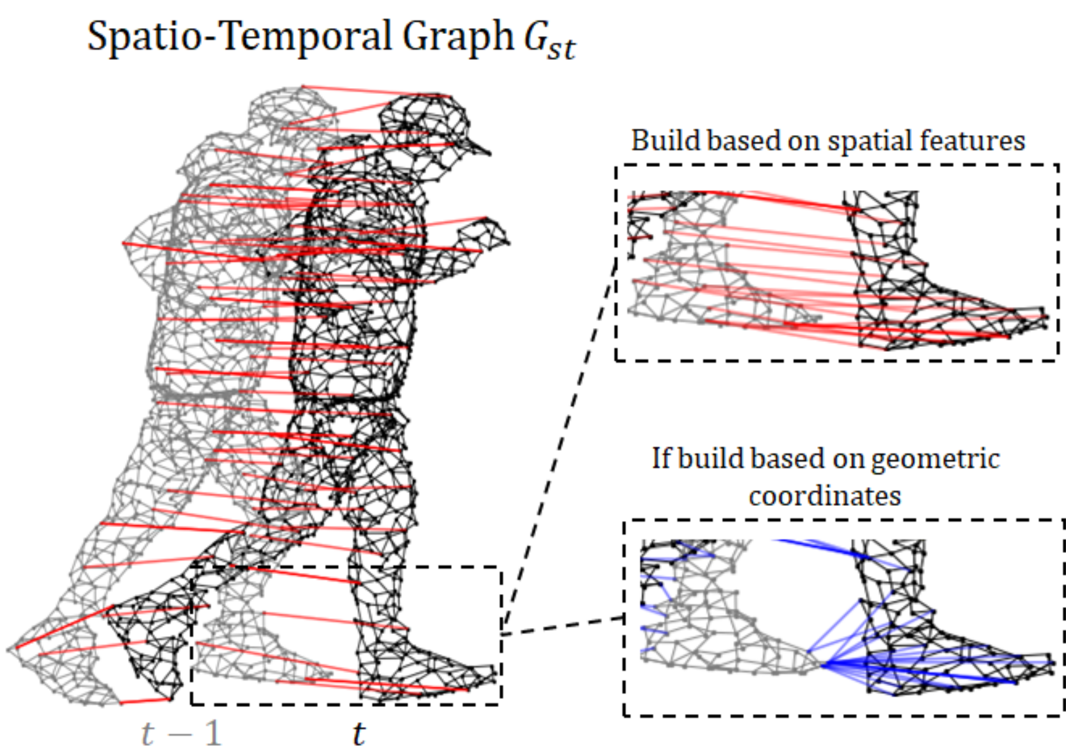
Graph-based regularizer as auxiliary loss

$$\gamma p^T L p$$

P. Gomes, S. Rossi, **L. Toni**, "AGAR: Attention Graph-RNN for Adaptative Motion Prediction of Point Clouds of Deformable Objects", submitted ACM TOMM, Arxiv 2023

Spatial-Structure GNN





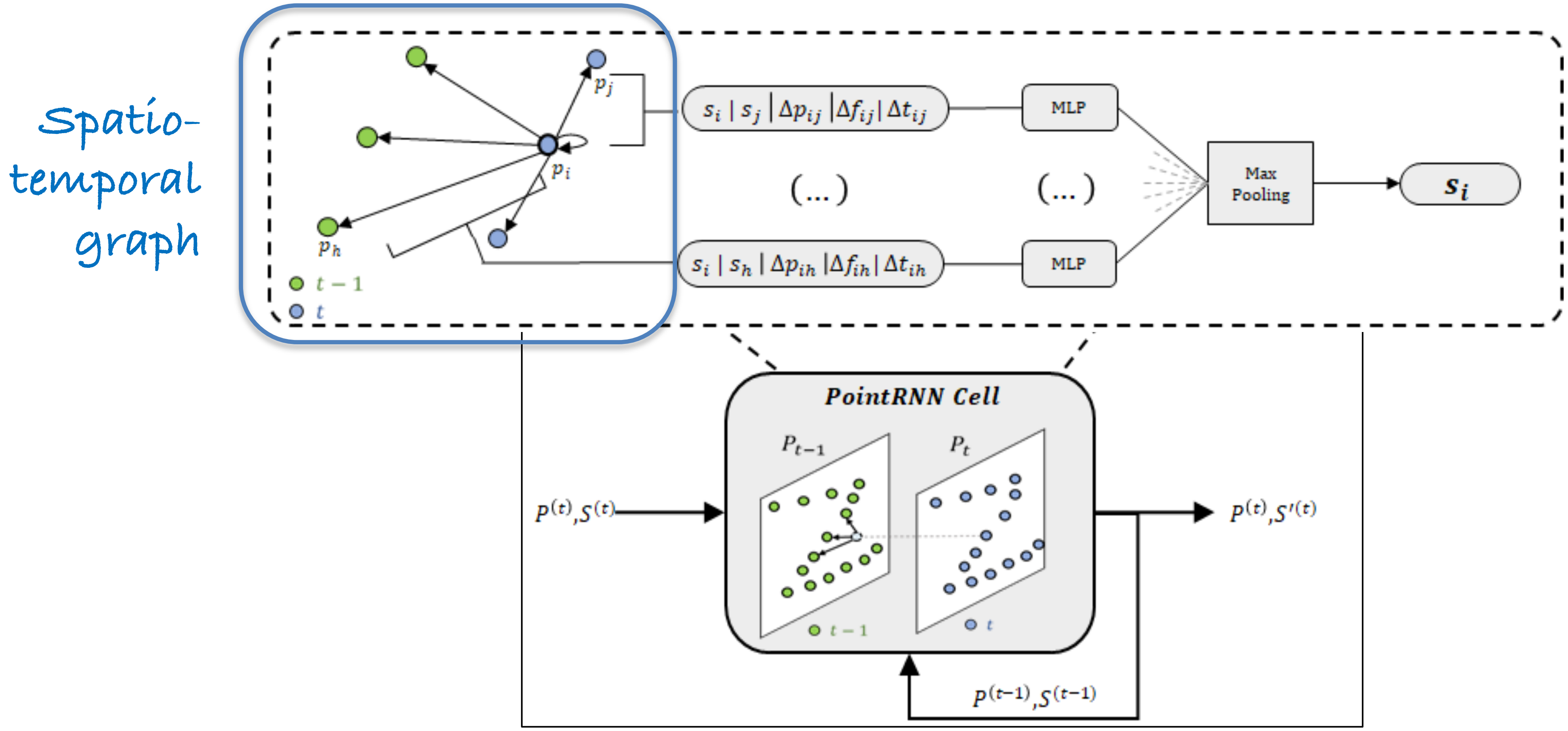
- K-NN graph built based on geometry (coordinates)
- SS-GNN composed of 3 layers, each performing a graph message-passing convolution

$$m_{ij,t}^{h} = \Theta_S^h(s_{i,t}^{h-1}\ ; p_{i,t}^{l}\ ; \Delta p_{ij})$$

$$s_{i,t}^{h} = \bigoplus_{j \in \mathcal{E}_i^C} \{m_{ij,t}^{h+1}\}$$

P. Gomes, S. Rossi, **L. Toni**, "AGAR: Attention Graph-RNN for Adaptative Motion Prediction of Point Clouds of Deformable Objects", submitted ACM TOMM, Arxiv 2023

Graph Based RNN

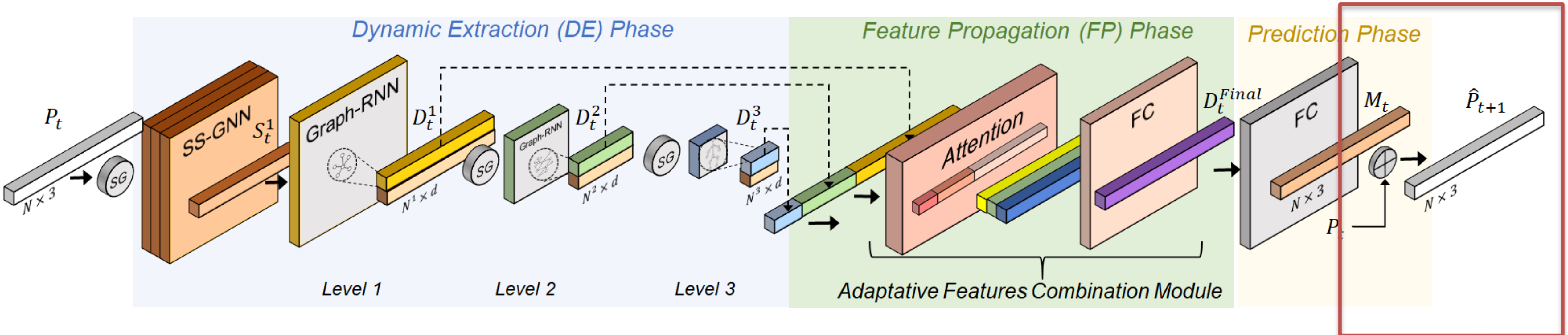


The graph-RNN extracts dynamic features by performing a message-passing convolution between a point and its neighbourhoods in the **spatio-temporal graph**

$$m^l_{ij,t} = \Theta^l_D(d^l_{i,t};\ d^l_{j,t'};\ \Delta p_{ij};\ \Delta s_{ij};;\Delta t_{ij})$$

$$d^{l+1}_{i,t} = \bigoplus_{j \in \mathcal{E}^{ST}_i} \{m^l_{ij,t}\}$$

Laplacian-Regularizer as Auxiliary Loss



Training Loss + **Auxiliary Loss**

$$d_{CD}(P,\hat{P}) = \frac{1}{n}\sum_{p\in P}\min_{\hat{p}\in\hat{P}}||p-\hat{p}||^2 + \frac{1}{n}\sum_{p\in\hat{P}}\min_{p\in P}||\hat{p}-p||^2$$

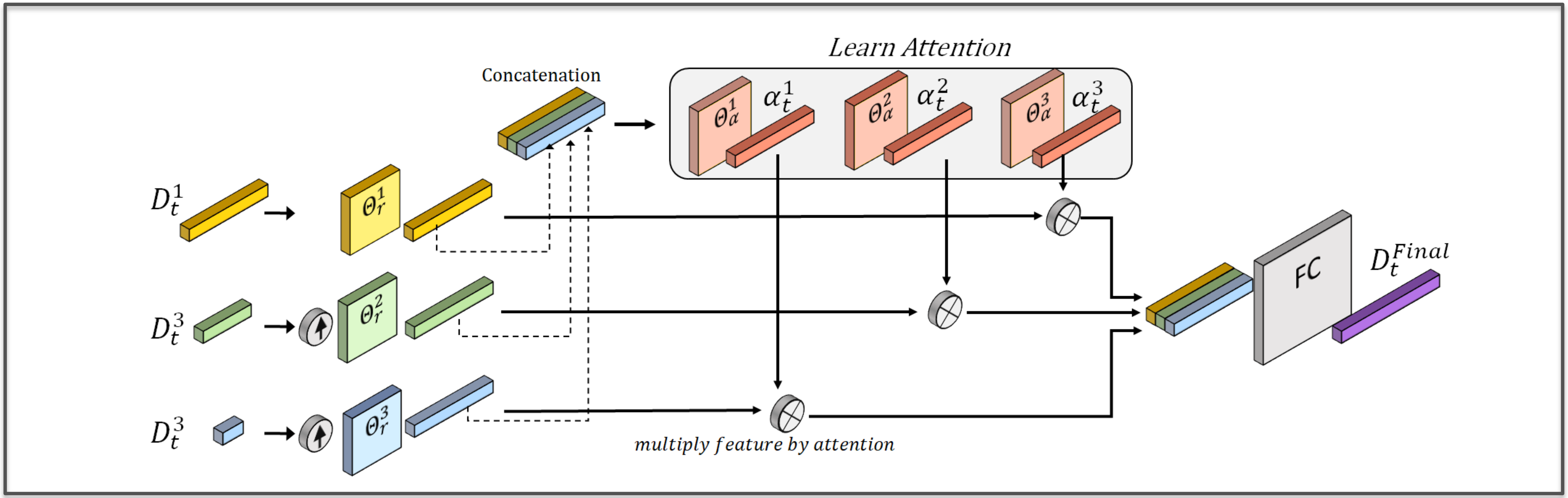
$$d_{EMD}(P,\hat{P}) = \min_{\theta:P\longrightarrow\hat{P}}\sum_{p\in P}||p-\theta(p)||^2.$$

graph denoising

$$\gamma p^T Lp$$

P. Gomes, S. Rossi, **L. Toni**, “AGAR: Attention Graph-RNN for Adaptative Motion Prediction of Point Clouds of Deformable Objects”, submitted ACM TOMM, Arxiv 2023

