

## Étude et classification des méthodes de compression par extraction de motifs et k2-trees

**Nom et Prénom** : **Hafsa Bousbiat**

**Ihadadene Sana**

**Spécialité** : **Systèmes Informatiques**

**Organisme d'accueil** : **Laboratoire LCSI**

**Encadré par** : **Mme Amrouche (ESI)**  
**Mme Seba (Univ Lyon1)**

### Laboratoire LCSI

Laboratoire de la Communication dans les systèmes Informatiques (LCSI) : Le laboratoire LCSI a été créé en 2010 suite à l'arrêt ministériel n93. Il est structuré en plusieurs équipes de recherche traitant des thématiques diverses qui sont en rapport avec la communication dans les systèmes informatiques. En 2017, cinq articles de recherches ont été publiés dans des revues internationales par le laboratoire avec 11 communications internationales et une communication nationale.

**Création** 25 Mars 2010 (9 ans)

**Forme juridique** Laboratoire de Recherche

**Siège social** Oued Smar, Alger

**Direction** D.E Zegour

**Financement** État algérien (100 %)

**Activité** Encadrement de thèse de doctorat/Master,  
Séminaire, Formation, ...

**Partenaires** USTHB, univ. Chlef, univ Toulouse, UPEC, U. Lyon, ...

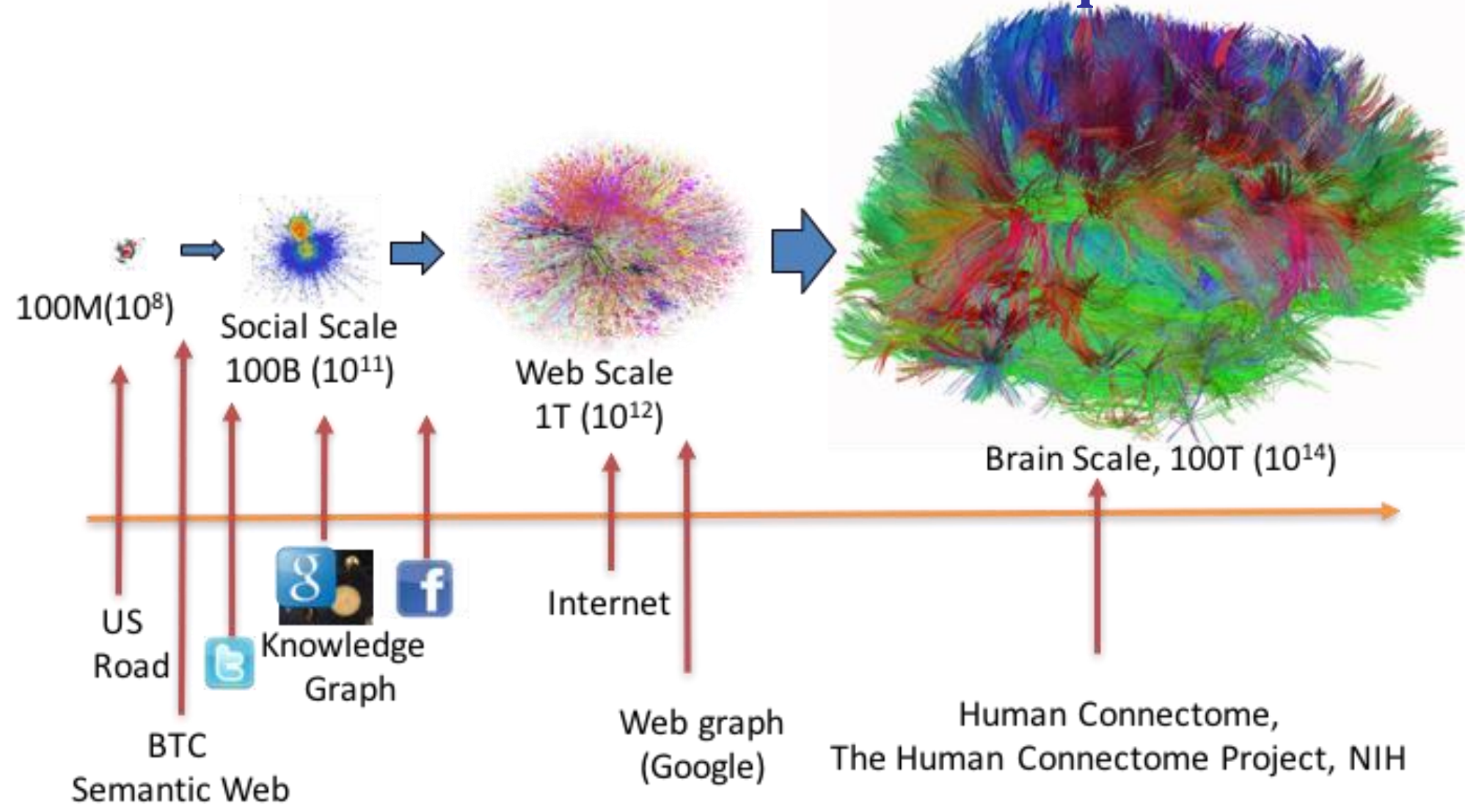
**Nombre d'équipe de recherche** 5

**Effectif** 20 chercheurs permanents, 60 doctorants

**Site web** lcsi.esi.dz



### Contexte Motivation et Problématique



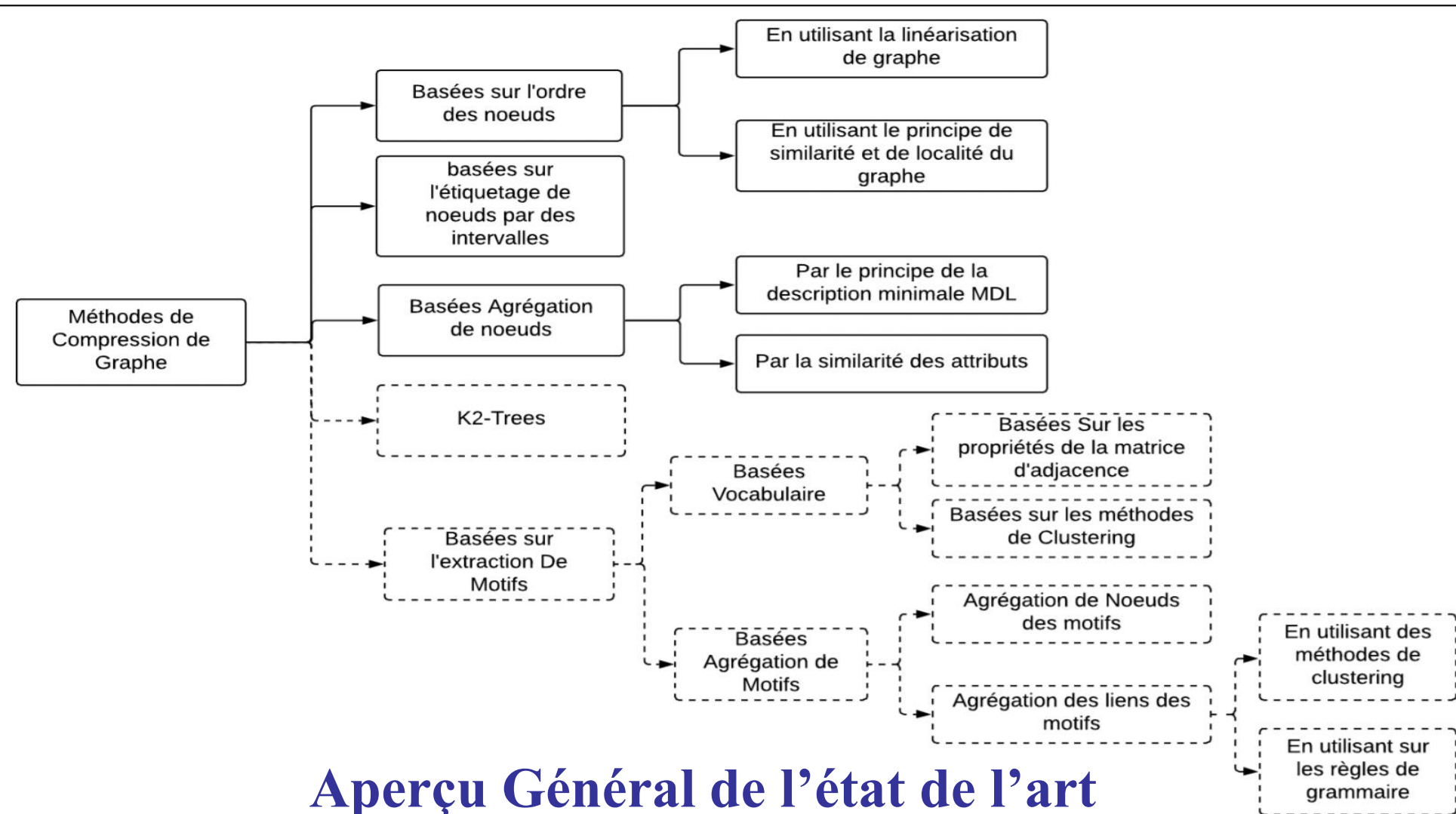
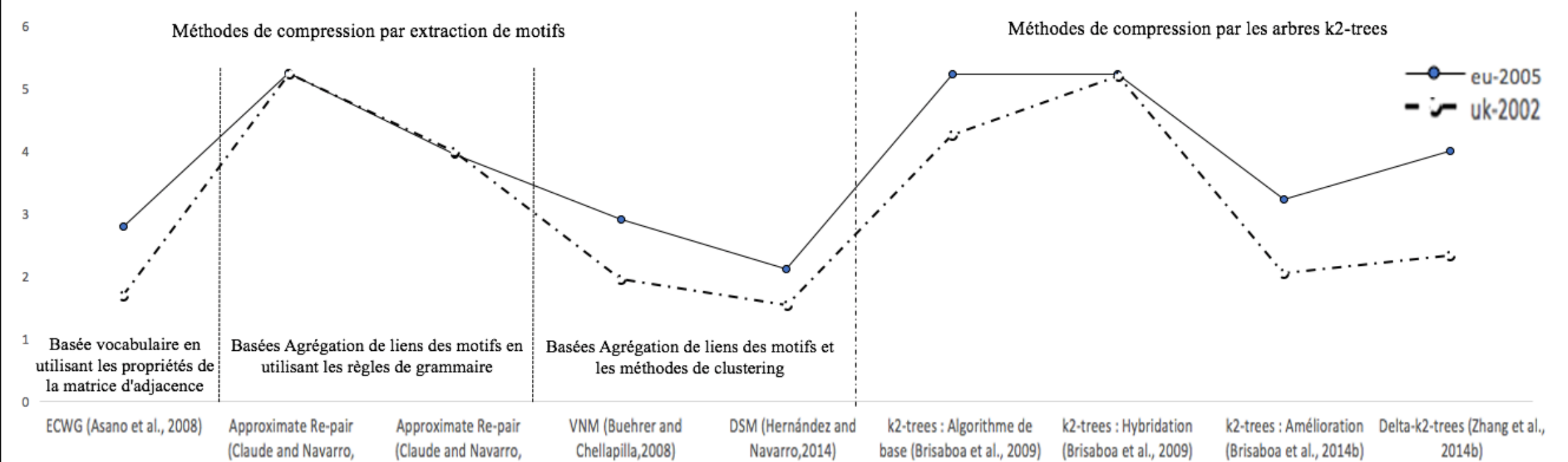
#### Problématique

