

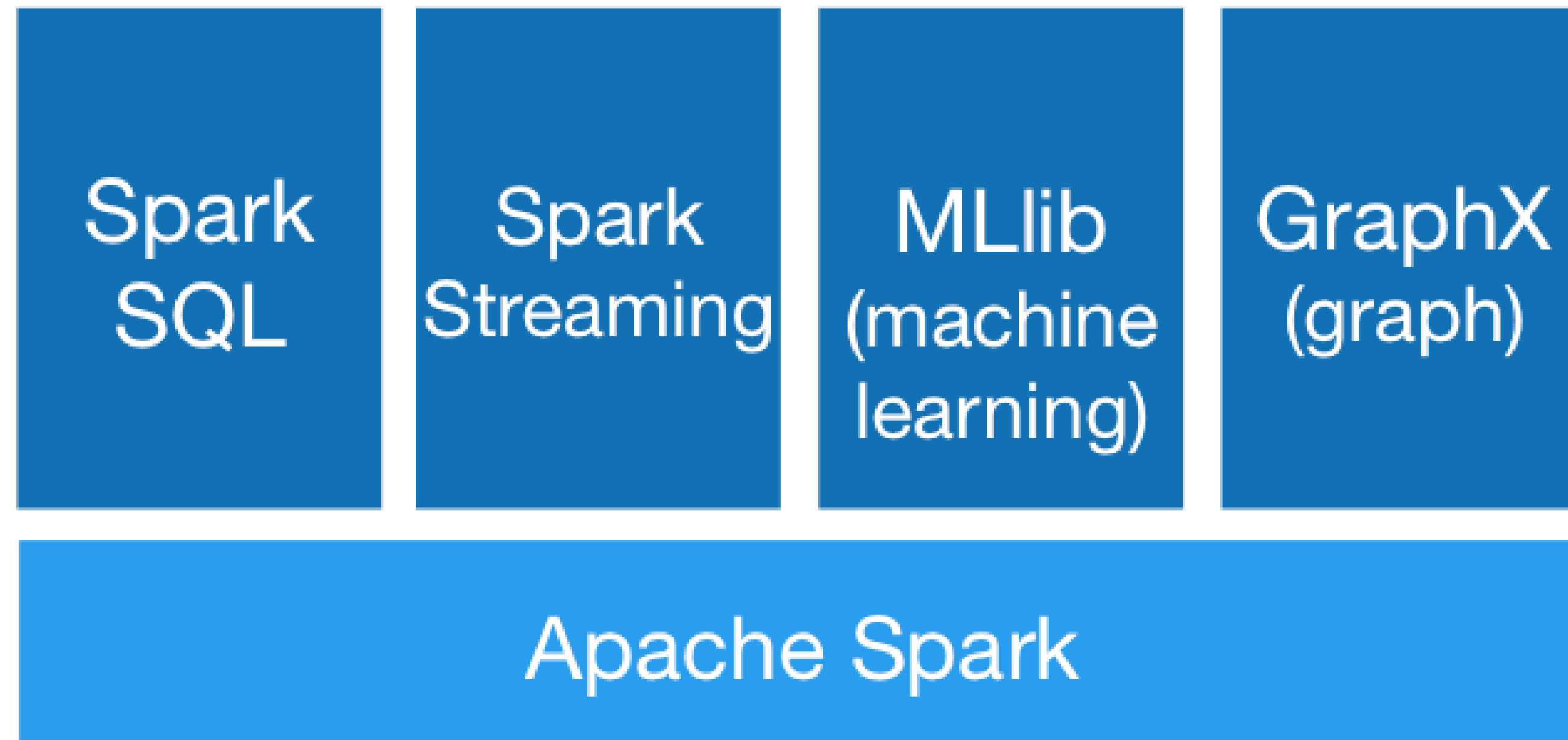


Big Data Analytics

« Pilotage de la performance pour une bonne gouvernance des entreprises »

