



LISTIC

Laboratoire d'Informatique, Systèmes,
Traitement de l'Information et de la Connaissance

Computer Science, Systems,
Information and Knowledge Processing Laboratory

Rencontre LAPP - LISTIC
22 janvier 2021

Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance

- 1 équipe : 36 enseignants chercheurs, 10-15 doctorants, 3 IATOS
- 2 thèmes :
 - **AFuTé : Apprentissage, Fusion et Télédétection**
 - **Apprentissage automatique : neuronal**, statistique, **fouille de données**...
 - Fusion de données incertaines : approches possibilistes, crédibilistes...
 - Télédétection : images satellites optiques et radar, géoradar, photogrammétrie...
 - **ReGaRD : Représentation, Gestion et tRaitement des Données pour l'humain**
 - Réseaux et systèmes distribués : **graphes**, **parallélisme**, cybersécurité...
 - **Gestion, placement et transport des grandes masses de données**
 - Traitement et l'analyse pour l'aide à la personne, et l'aide à la décision

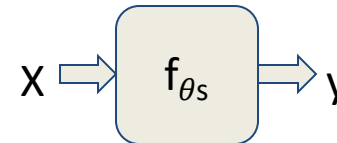
Activités sur l'apprentissage profond

Apprentissage classique (appliqué) et problématiques :

f_{θ_i} non linéaire
 θ_i très nombreux
 optimisation avec $X \in \Omega$ grand, varié

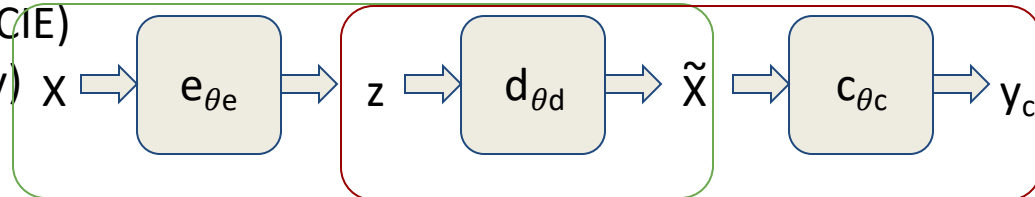
Supervisé

- ✦ détection et segmentation d'objets (Total)
- ✦ transfert de représentations (SmarterPlan)
- ✦ classification/regression (GammaLearn)



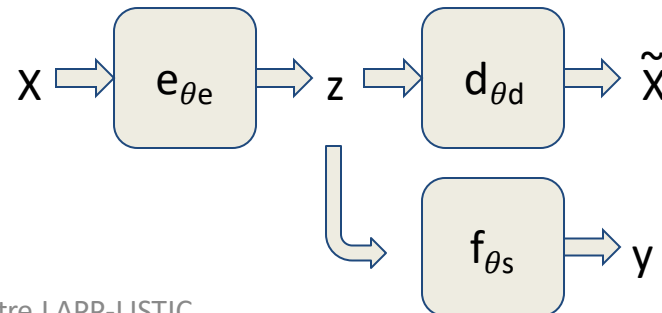
Non supervisé autoencodage/génératif

- ✦ représentation des données (LOCIE)
- ✦ détection d'anomalies (HelioCity)
- ✦ inversion (REPED-SARIX)



Semi supervisé

- ✦ adaptation de domaine (GammaLearn++)

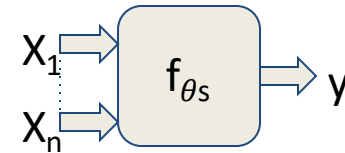


Activités sur l'apprentissage profond, problématiques spécifiques 1

🔗 Fusion de données hétérogènes (précoce, tardive)

- 🌿 Echelle des données
- 🌿 Niveau sémantique des données
- 🌿 Relation entre les sources

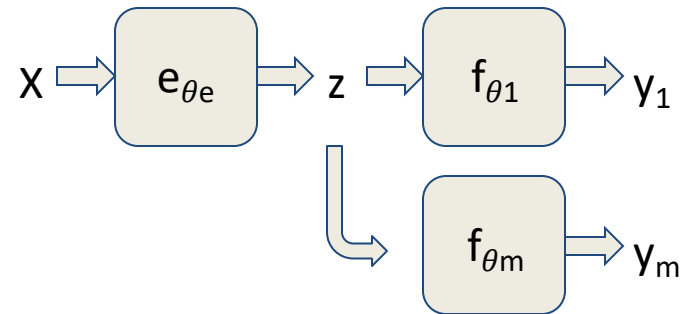
Ex : Projet Total



🔗 Multitâches

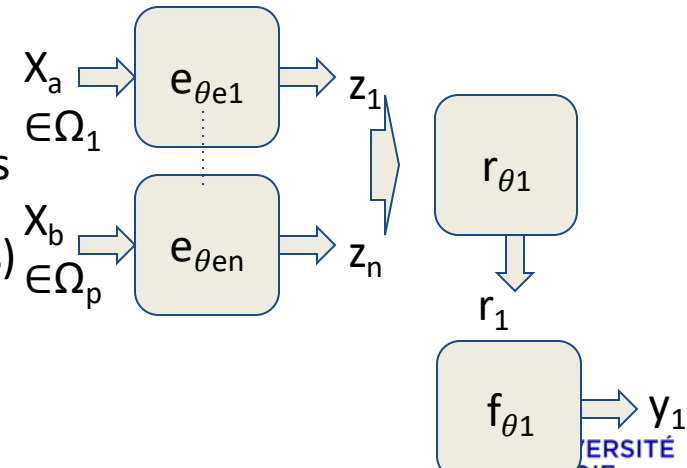
- 🌿 représentation latente et tâches spécifiques
- 🌿 contraindre la représentation latente
- 🌿 régularisation des priorités des tâches
- 🌿 transfert de représentation

Ex : Gammalearn



🔗 Apprentissage distribué/fédéré (asynchrone)

- 🌿 représentation commune/spécialisée de différentes sources de données
- 🌿 fusion de données (redondantes, complémentaires)
- 🌿 répartition des tâches tout au long du workflow complet, du capteur à l'utilisateur
 - (prétraitement, fusion, représentation, prédiction)