- La puissance des processeurs augmente plus vite que les capacités de stockage : un déséquilibre entre le volume des données qu'il est possible de traiter et de stocker.
- Capacité mémoire : les graphes ne peuvent plus résider en mémoire vu la quantité d'informations qui modélisent de nos jours.
- Les algorithmes classiques de traitement des graphes ne sont plus adaptés

→ **Comment gérer cette grande masse de données tout en permettant une exécution efficace des algorithmes classiques de traitement des graphes?**

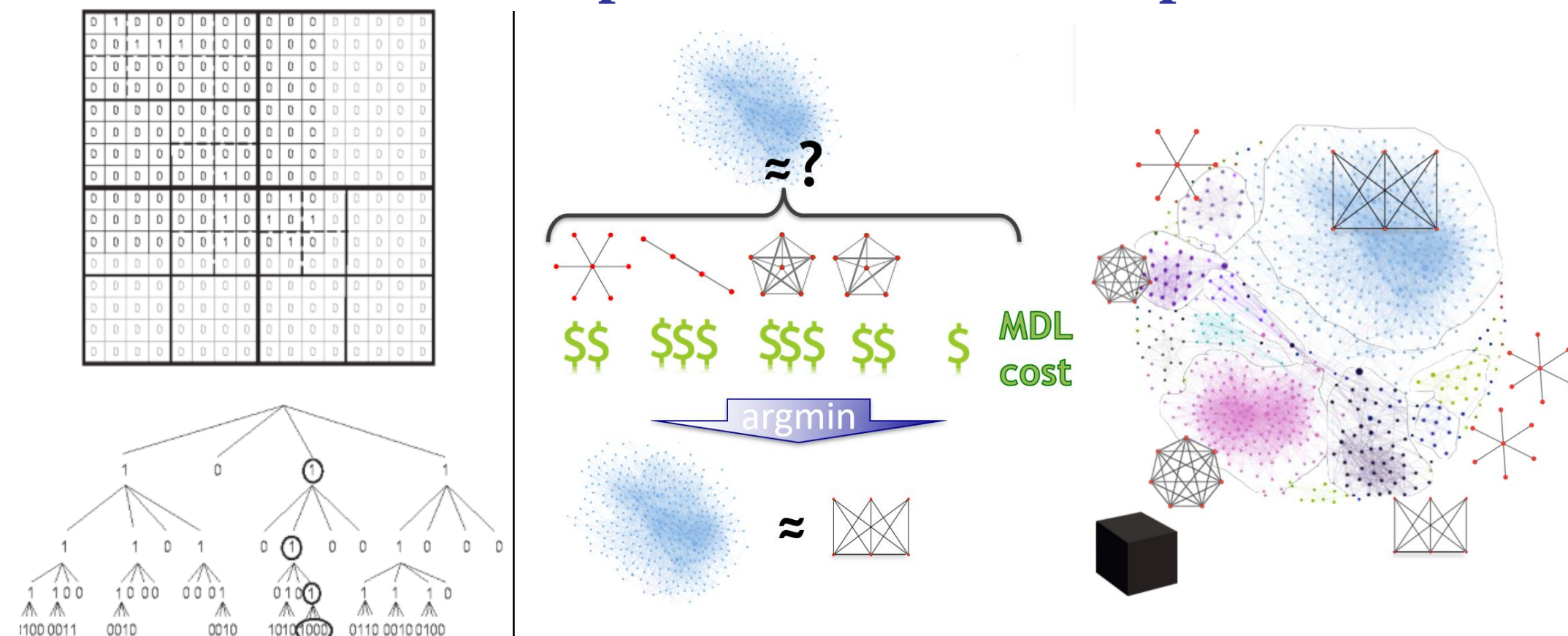
### Analyse des solutions existantes et comparaison

			Graphe En entrée				Type de Compression		Structure en sortie		
			Statique	Dynamique	Attribué	Étiqueté	Sans Perte	Avec Perte	Succinte	Structurelle	
Méthode basées sur les arbres k2-trees											
Méthode basées sur l'extraction de motifs	Basées Vocabulaire	Basées sur les propriétés de la matrice d'adjacence									
		Basées sur les méthodes de clustering									
	Basées Agrégation des motifs	Agrégation des nœuds des motifs									
		Agrégation des liens des motifs	En utilisant des règles de grammaire								
			En utilisant les méthodes de clustreing								



### Aperçu Général de l'état de l'art

### Focus sur une ou plusieurs solutions importantes



- K2-tree est une méthode de compression sans perte qui appliqué sur des graphes statiques non pondérés, elle est destinée pour les graphes du web. Elle exploite les propriétés de la matrice d'adjacence pour réduire la taille du graphe.
- VoG est une méthode de compression sans perte basé sur les méthodes de clustering, qui utilise le principe MDL pour choisir un ensemble de structures appelé modèle permettant d'obtenir une représentation succincte du graphe initial.

#### Reference des methodes

- Brisaboa, N. R., Ladra, S., and Navarro, G. (2009). k 2-trees for compact web graph representation. In International Symposium on String Processing and Information Retrieval, pages 18–30. Springer

### Perspectives

#### A court terme

- Conception d'un outil de compression de graphe par extraction de motifs et k2-trees.
- Implémentation de plusieurs méthodes existantes.
- Evaluation des performances de chaque approche en utilisant des benchmarks connus.
- Comparaison entre les méthodes de différentes classes.

#### A long terme

Développement de nouvelles méthodes de compression compétitives en temps d'exécution et de taux de compression tout en se basant sur l'extraction de motifs et les arbres k2-trees.

### Références

#### Bibliographie

- Brisaboa, N. R., Ladra, S., & Navarro, G. (2009, August). k 2-trees for compact web graph representation. In *International Symposium on String Processing and Information Retrieval* (pp. 18-30). Springer, Berlin, Heidelberg.
- Koutra, D., Kang, U., Vreeken, J., and Faloutsos, C. (2015). Summarizing and understanding large graphs. *Statistical Analysis and Data Mining : The ASA Data Science Journal*, 8(3) :183–202.
- Liu, Y., Safavi, T., Dighe, A., and Koutra, D. (2018a). Graph summarization methods and applications : A survey. *ACM Computing Surveys (CSUR)*, 51(3) :62.