CHAPITRE 6 – Spark GraphX

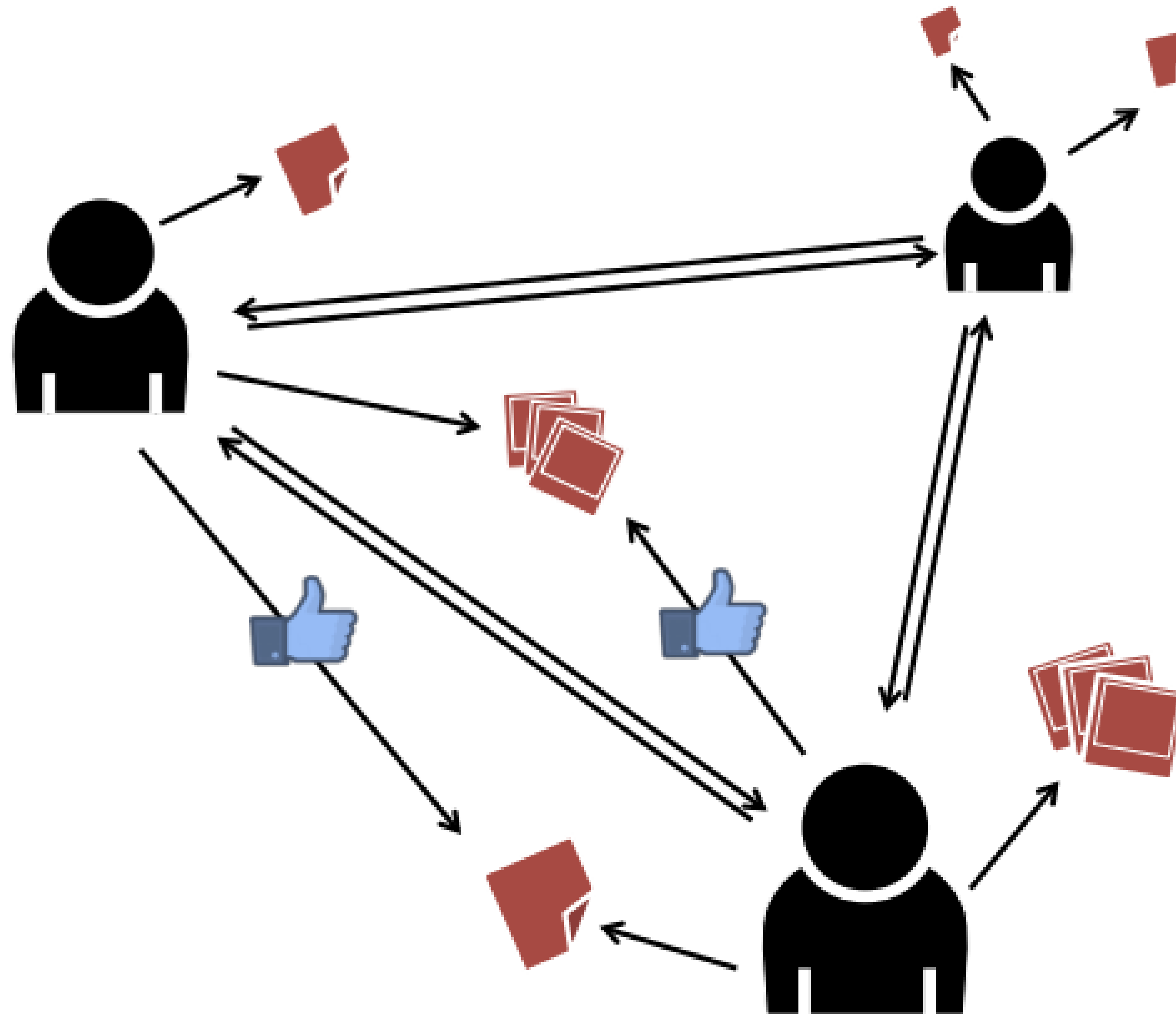
Spark Stack



- **Spark SQL** : pour le traitement de données (SQL et non structuré)
- **Spark Streaming** : traitement de flux de données en direct (live streaming)
- **MLlib** : Algorithmes Machine Learning
- **GraphX** : Traitement de graphes