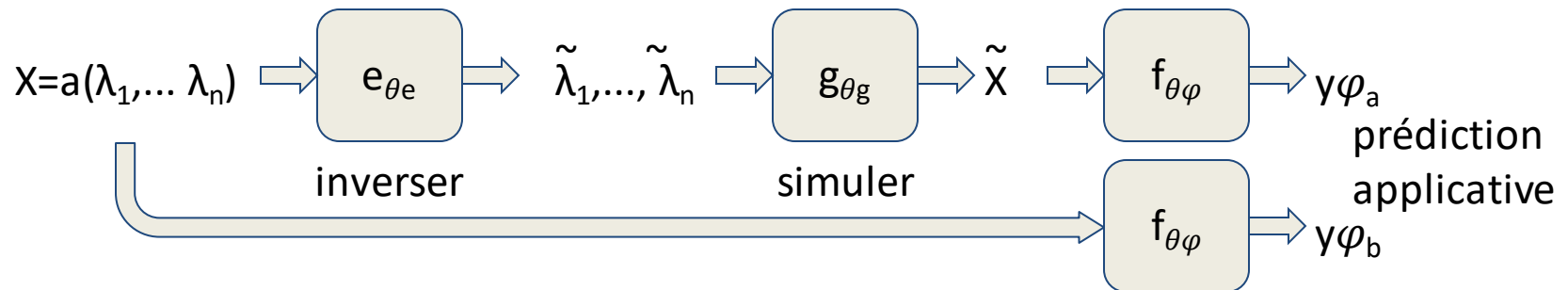


Activités sur l'apprentissage profond, problématiques spécifiques 2

🔗 Inversion de modèles physiques

🌿 introduction de contraintes

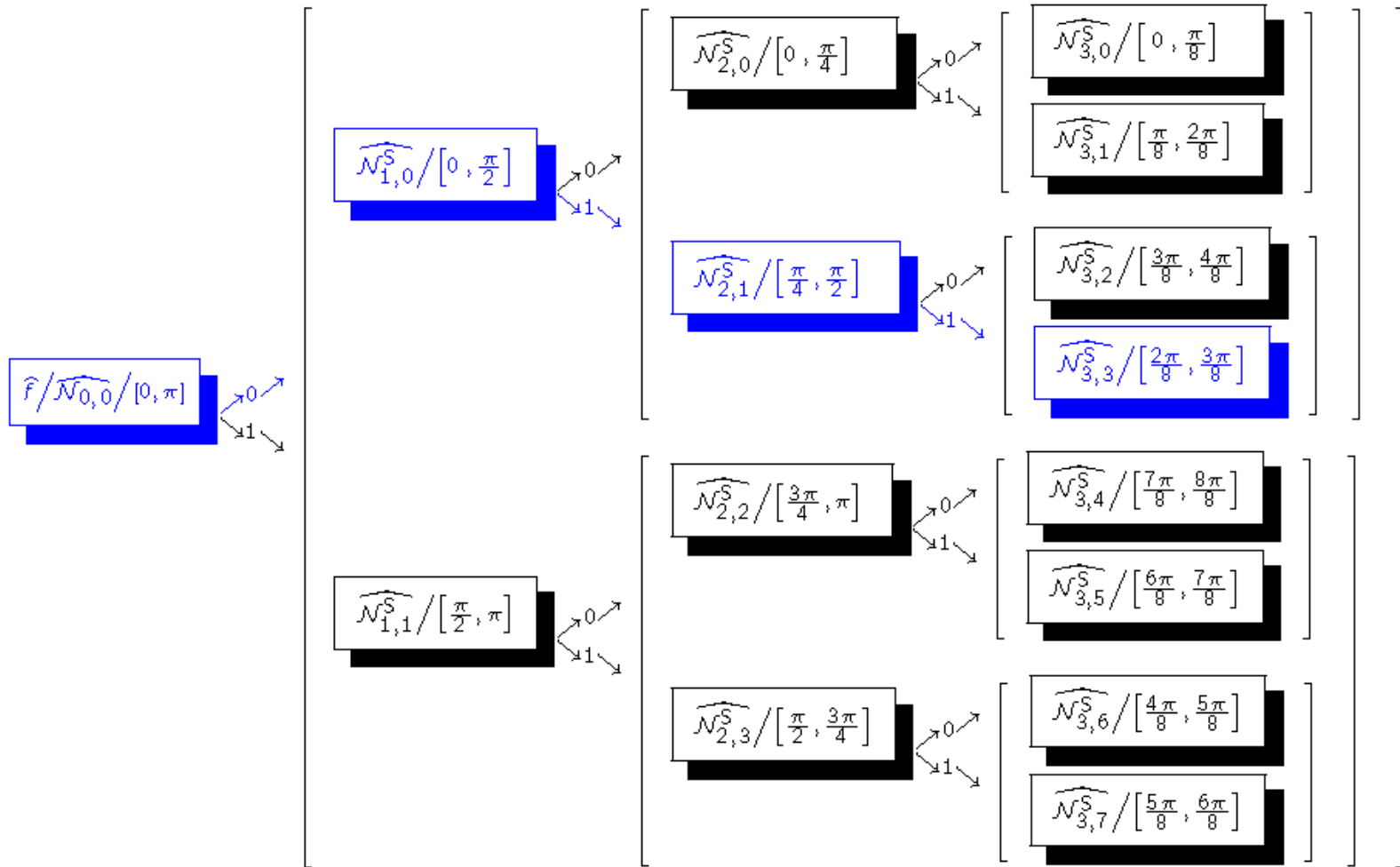
- grandeurs physiques
- encodage/représentation constraints



🔗 Explicabilité des modèles

- 🌿 prise en compte de la nature des données
- 🌿 représentations intermédiaires contraintes
- 🌿 approches data mining, identifier les patterns d'activation typiques
- 🌿 représentation hiérarchique des données (approche ondelettes)

Series of Shannon-Nyquist functionals



Deep multi-serial functionals

1. **Wavelet Packets**: multiserial convolutions / $1 \leq m \leq M_{j+1}$ / $0 \leq n \leq \prod_{\ell=1}^j M_{\ell} - 1$:

$$\mathfrak{e}_{j+1, M_{j+1}n+m}[k, \ell] = \sum_{p, q \in \mathbb{Z}} h_{j+1, m}[p, q] \mathfrak{e}_{j, n}[s_{j+1}k - p, t_{j+1}\ell - q] \quad (1)$$

2. **CNN** (Convolutional Neural Network): Depth $j + 1$ maps (expanded CNN) are:

$$\mathfrak{D}_{j+1, M_{j+1}n+m}[k, \ell] = \Upsilon_{j+1, M_{j+1}n+m} \left(\sum_{p, q \in \mathbb{Z}} h_{j+1, m}[p, q] \mathfrak{D}_{j, n}[s_{j+1}k - p, t_{j+1}\ell - q] \right) \quad (2)$$

for $0 \leq m \leq M_{j+1} - 1$ and $0 \leq n \leq M_1 M_2 \cdots M_j - 1$.