Attention Mechanism

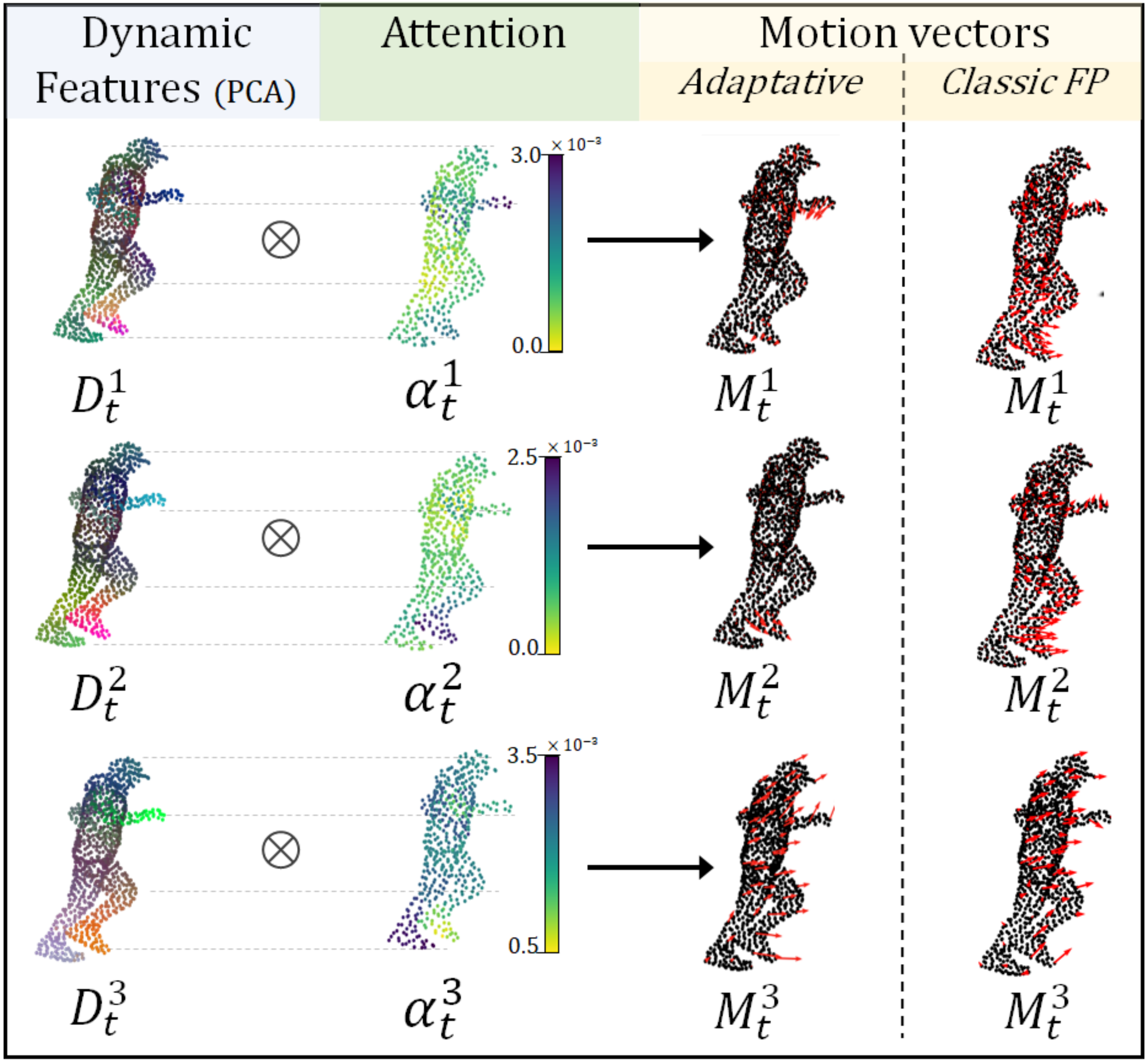


$$\alpha_{i,t}^l = \sigma\left(\Theta_\alpha^l\{\psi(d_{i,t}^1);\psi(d_{\tilde{i},t}^2);\psi(d_{\tilde{i},t}^3)\}\right)$$

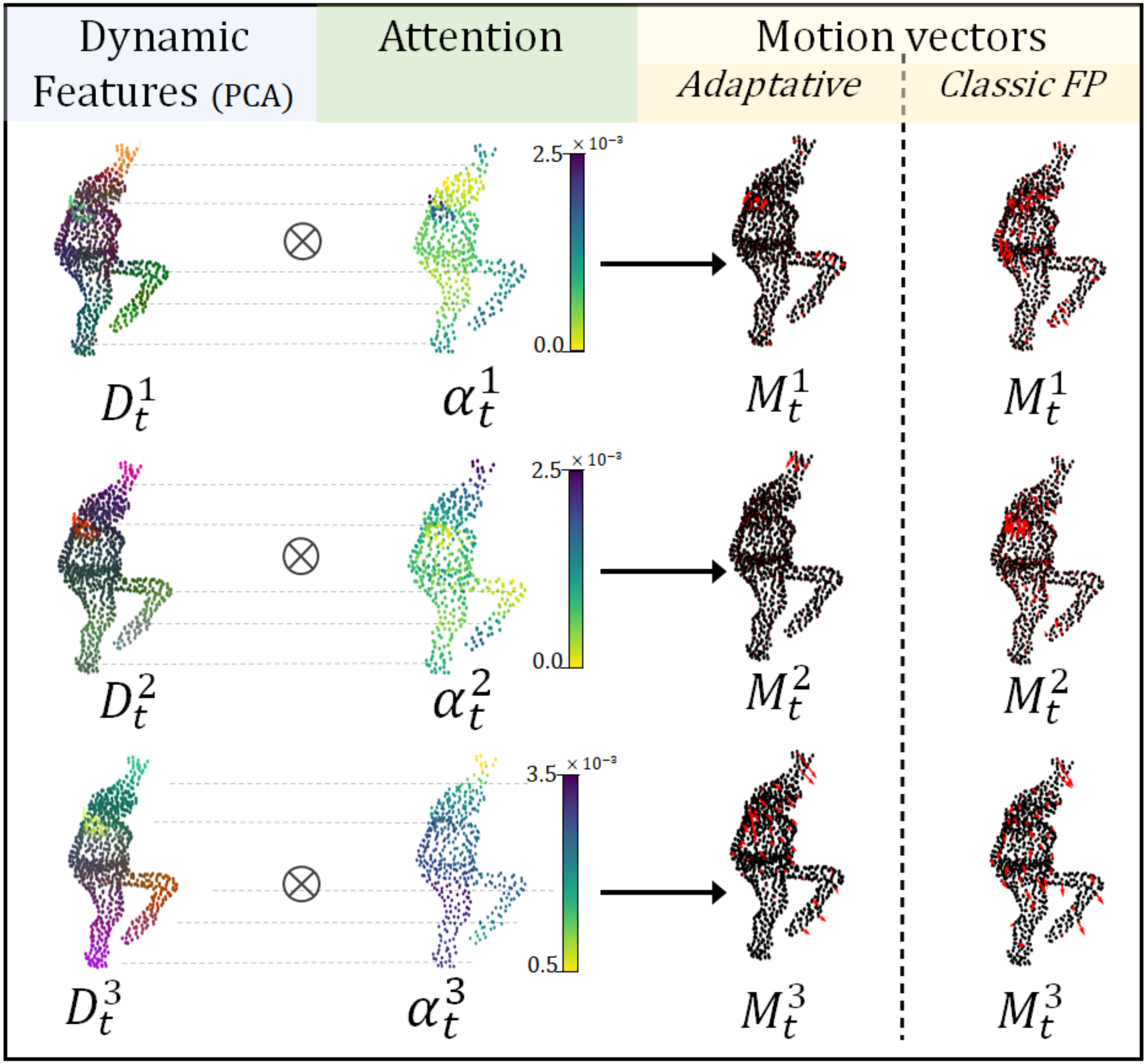
$$d_{i,t}^{Final} = \sigma\left(\Theta_{FC}\{\Psi(d_{i,t}^1);\Psi(d_{\tilde{i},t}^2);\Psi(d_{\tilde{i},t}^3)\}\right)$$

$$\Psi(d_{i,t}^l) = \psi(d_{i,t}^l) \times \alpha_{\tilde{i},t}^l$$

Local and Global MVs



Local and Global MVs



Results

Mixamo (Synthetic Human bodies dataset)				
Model		CD	EMD	CD Top 5%
Copy-Last-input		0.1056	123.4	0.2691
PointPWC-Net-*pred* [43]		0.09358	118.5	0.2601
FlowStep3D-*pred* [15]		0.09153	115.6	0.2575
PSTNet-*pred* [8]		0.08984	114.1	0.2556
PointRNN [6]		0.00351	68.0	0.1593
AGAR	*Classic-FP*	0.00262	59.6	0.1412
	Adaptative	**0.00254**	**58.2**	**0.1346**

JPEG and CWIPC-SXR Real-world human bodies dataset							
Method		JPEG			CWIPC-SXR		
		CD	EMD	CD Top 5%	CD	EMD	CD Top 5%
Copy Last Input		0.00118	42.0	0.09001	0.00295	43.2	0.12915
PointRNN		0.00109	41.3	0.083461	0.00157	43.4	0.10973
AGAR	*Classic-FP*	0.00101	38.6	0.08172	**0.00150**	40.8	**0.10655**
	Adaptative	**0.00095**	**37.4**	**0.07754**	0.00155	**39.8**	0.10760

Take Home Message

- Deep learning models are **black boxes that can be opened**
- Graphs can help in
  - Building meaningful neighborhoods
  - Learning ST features with geometry information embedded

- **Can graphs help further?**

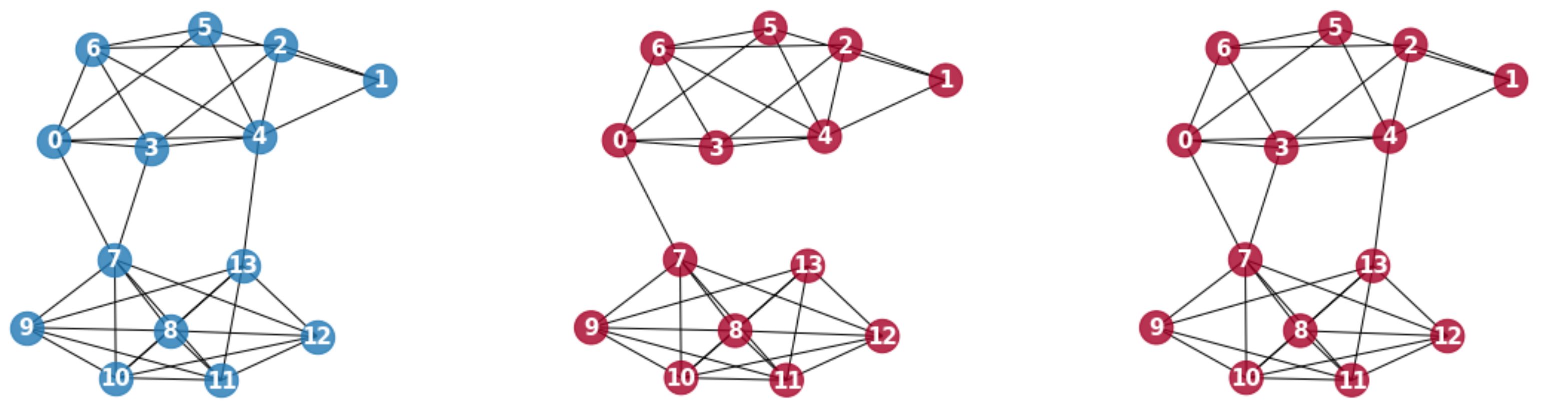
Graph-based ML Improves Performance

- Graphs can help in
  - Building meaningful neighborhoods
  - Learning ST features with geometry information embedded

- **Can graphs help further?**

Graph-based ML Improves Performance

- Deformation during reconstruction remains a problem → it can be addressed by imposing graph-based constraints (graph based optimal transport)



(a) $\mathcal{G}_1$

(b) $\mathcal{G}_2$: $\|L_1 - L_2\|_F = 2.828$, $\mathcal{W}_2^2(\nu^{\mathcal{G}_1}, \mu^{\mathcal{G}_2}) = 0.912$