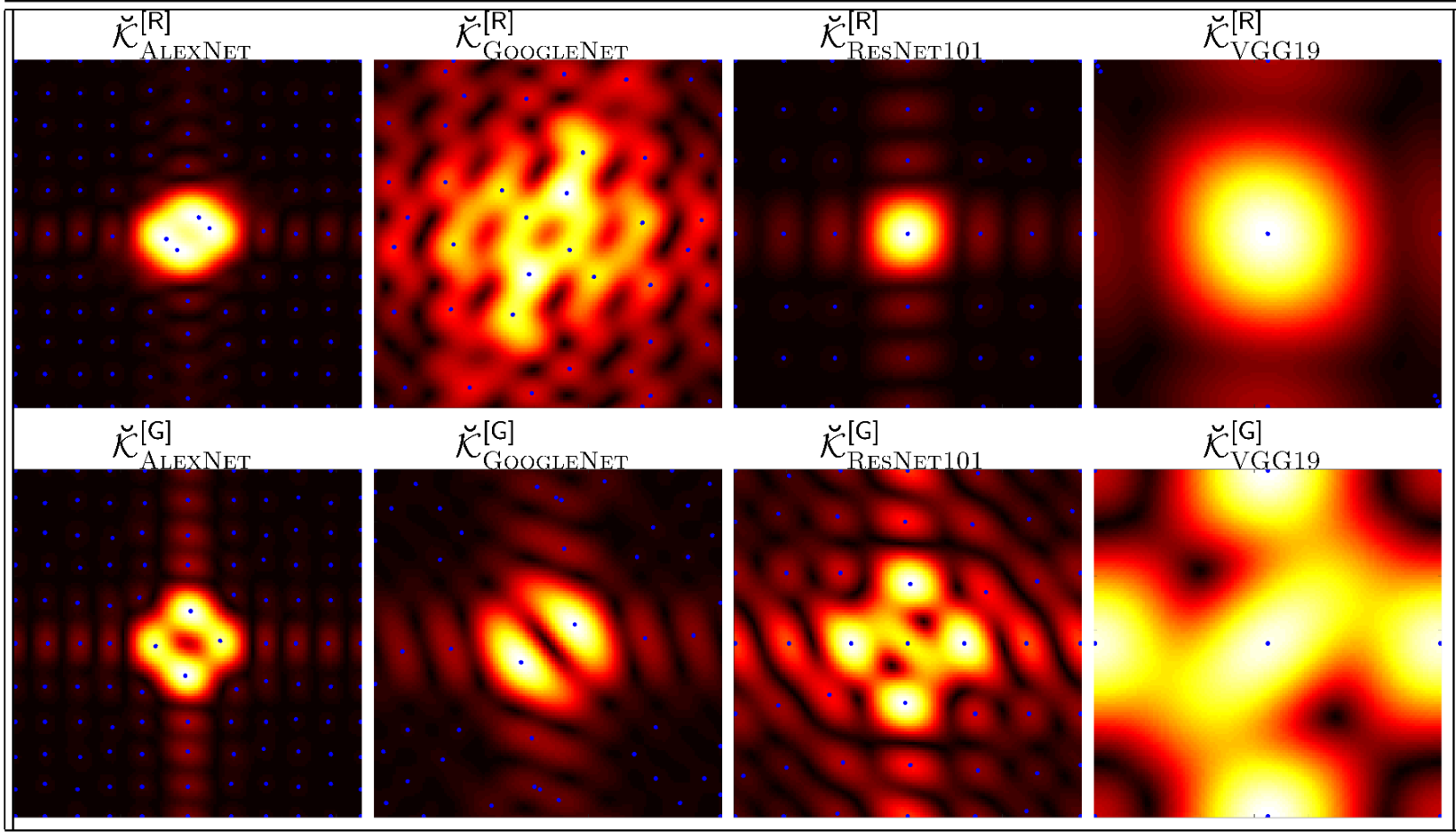
3. **CNN (deep concern)**: Depth $j + 1$ fused multiserial neuro-convolutional maps are:

$$\mathfrak{D}_{j+1, m_{j+1}}[k, \ell] = \Upsilon_{j+1, m_{j+1}} \left(-\theta_{j+1, m_{j+1}} + \bigoplus_{m_j=0}^{M_j-1} \sum_{p, q \in \mathbb{Z}} h_{m_{j+1}}[p, q] \mathfrak{e}_{j, m_j} \mathfrak{D}_{j, m_j}[s_{j+1}k - p, t_{j+1}\ell - q] \right) \quad (3)$$

where $m_{j+1} \in \{0, 1, 2, \dots, M_{j+1} - 1\}$.

Example of characterizations for filters discovered by CNN

$|H_{1,n}^c|$ for⁴ some convolution kernels having largest numbers of regional maxima.



⁴ \check{K} denotes the convolution kernel series, $H_{1,n}^c$ is the Fourier transform of $h_{1,n}^c$, $c \in \{\text{Red, Green, Blue}\} = \{[R], [G], [B]\}$

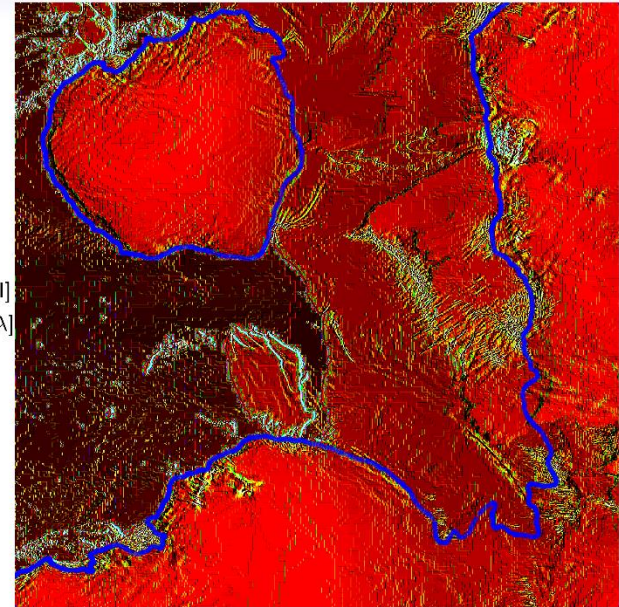
Best convkernel $h \in \check{\mathcal{K}}_\lambda$?



Landsat 8 OLI

Antarctic ice-sheet

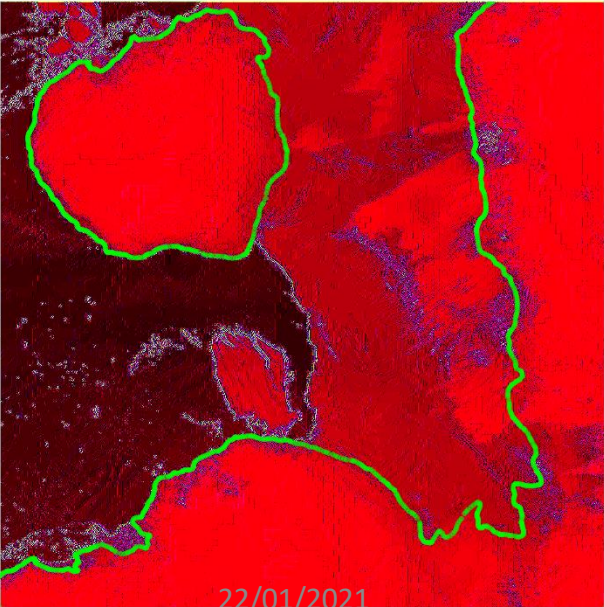
Rank#1 $h \in \check{\mathcal{K}}_\lambda$ where $\lambda \in \{\text{ALEXNET}, \text{GOOGLENET}, \text{RESNET101}, \text{VGG19}\}$



$$\mathcal{M}_{\check{\mathcal{K}}_{\lambda(m)}}[I] = 1 * h_m[\lambda]$$

$\lambda = \text{VGG19}$
 $m = 59 \Rightarrow h =$
 $\left\{ \begin{array}{l} \text{Distortlet} \\ \text{Differencelet} \\ \text{Distortlet} \end{array} \right\}$

Rank#2 $h \in \check{\mathcal{K}}_\lambda$ where $\lambda \in \{\text{ALEXNET}, \text{GOOGLENET}, \text{RESNET101}, \text{VGG19}\}$



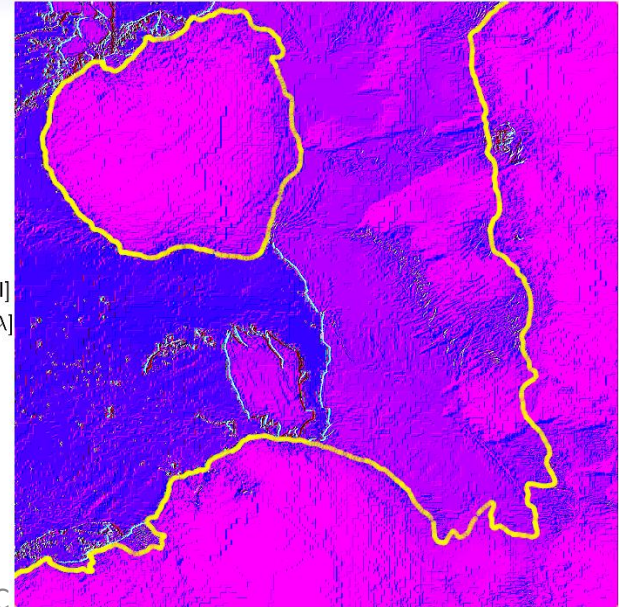
$$\mathcal{M}_{\check{\mathcal{K}}_{\lambda(m)}}[I] = 1 * h_m[\lambda]$$

$\lambda = \text{RESNET101}$
 $m = 22 \Rightarrow h =$
 $\left\{ \begin{array}{l} \text{Differencelet} \\ \text{Distortlet} \\ \text{Differencelet} \end{array} \right\}$

22/01/2021

Rencontre LAPP-LISTIC

Rank#3 $h \in \check{\mathcal{K}}_\lambda$ where $\lambda \in \{\text{ALEXNET}, \text{GOOGLENET}, \text{RESNET101}, \text{VGG19}\}$



$$\mathcal{M}_{\check{\mathcal{K}}_{\lambda(m)}}[I] = 1 * h_m[\lambda]$$

$\lambda = \text{VGG19}$
 $m = 52 \Rightarrow h =$
 $\left\{ \begin{array}{l} \text{Distortlet} \\ \text{Distortlet} \\ \text{Distortlet} \end{array} \right\}$