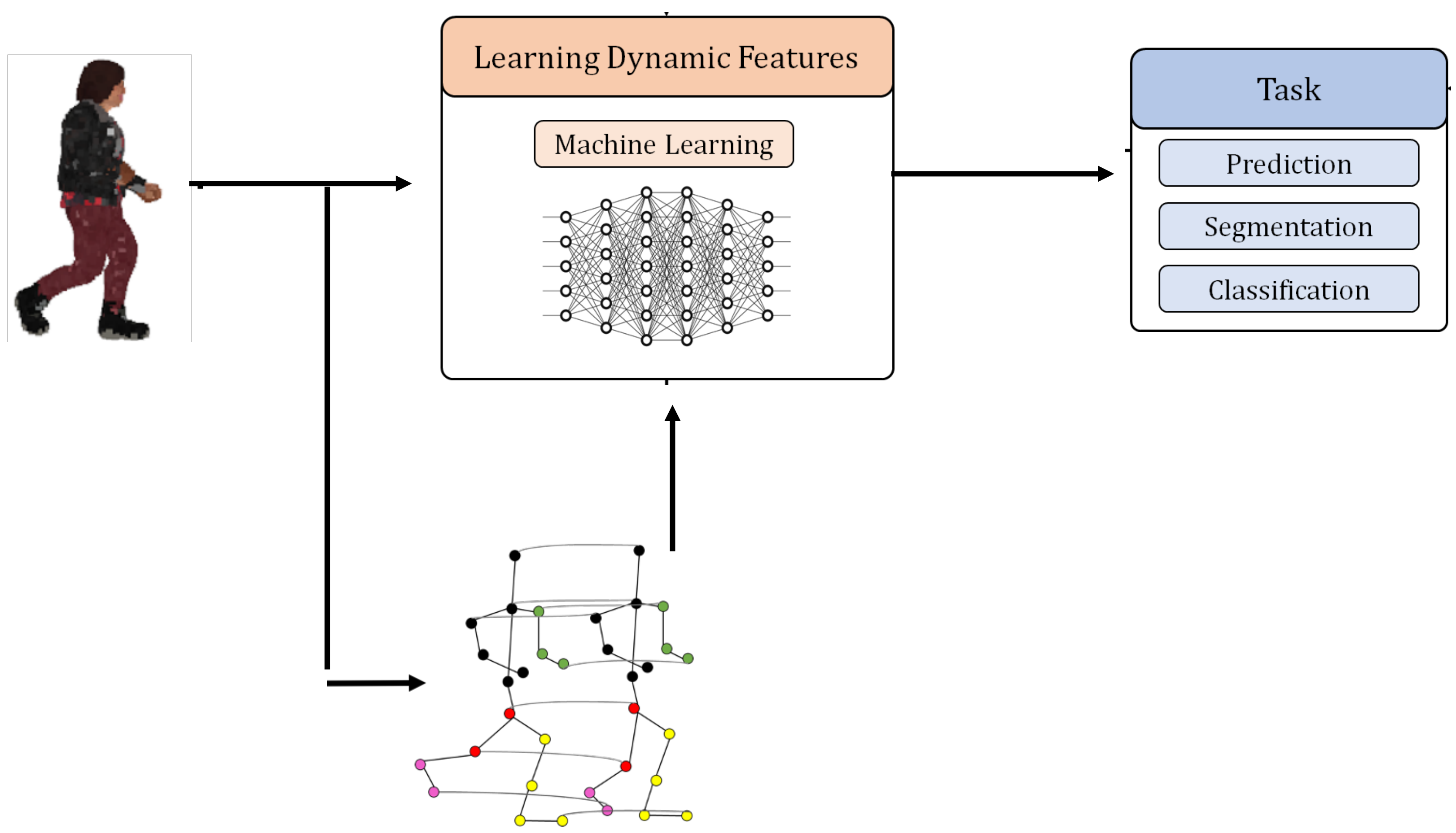
(c) $\mathcal{G}_3$: $\|L_1 - L_3\|_F = 2.828$, $\mathcal{W}_2^2(\nu^{\mathcal{G}_1}, \mu^{\mathcal{G}_3}) = 0.013$

- K Yang, X Dong L Toni, “Laplacian-regularized graph bandits: Algorithms and theoretical analysis”, AISTATS 2020
- L. Toni, P. Frossard, “Online network source optimization with graph-kernel MAB”, ECML 2023
- H. Maretic et al., “GOT: An Optimal Transport framework for Graph comparison”, 2019

How much can graph-based ML help?

- Deformation during reconstruction remains a problem → it can be addressed by imposing graph-based constraints (graph based optimal transport)

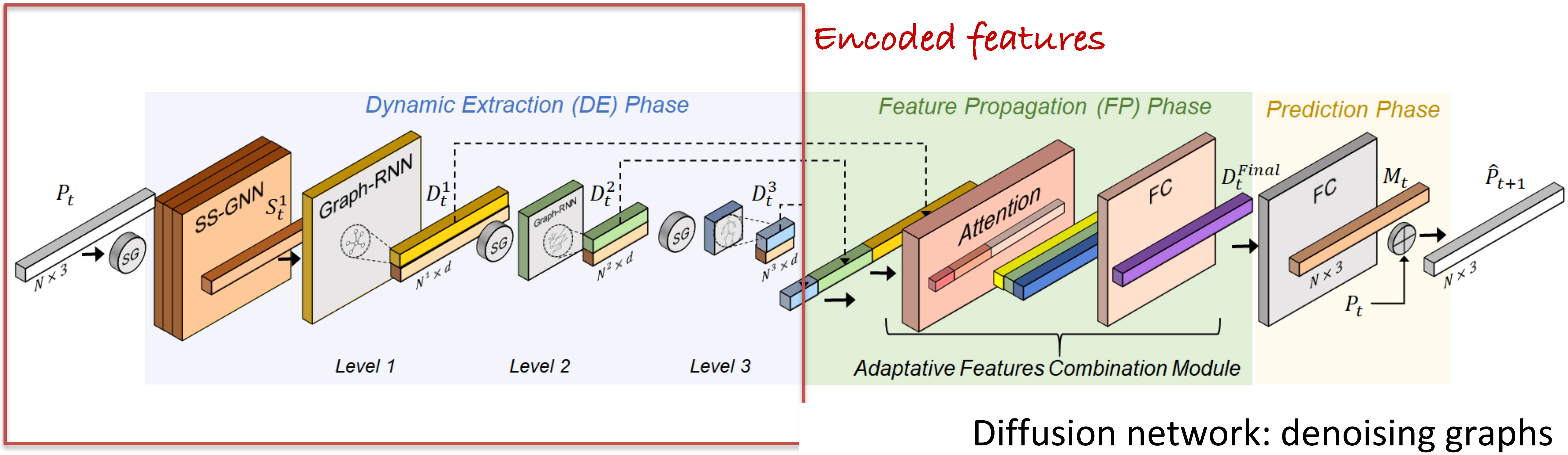
- Building a spatial-temporal skeleton as prior knowledge



- K Yang, X Dong L Toni, “Laplacian-regularized graph bandits: Algorithms and theoretical analysis”, AISTATS 2020
- L. Toni, P. Frossard, “Online network source optimization with graph-kernel MAB”, ECML 2023
- H. Maretic et al., “GOT: An Optimal Transport framework for Graph comparison”, 2019

Can Graphs Help Further?

- Deformation during reconstruction remains a problem → it can be addressed by imposing graph-based constraints (graph based optimal transport)
- Building a spatial-temporal skeleton as prior knowledge
- **Dynamic point cloud compression** based on new ML model

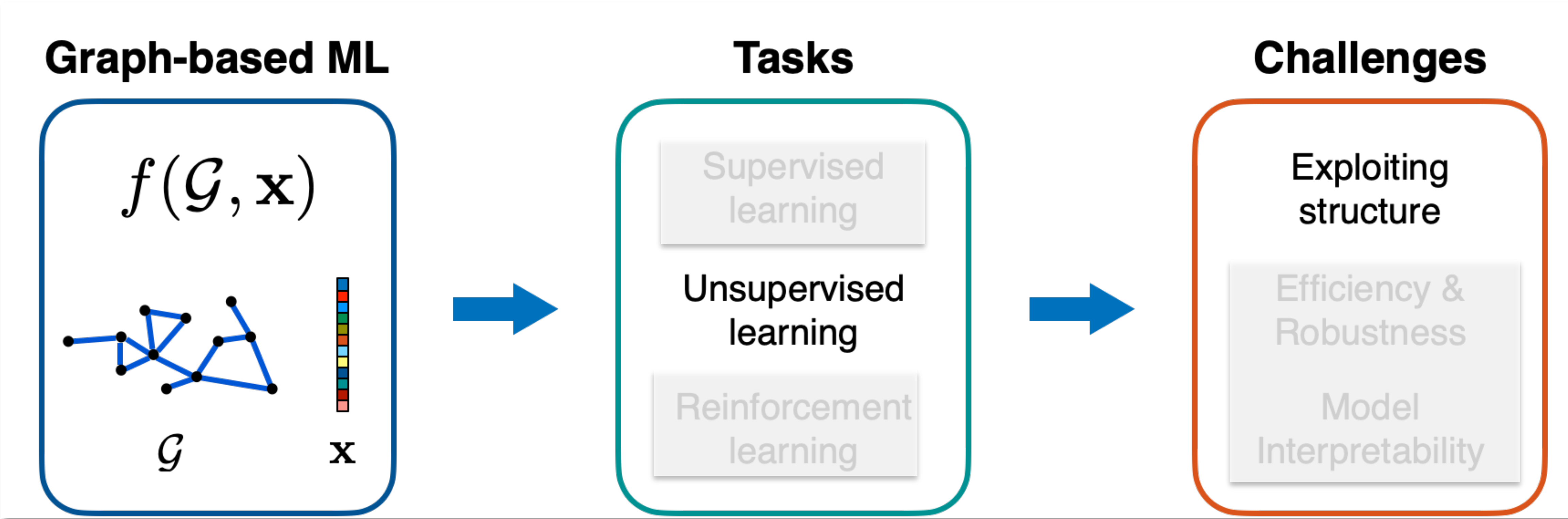


Diffusion network: denoising graphs

https://towardsdatascience.com/predictions-and-hopes-for-geometric-graph-ml-in-2022-aa3b8b79f5cc
J. Jumper et al., Highly accurate protein structure prediction with AlphaFold, Nature 596:583–589, 2021.

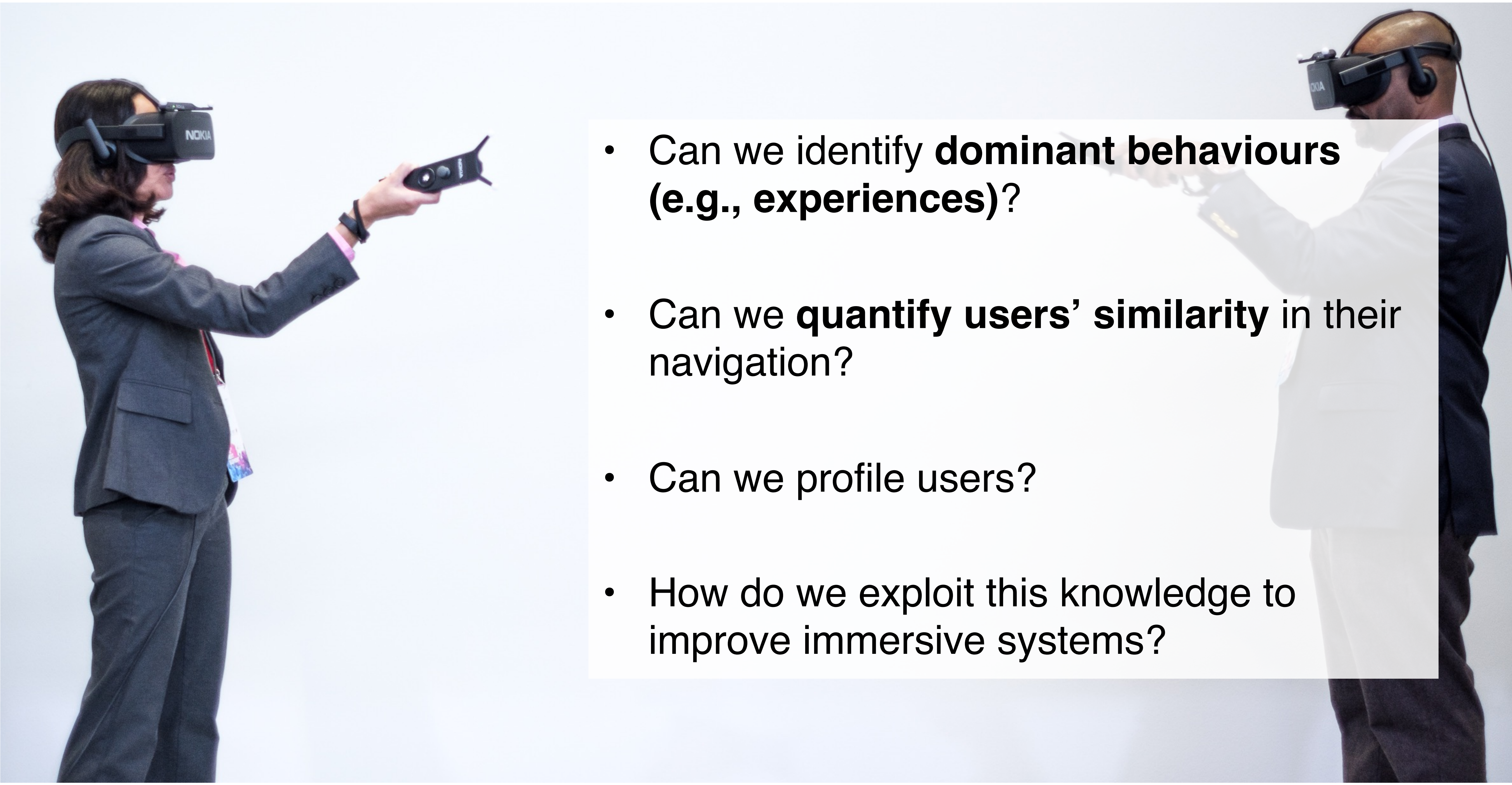
Behavioral Study of Interactive Users and its Application in Immersive Communications