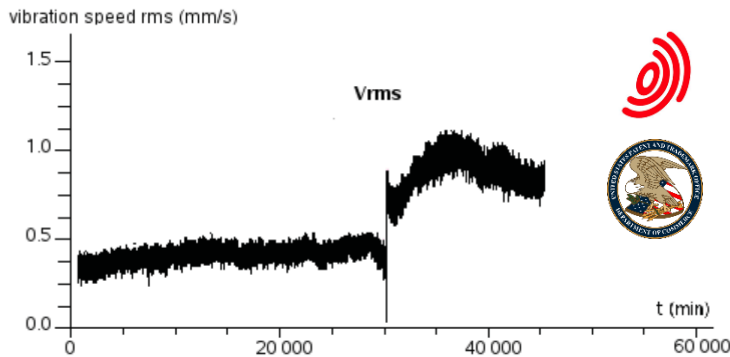
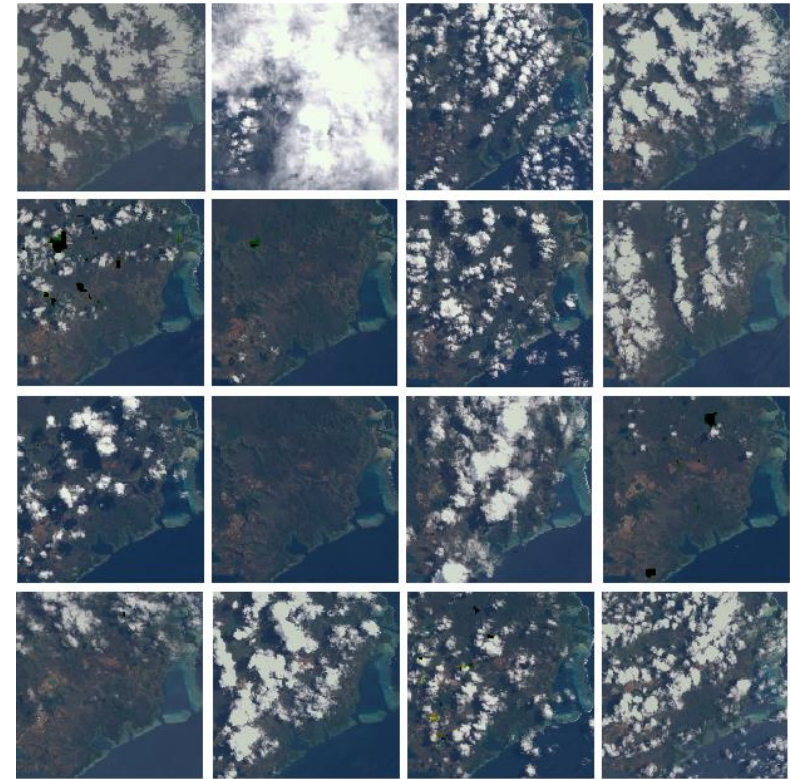
Extraction de connaissances dans les données LISTIC

fouille de motifs séquentiels



description prévision

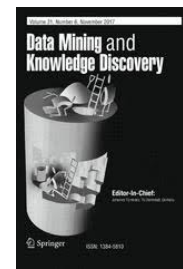
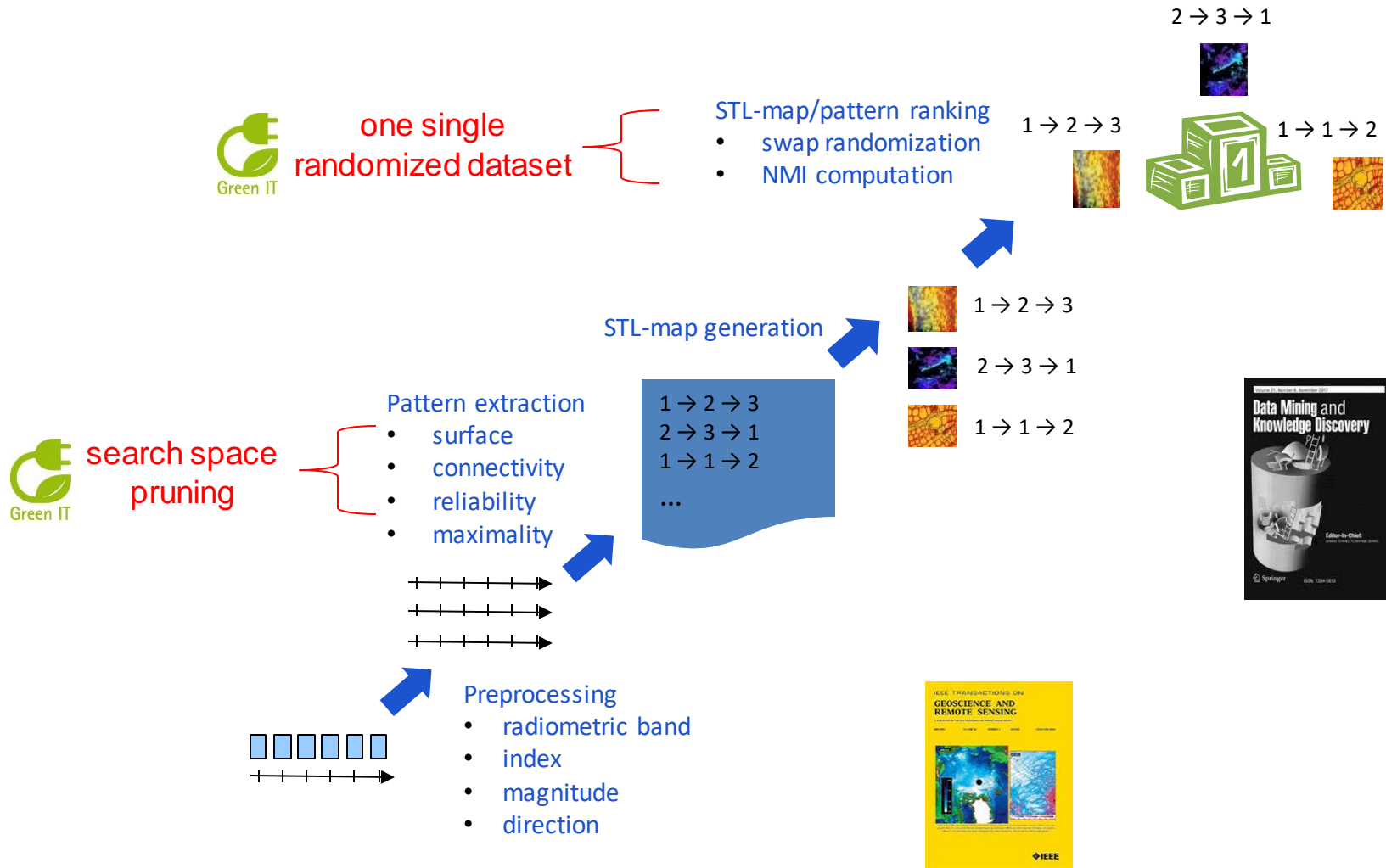
complétude, justesse, signification statistique, interprétabilité



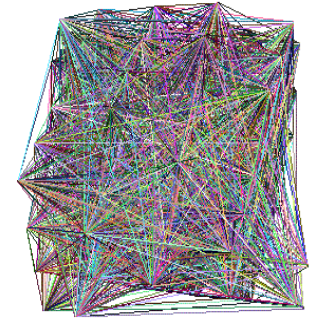
adixen
by PFEIFFER VACUUM



Satellite Image Time Series



Large graphs monitoring



Graph monitoring is the process of deciding if a local change will lead to global changes or not ?