Taxonomy



Graph-based ML
$f(\mathcal{G}, \mathbf{x})$
$\mathcal{G}$
$\mathbf{x}$
Tasks
Supervised learning
Unsupervised learning
Reinforcement learning
Challenges
Exploiting structure
Efficiency & Robustness
Model Interpretability

Our Main Goal

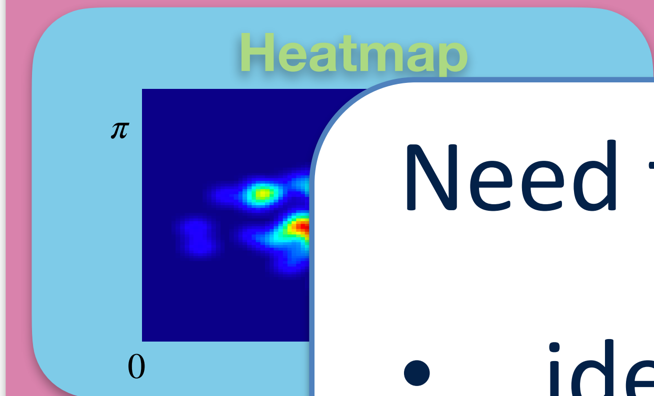


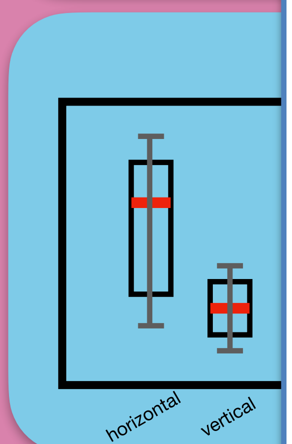
- Can we identify **dominant behaviours (e.g., experiences)**?
- Can we **quantify users' similarity** in their navigation?
- Can we profile users?
- How do we exploit this knowledge to improve immersive systems?

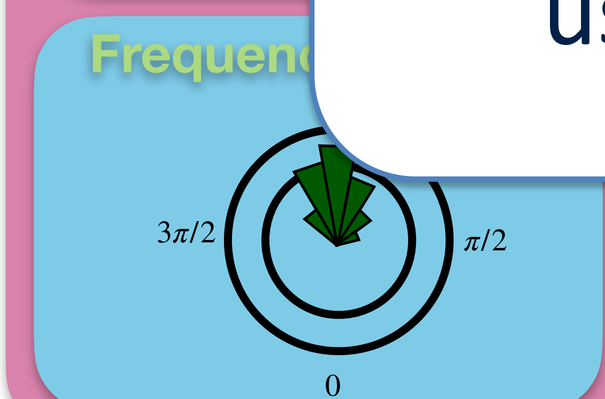
Image Credit: https://upload.wikimedia.org/wikipedia/commons/0/04/Mobile_World_Congress_2017_%2838277560286%29.jpg

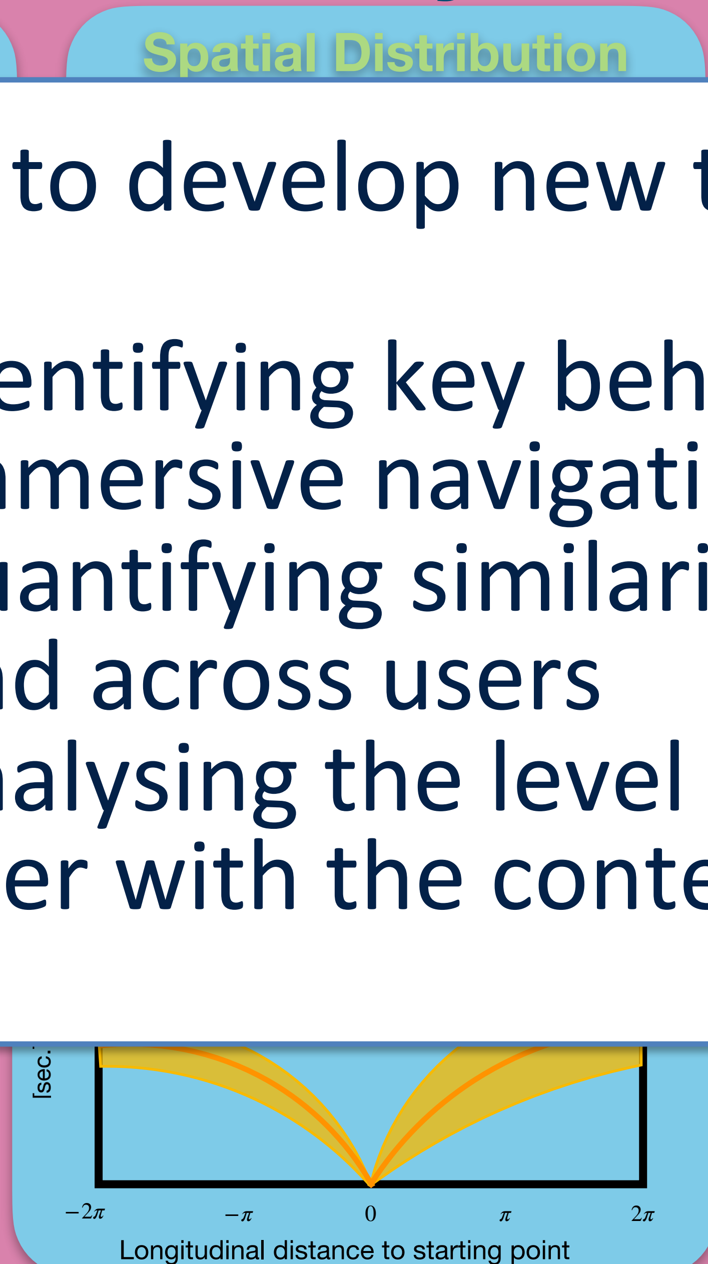
Traditional vs Trajectory-Based Data Analysis

Traditional Data Analysis









Trajectory-based Analysis

Need to develop new tools and metrics for

- identifying key behavioral aspects of immersive navigation
- quantifying similarities across contents and across users
- analysing the level of interaction of the user with the content

Behavioral Analysis Overlooked

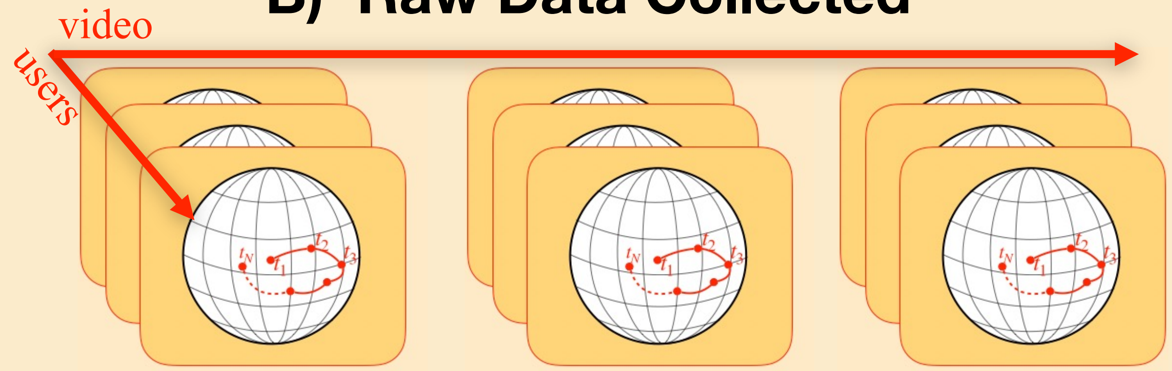
Survey	Acquisition	Compression	Delivery	Rendering	Quality Assessment	Prediction	Behavioural Analysis
Chen *et al.* [15]		■■■	■■		■■■		
He *et al.* [16]		■■	■■			■■	
Fan *et al.* [17]	■■	■■■	■■■		■■		
Zink *et al.* [18]	■	■	■	■	■		
Azevedo *et al.* [19]	■■■	■■	■	■■■	■■		
Yaqoob *et al.* [20]		■■	■■■		■■	■■	
Shafi *et al.* [21]		■■■	■■■	■■	■■■		
Xu *et al.* [22]		■■■	■■		■■■	■■	■
Ruan *et al.* [23]	■	■	■	■			
Chiarotti [24]		■■	■■■		■■■	■■■	
Our chapter		■■	■■■			■■■	■■■

S. Rossi, A. Guedes, and **L. Toni,** "Coding, Streaming, and User Behaviour in Omnidirectional Video" Elsevier, 2022.

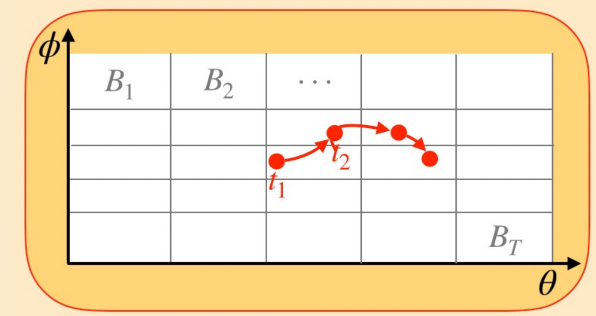
User Behaviour Analysis in VR system

A) Experiments

B) Raw Data Collected

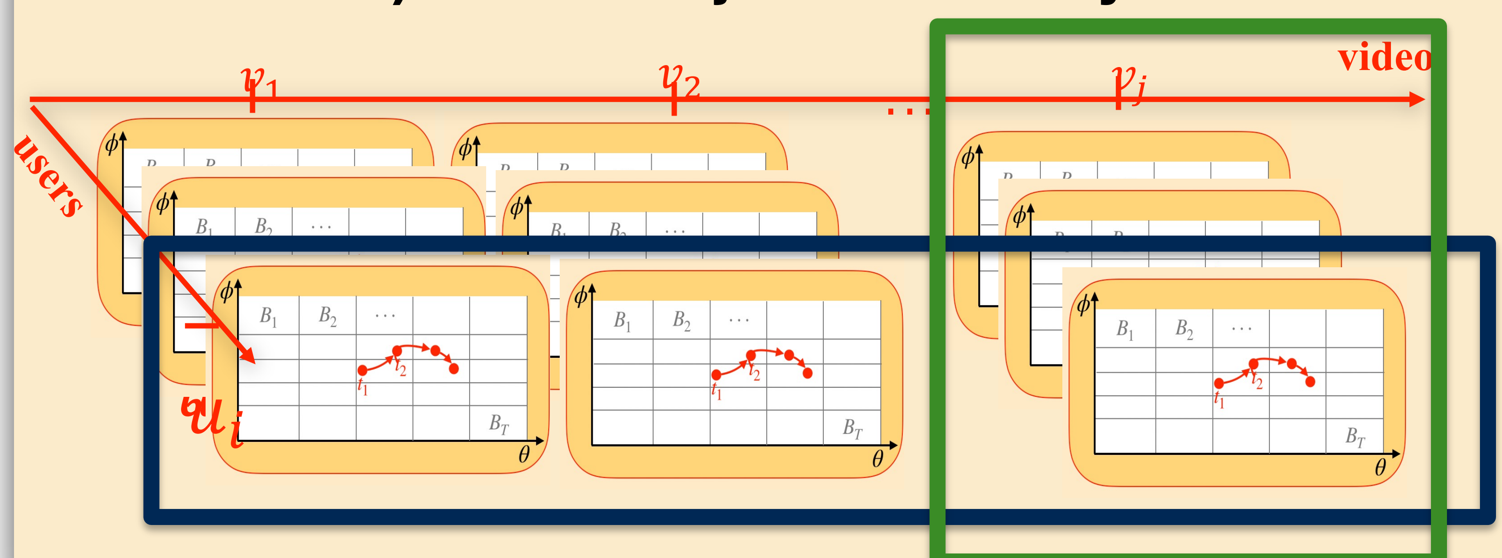


C) Pre-Processing



$u_i = < (x_1, t_1), \ldots, (x_n, t_n) >$

D) User's Trajectories Analysis



Intra-user behaviour analysis:

To characterise the navigation of each user over time against different video contents.

Actual Entropy
Fixation map Entropy

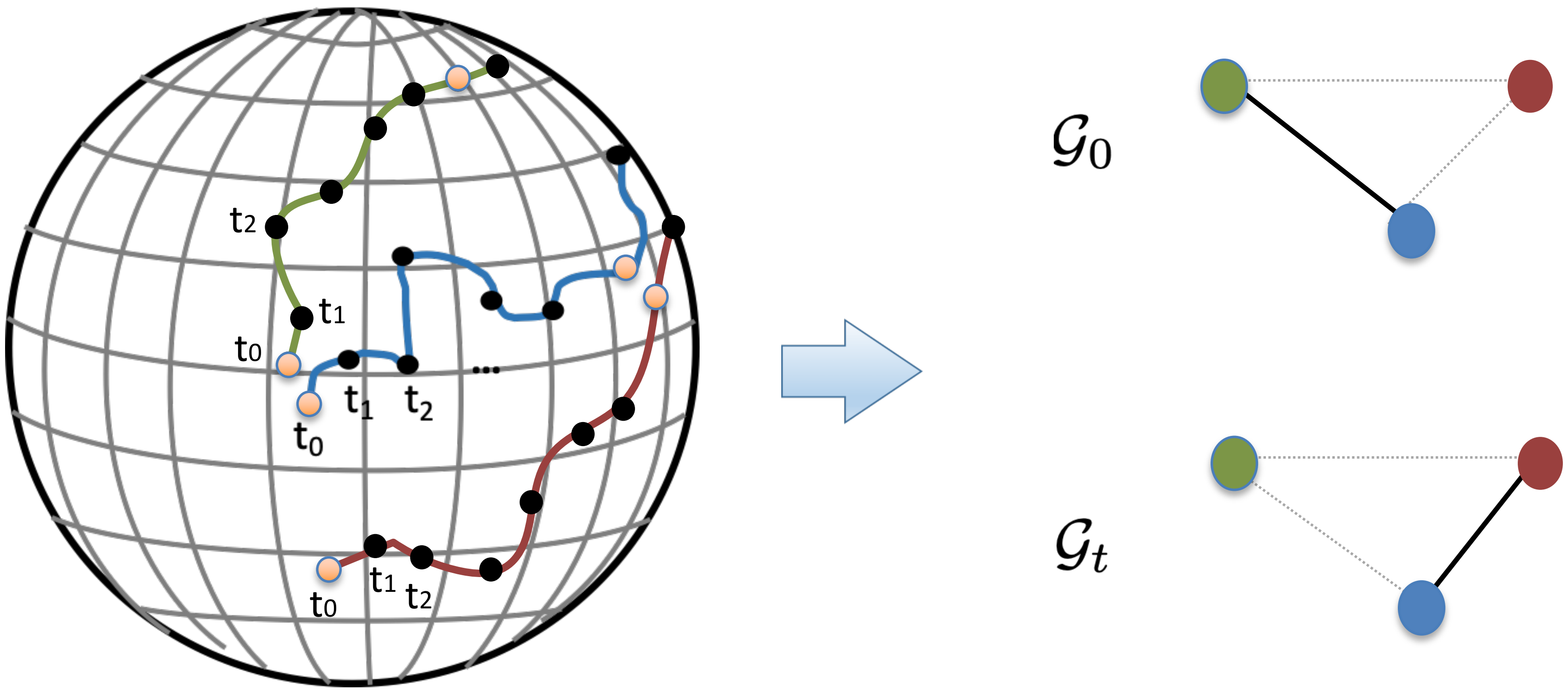
Inter-user behaviour analysis

To study the behaviour of a single user in correlation with others in the same content.

User Affinity Index

UCL

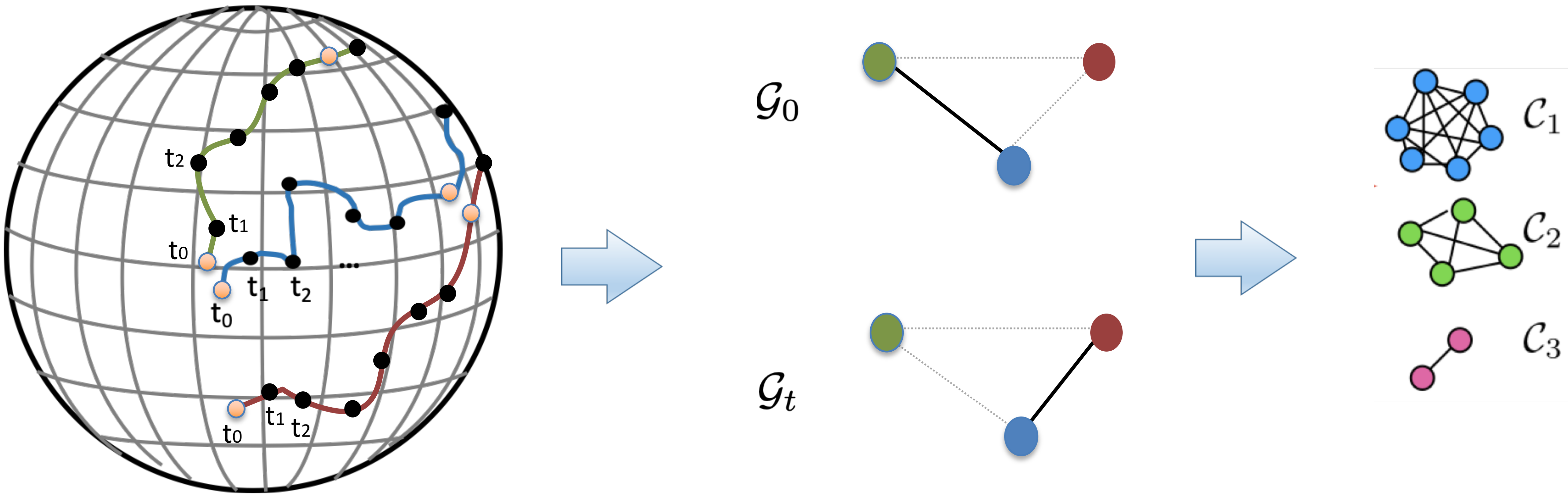
A Graph Approach



Proxy for viewport overlap: distance between viewport center

S. Rossi, et al., “Spherical clustering of users navigating in 360-degree content.”, ICASSP 2019.

A Graph Approach



Our Goal: To identify a metric able to measure the similarity (in terms of navigation) between users.

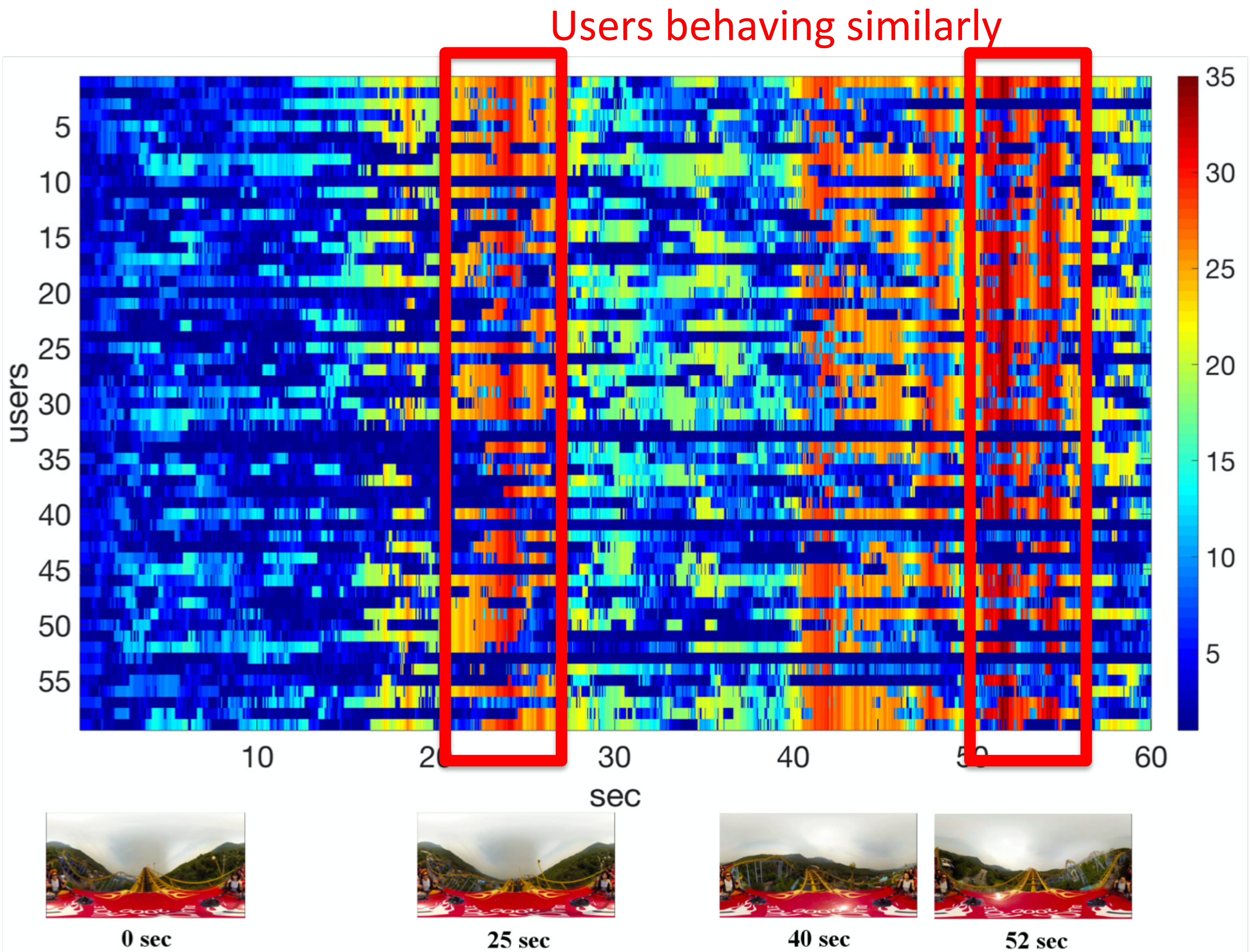
Affinity metric

$$UAI = \frac{\sum_{i=1}^{C} x_i \cdot w_i}{\sum_{i=1}^{C} w_i}$$

- *C*: number of clusters detected in a frame by the clique-clustering
- x_i : % of users in cluster *i*
- w_i : number of users in cluster *i*