- Large set of applications
 - Computer Networks, biology, social networks , complex systems

How can we know if a local change is scaling into global?

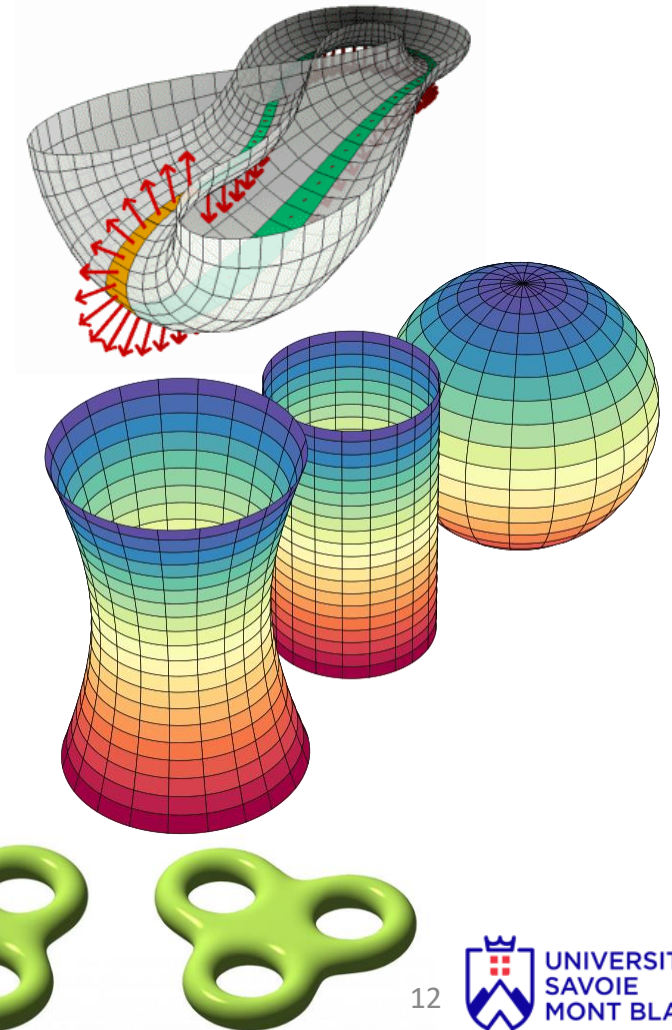
- Complex relation between intrinsic factors (the graph structure) and extrinsic ones (the change)

Gauss question:

- Do an ant moving on a shape can figure out what is the shape ?

Gauss-Bonnet theorem

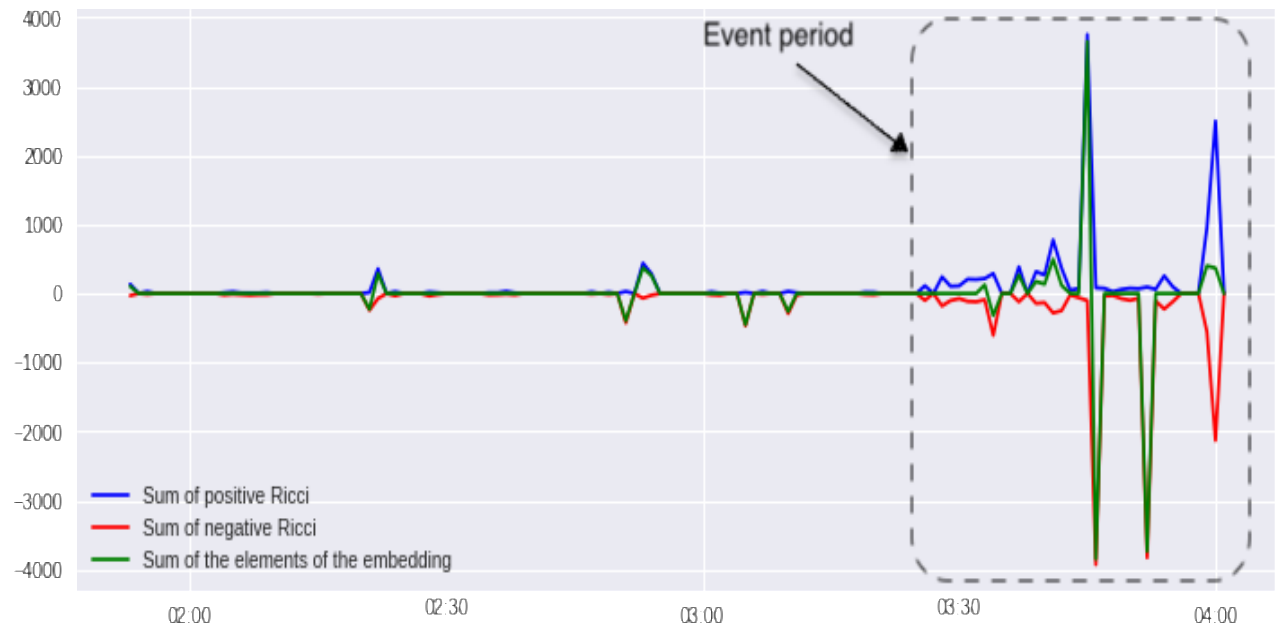
$$\int_M K dA + \int_{\partial M} k_g ds = 2\pi\chi(M).$$