S. Rossi, et al., “Spherical clustering of users navigating in 360-degree content.”, ICASSP 2019.

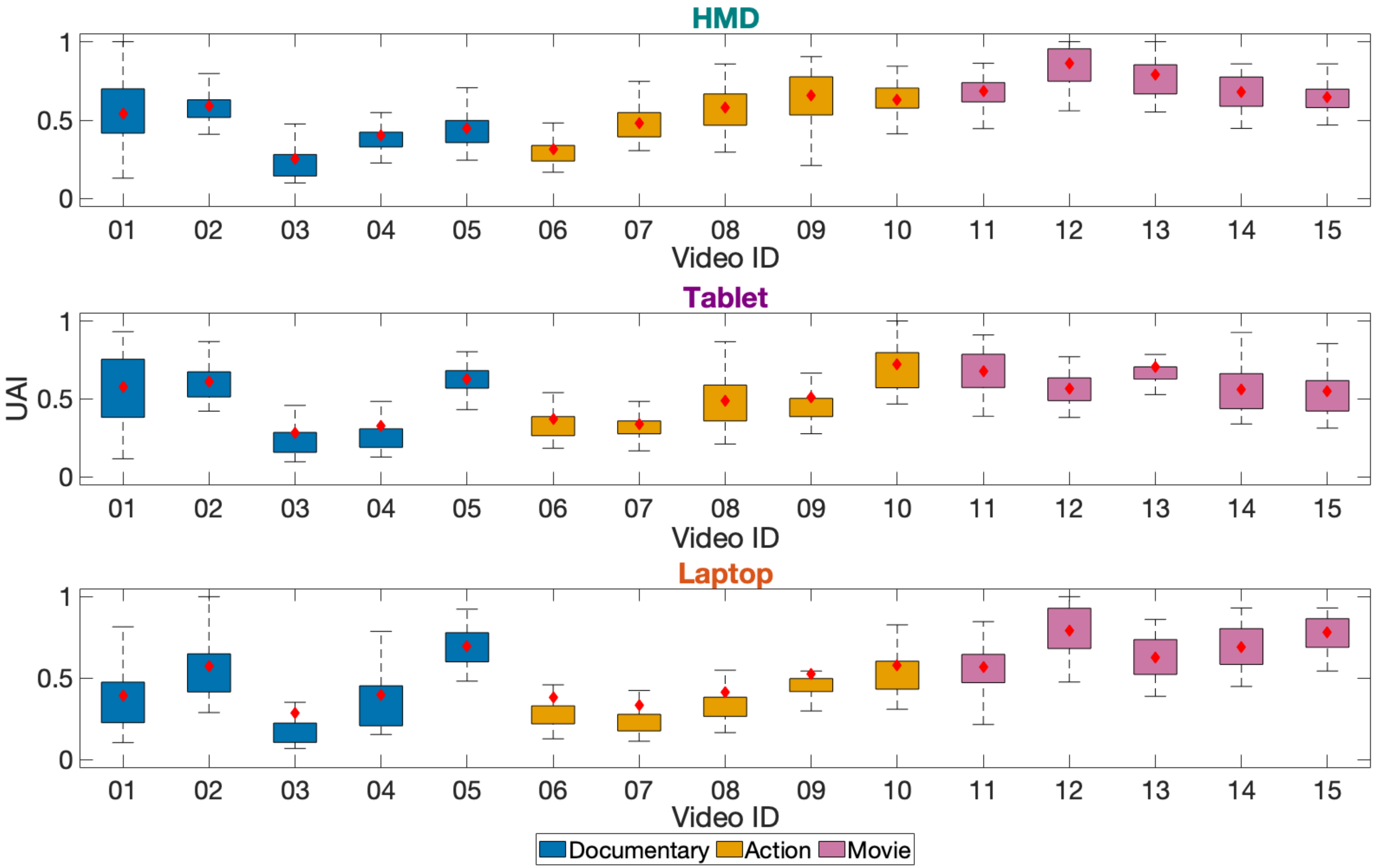
Analysis based on Clusters



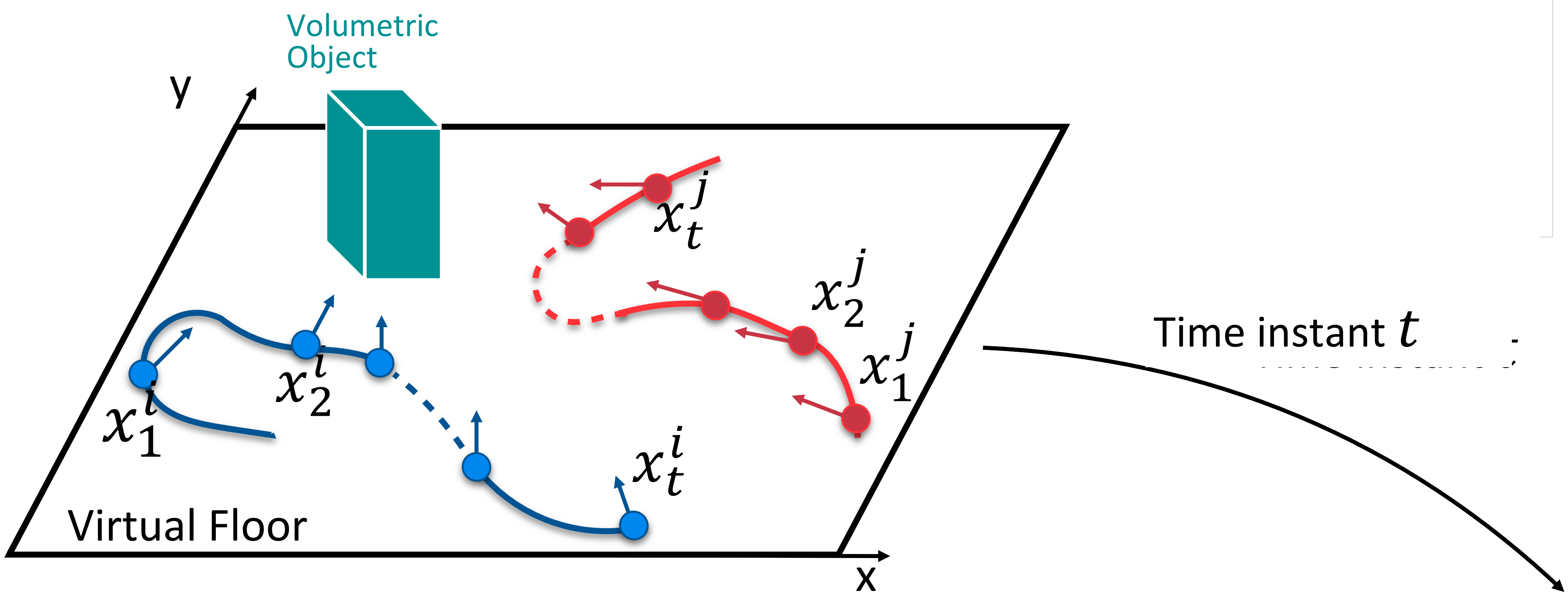
(a) Rollercoaster video

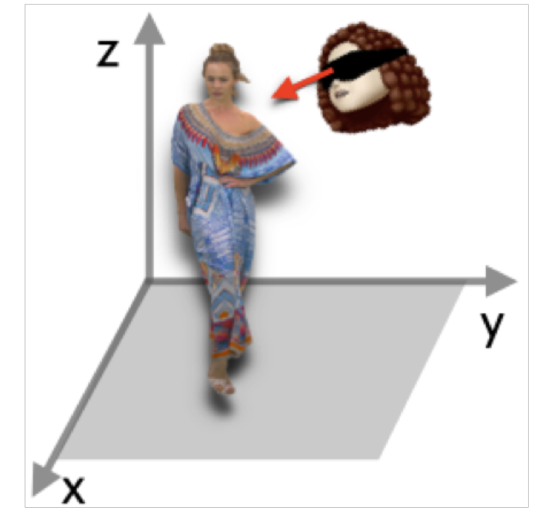
S. ROSSI, C. OZCINAR, A. SMOLIC, L. TONI, "Do Users Behave Similarly in VR? Investigation of the User Influence on the System Design ", ACM TOMM 2020

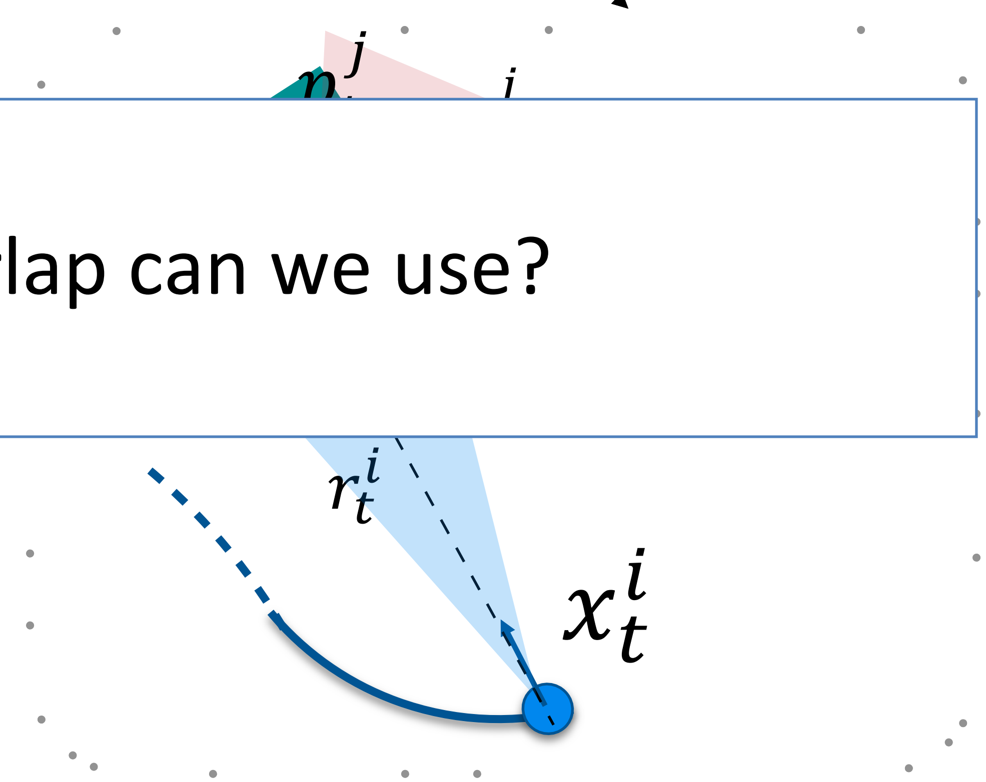
Users Similarities vs Displaying Devices



User Navigation in 6-DoF







Which proxy of the viewport overlap can we use?

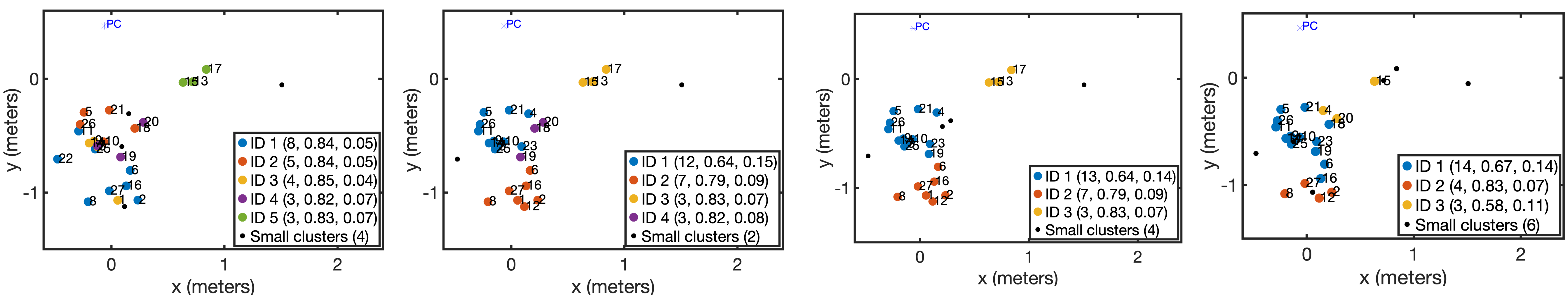
the volumetric object

- r $\rightarrow$ relative distance between user and volumetric content

Seeking a New Metrix for 6-DoF

Table 1. Definition of distance features and measurements.

	Symbol	Definition
Distance Features	x	user position on the VR floor
	p	viewport center projected on the volumetric content
	r	relative distance between user and volumetric content
	v	vector of the viewing direction
Distance Measurements	$\mathrm{L}(\cdot,\cdot)$	difference of relative distance between two users
	$\mathrm{E}(\cdot,\cdot)$	Euclidean distance
	$\mathrm{G}(\cdot,\cdot)$	Geodesic distance
	$\theta(\cdot,\cdot)$	Angle between two vectors



(a) Ground-truth (O_{th} = 75%)

(b) w_1 (single feature metric)

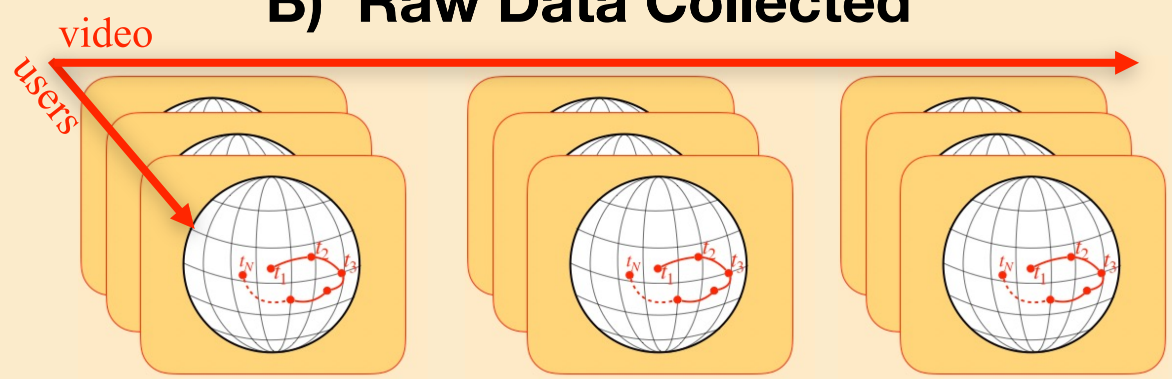
(j) w_9 (multi-feature metric)

(k) w_{10} (multi-feature metric)

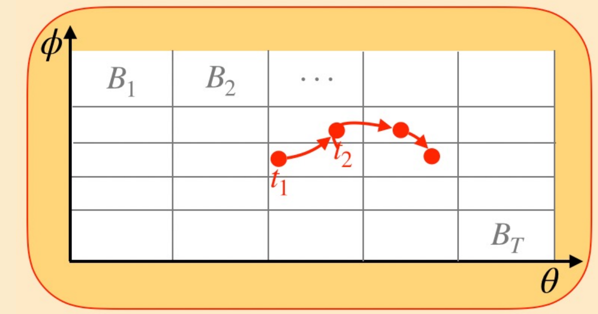
User Behaviour Analysis in VR system

A) Experiments

B) Raw Data Collected

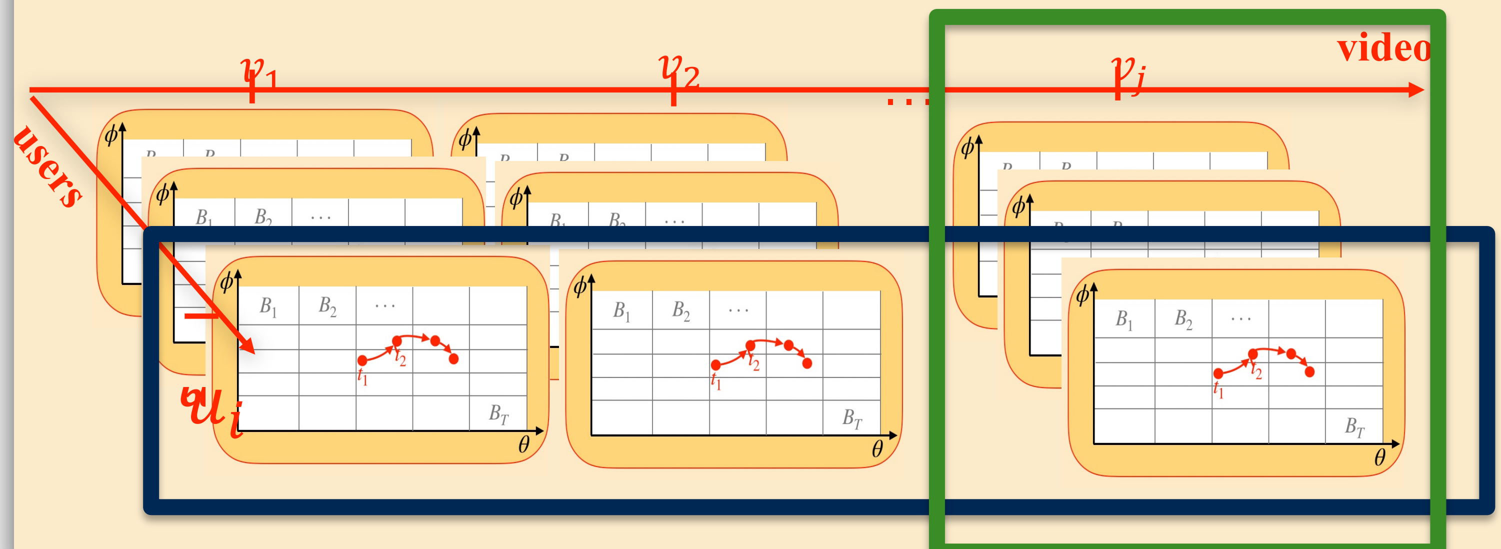


C) Pre-Processing



$u_i =< (x_1, t_1), .., (x_n, t_n) >$

D) User's Trajectories Analysis



Intra-user behaviour analysis:

To characterise the navigation of each user over time against different video contents.

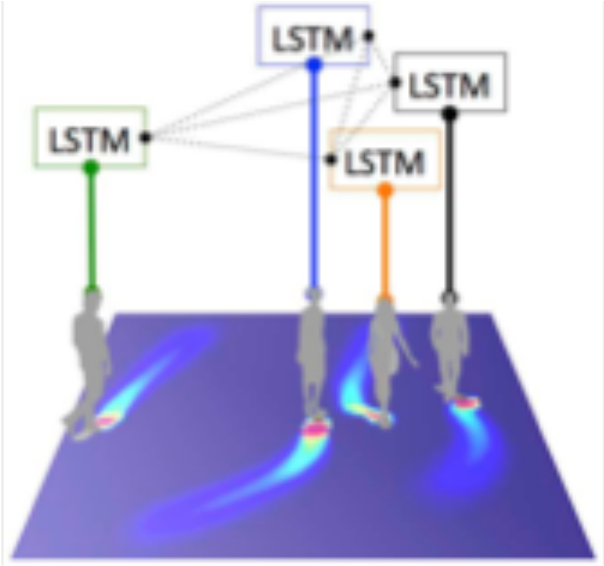
Actual Entropy
Fixation map Entropy

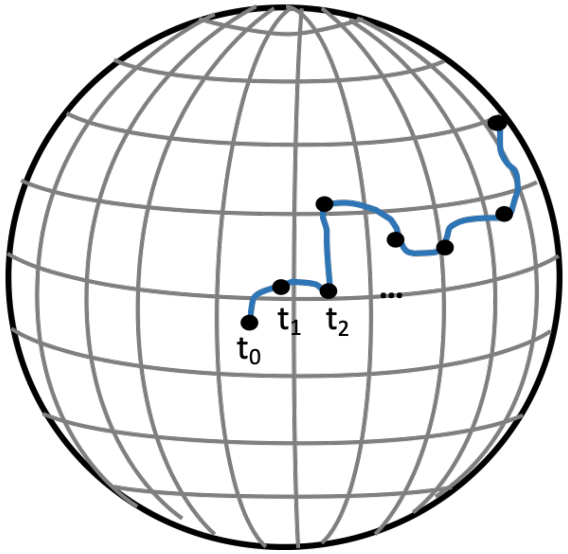
Inter-user behaviour analysis

To study the behaviour of a single user in correlation with others in the same content.

User Affinity Index

Intra-User behaviour metrics





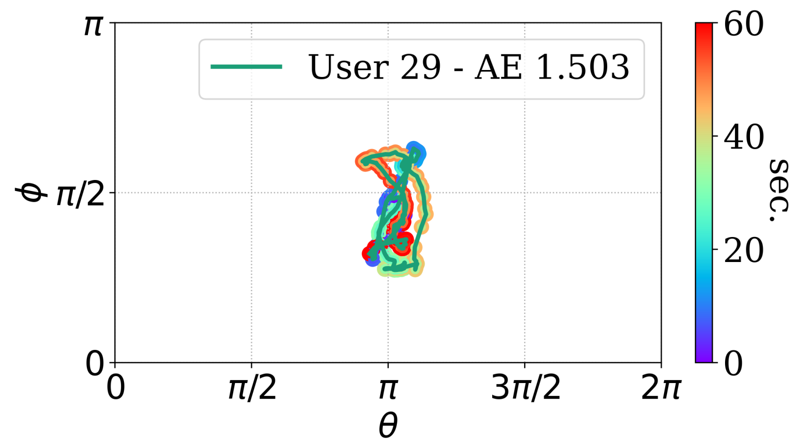
Key basics from the study of human behavior / pedestrian

Introduced as a proxy of predictability of human mobility patterns [1], the actual entropy **quantifies** the **information** carried within a given **trajectory**.

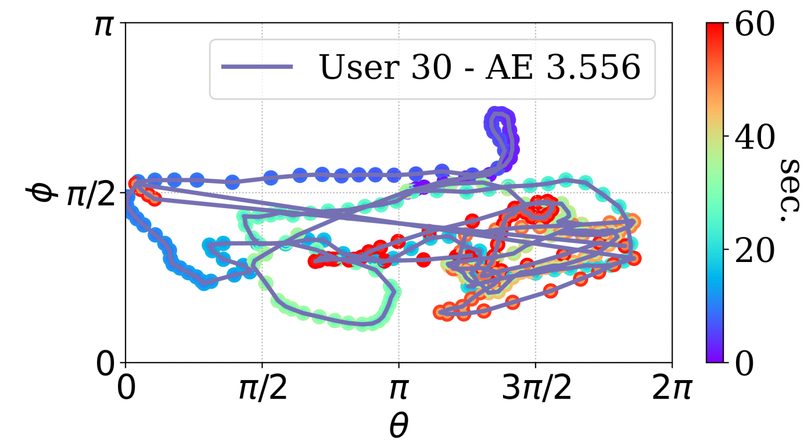
Alexandre Alahi et al, “Social LSTM: Human Trajectory Prediction in Crowded Spaces”, CVPR 2016
C. Song, Z. Qu, N. Blumm, and A. Barabási. 2010. **Limits of predictability in human mobility**. Science.

Actual Entropy for Users' Behaviour

(a) Video "Diving" from 2_MMsys2017_2 [17]

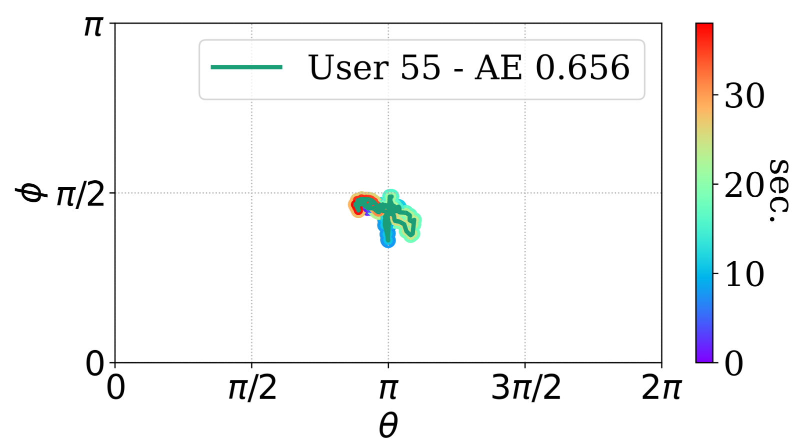


(b) Low Entropy User in "Diving"

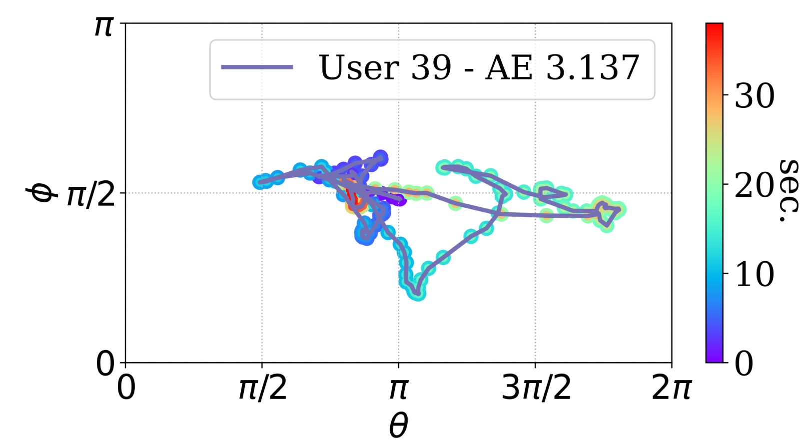


(c) High Entropy User in "Diving"

(d) "AcerPredator" from 6_PAMI2018 [38]