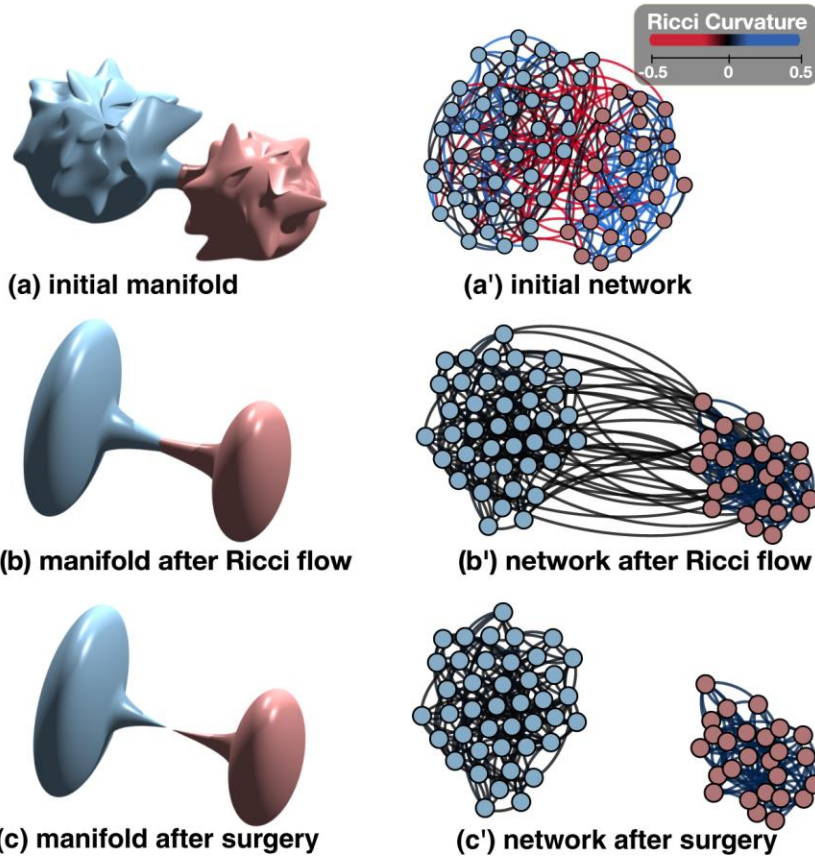
Ollivier-Ricci Curvature monitoring system

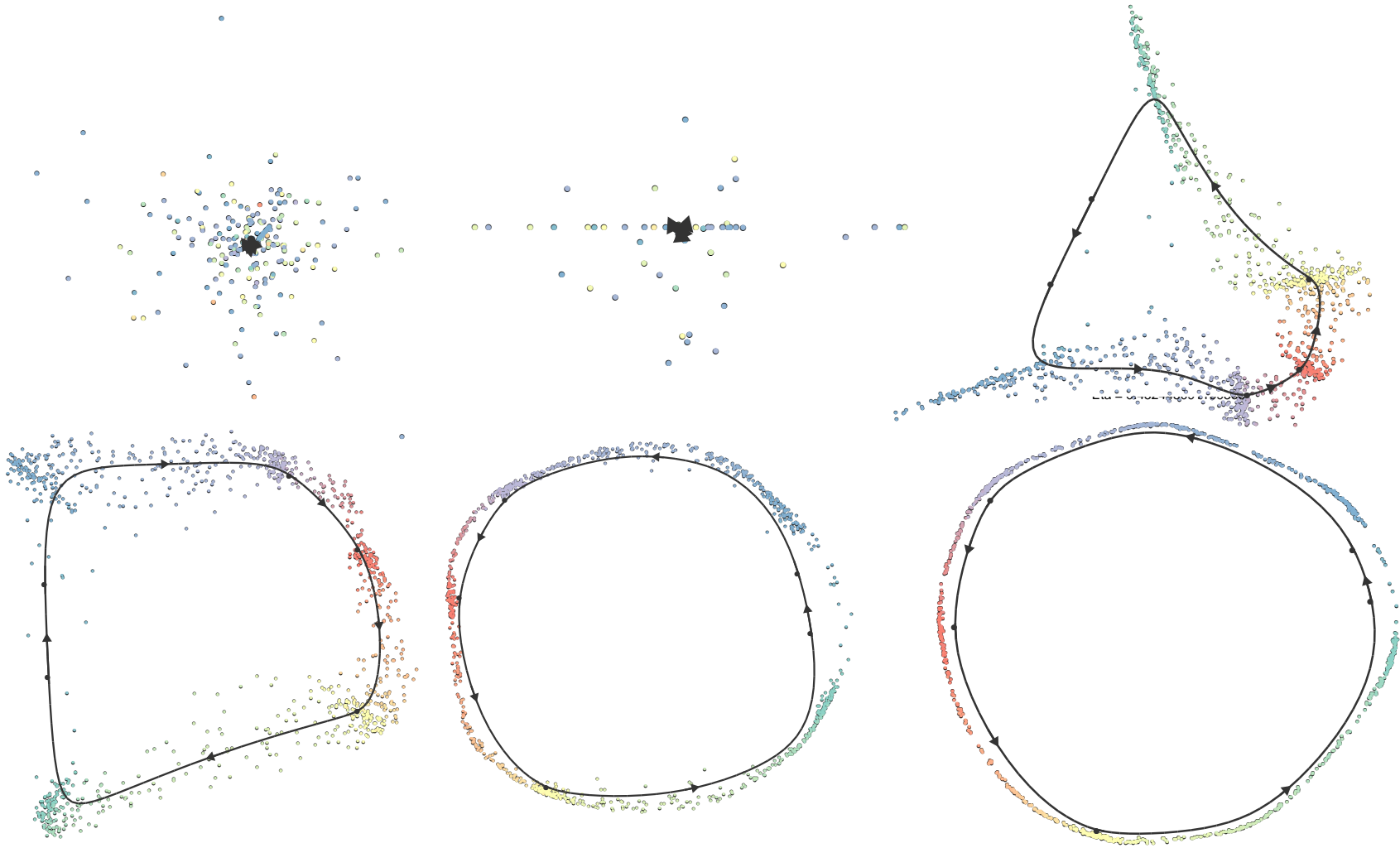
Compare the Ollivier- Ricci between all nodes of two snapshots of the graph
Evaluate the importance of the change by the magnitude of the change
Is Gauss-Bonnet theorem is valid ?