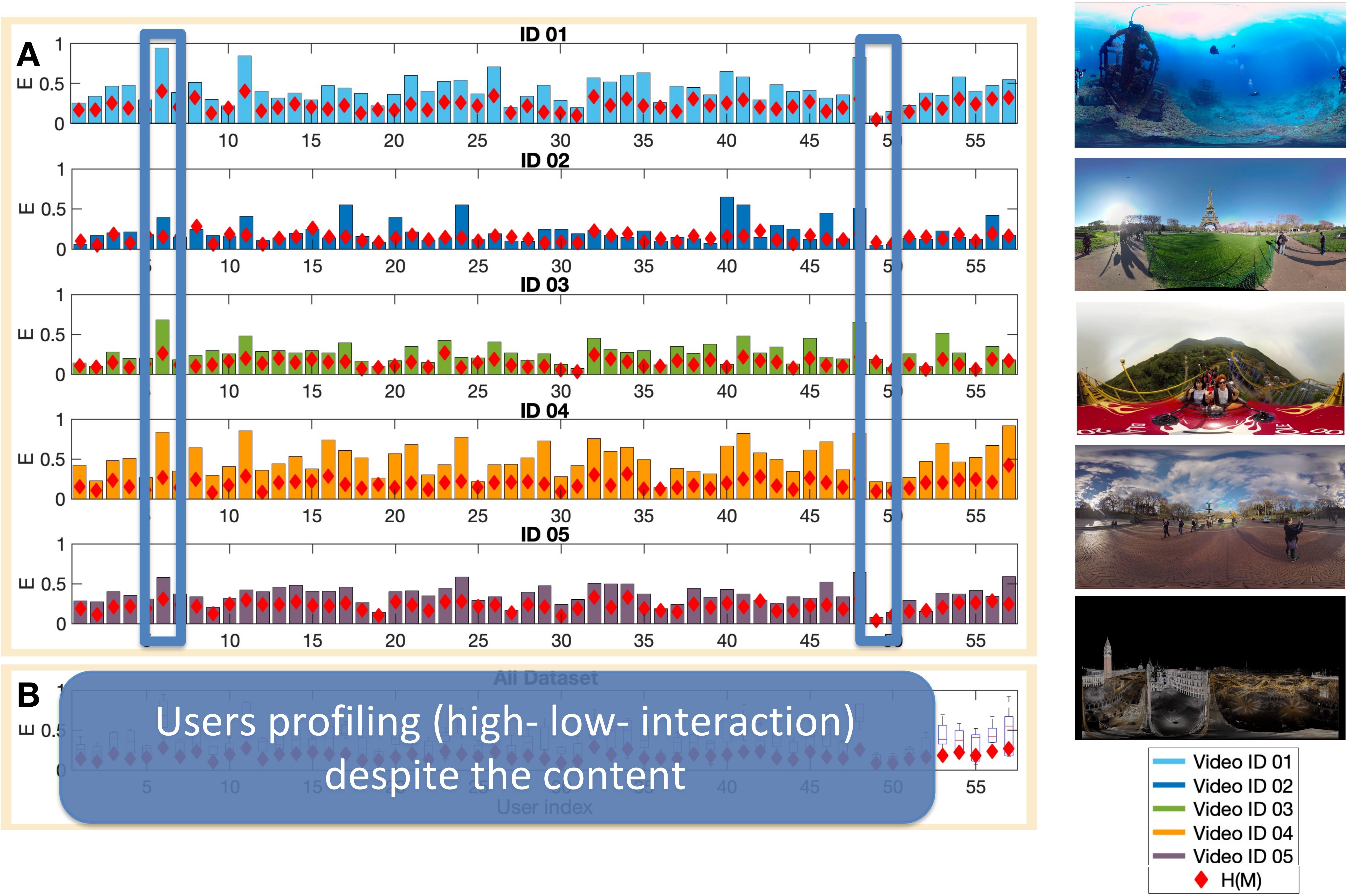
(e) Low entropy user in "AcerPredator"



(f) High Entropy User in "AcerPredator"

S. Rossi, **L. Toni**, P. Cesar, "Correlation between Entropy and Prediction Error in VR Head Motion Trajectories", IMX Workshop, ACM MM 2023

Intra-User behavior analysis

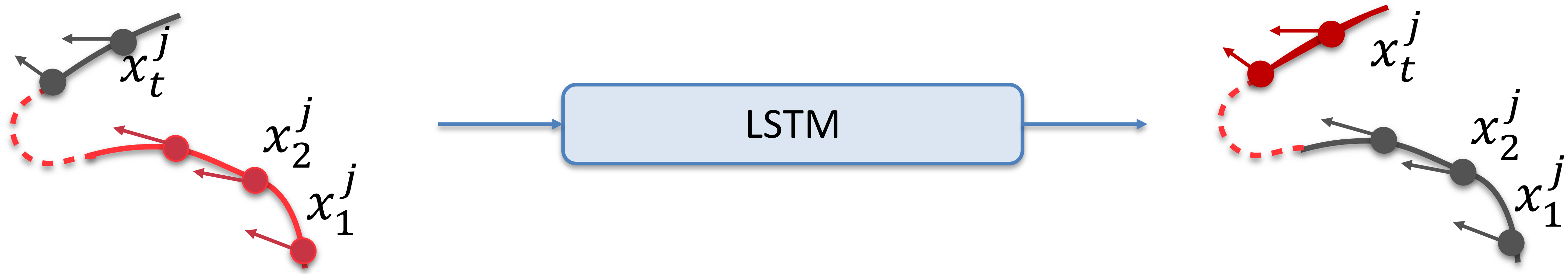


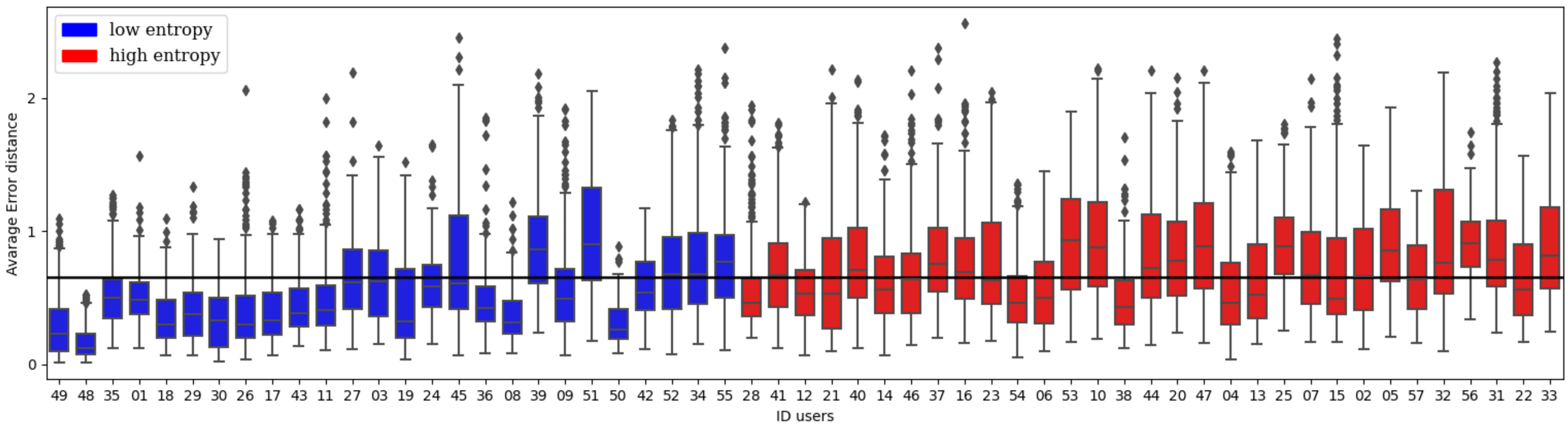
Users profiling (high- low- interaction) despite the content

S. Rossi, L. Toni, "Understanding user navigation in immersive experience: an information-theoretic analysis" MMVE 2020

UCL

Correlation with Prediction Error





S. Rossi, **L. Toni**, P. Cesar, "Correlation between Entropy and Prediction Error in VR Head Motion Trajectories", IMX Workshop, ACM MM 2023

Take Home Message

- **Graph-based clustering** and **information theory metrics** can be used as novel metrics for studying users' behaviour

- Metrics can be extended to 6-DoF systems (?)

- Content and display settings do impact the users' behaviour, but at the same time users tend have their own "style" of interaction
  - Can we profile users? With application toward the healthcare sector?
  - What can we infer from users' trajectory (privacy concerns)

- Link between Entropy and estimation error
  - Can we create better dataset to train prediction models?

Feature Works and Open Challenges

- Analysis of users behavior to develop **user-centric systems** (coding, quality, streaming etc. tailored to users)

- Graphs can further improve users analysis/prediction [1] and can help toward spherical processing [2]

[1] X Zhang, et al., "Graph Learning Based Head Movement Prediction for Interactive 360 Video Streaming" IEEE Transactions on Image Processing,
[2] T Maugey, "Graph Spectral 3D Image Compression", Graph Spectral Image Processing, 105, 2021

Feature Works and Open Challenges

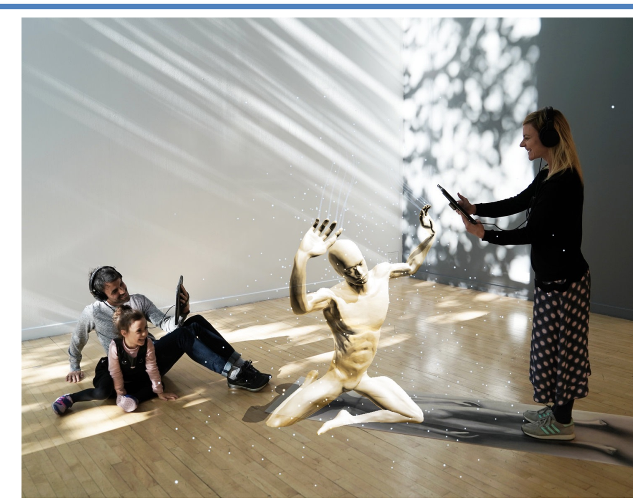
- Analysis of users behavior to develop **user-centric systems** (coding, quality, streaming etc. tailored to users)

- Graphs can further improve users analysis/prediction [1] and can help toward spherical processing [2]

- Metrics for understanding users behaviour can be extended to healthcare, performing arts, etc with HCI and social science interconnection

User engagement for art/cultural heritage and healthcare

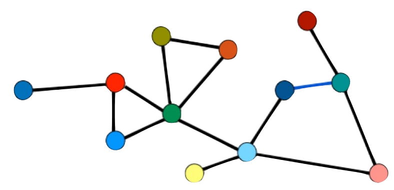
VR therapists



Live performance

Conclusions

- Metaverse with high potential impact
- Many Challenges need to be addressed
- We will focus on user-centric systems and new modalities



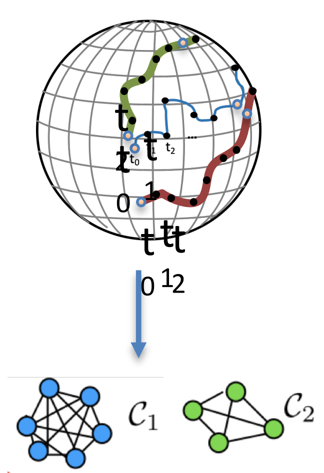
- GSP-based ML as growing fields
- Many Applications to MM

Dynamic Point Cloud Prediction

- Point-based processing algorithms. Current SOTA lack of structural relationship → graph processing is essential to overcome this limitation → Dynamic features extraction via GNN
- Complex motions requires attention based mechanism → Interpretability perspective to MVs

Future Directions

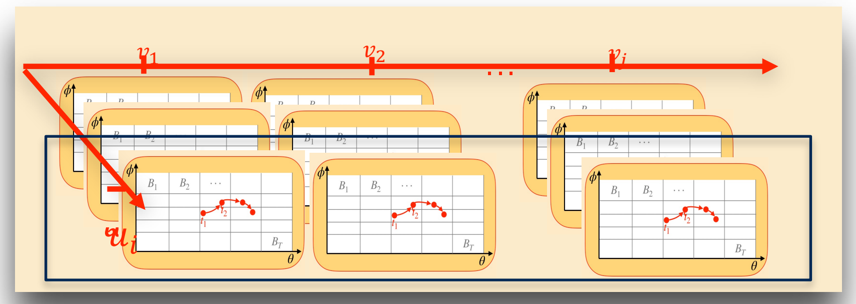
- Graph based Optimal transport to preserve geometry
- Graph-based generative models
- Applications to robotics/ new sensors/ modalities



Do users interact in similar way?

- Graph to capture users' similarity
- Graph-based clustering to detect meaningful clusters and derive quantitative similarity metric

- User affinity grows with focus of attention
- User affinity is affected by displaying device



Users Profiling?

- Users interact in consistent way
- Quantitative metric to capture this behaviour

Thank You! Questions?

Learning and Signal Processing Lab
UCL
https://lasp-ucl.github.io