- Seems to hold for self-healing networks



Ricci flow and clustering

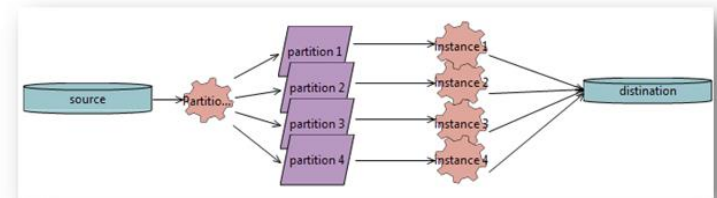
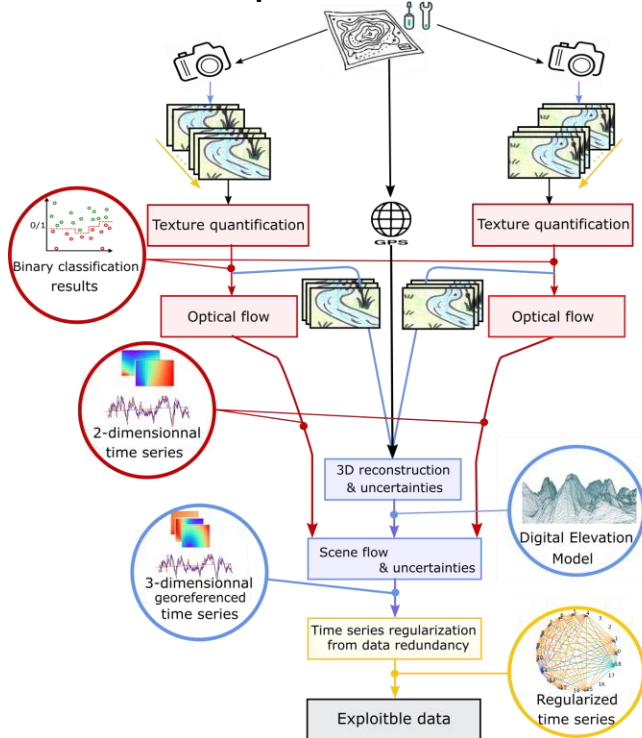




Distribution : les chaînes de traitements

Chaîne de traitements :

- Suite de traitements de données modélisée par un DAG
- Modélisation distribuée d'un traitement
- Composition de chaînes de traitements



Problématiques des chaînes de traitements :

- Description « user-friendly » des chaînes
- Gestion de la distribution
- Distribution des tâches et données de manière conjointe
- Recherche de l'optimalité de distribution
- Monitoring
- ...



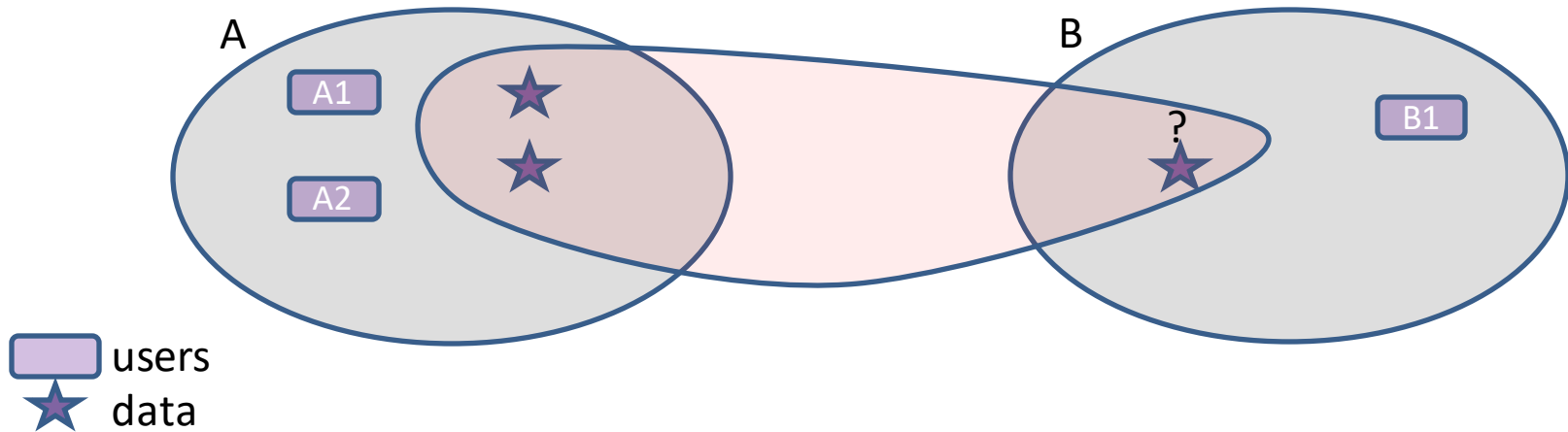
Data management for large-scale distributed systems

- Motivations
 - Massive data
 - Need for reliable, efficient and consistent data-management systems
=> What about the CAP theorem ?
- Targeted architectures
 - Data center/clouds
 - Fog
- Problems / challenges
 - Large scale (high number of nodes, huge volumes of data)
 - Dynamicity (hardware and software)
 - Virtualization (resources fragmentation)

Data management – Main contributions

- Key mechanism : data replication
 - Handling multiple copies of the same piece of data
- Three main goals
 - Fault tolerance (data durability/availability)
 - Access performance
 - Data consistency

=> Application dependant
- The « system » has to decide
 - How many data copies
 - Which consistency protocols (data types ?)
 - Data placement ?



■ Perspectives / ongoing work

- Gadget : Toward energy-aware data management systems

ANR 2021 submission (LISTIC, LIP, LIP6, CELESTE)

Conclusion

■ Spécificité AFuTE

- ❑ Apprendre sur des données temporelles et/ou imparfaites en développant des travaux méthodologiques basés sur une riche palette théorique pour des applications en télédétection.

■ Spécificité ReGaRD

- ❑ S'appuyer sur des connaissances multidisciplinaires pour construire des systèmes informatiques fiables et rapides afin de répondre aux besoins de l'humain.