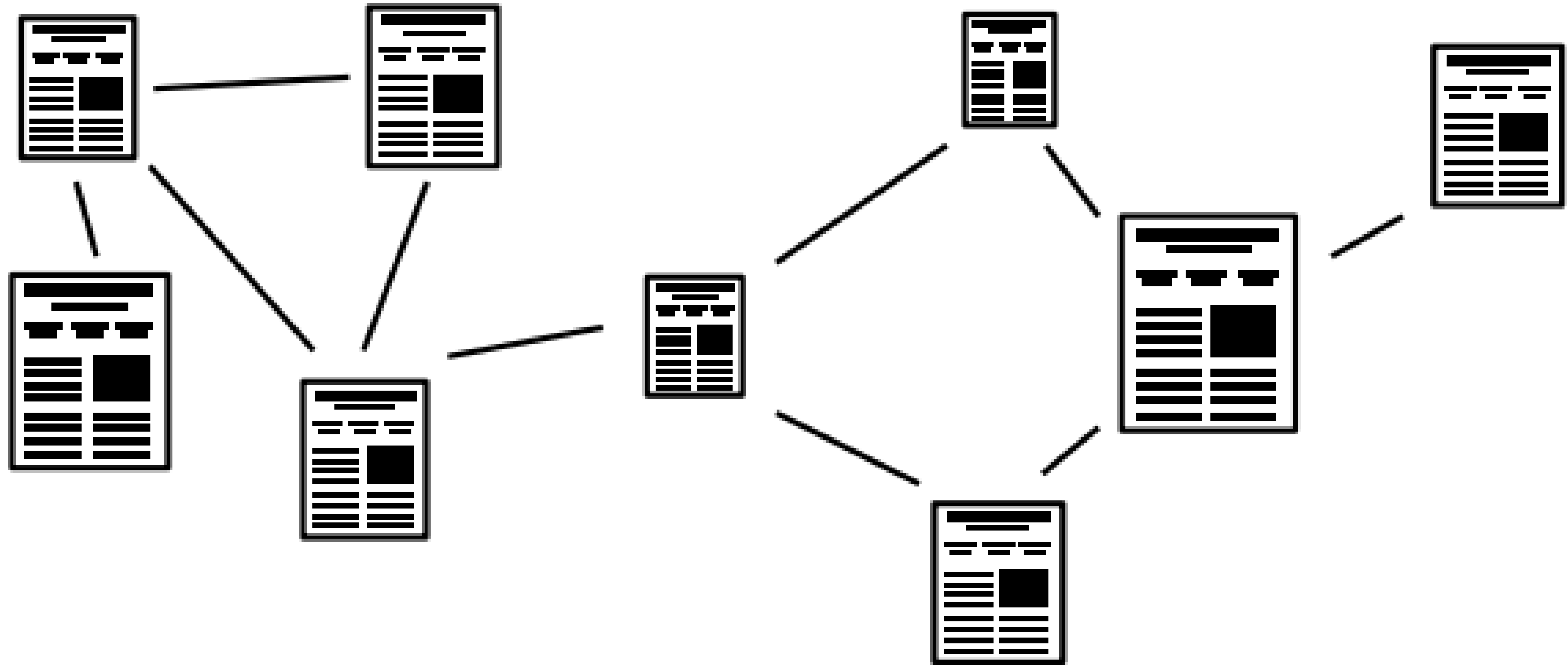
Graphes

Réseaux sociaux



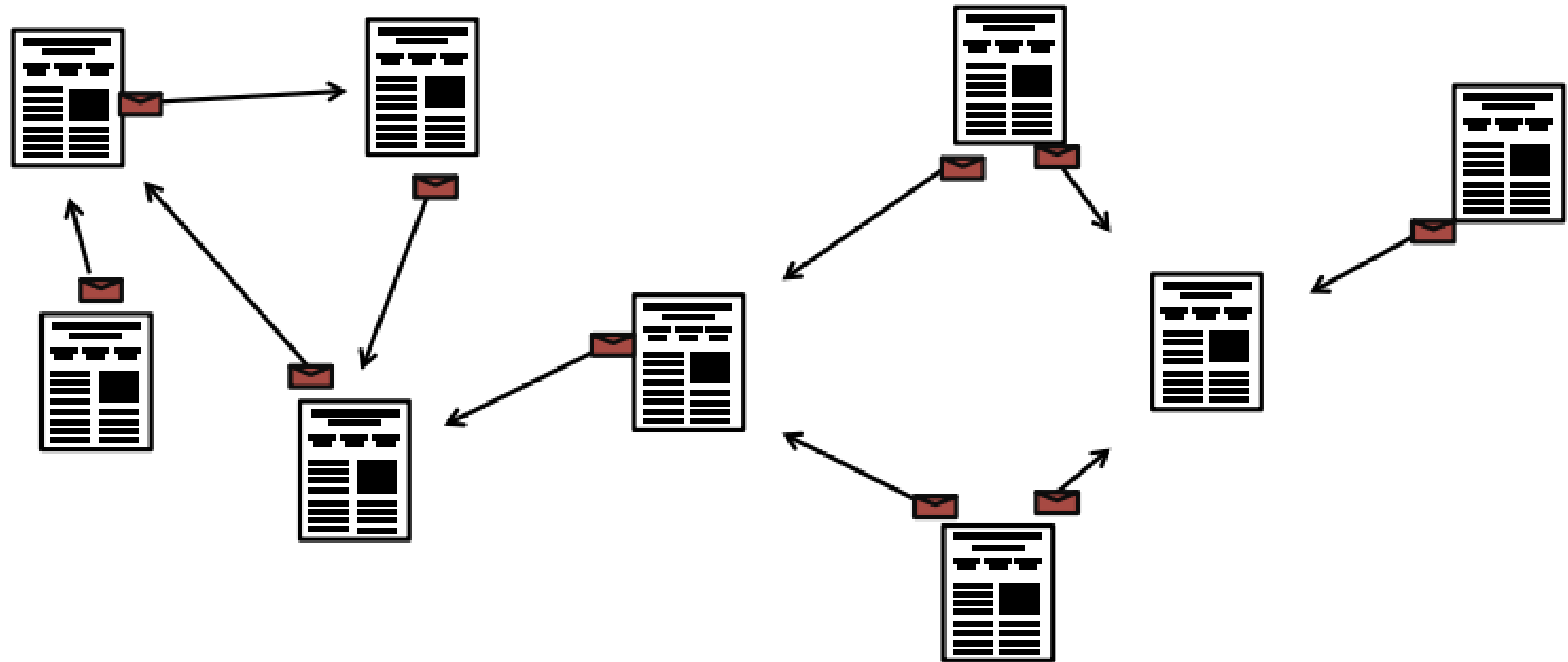
Graphes

Graphes Web



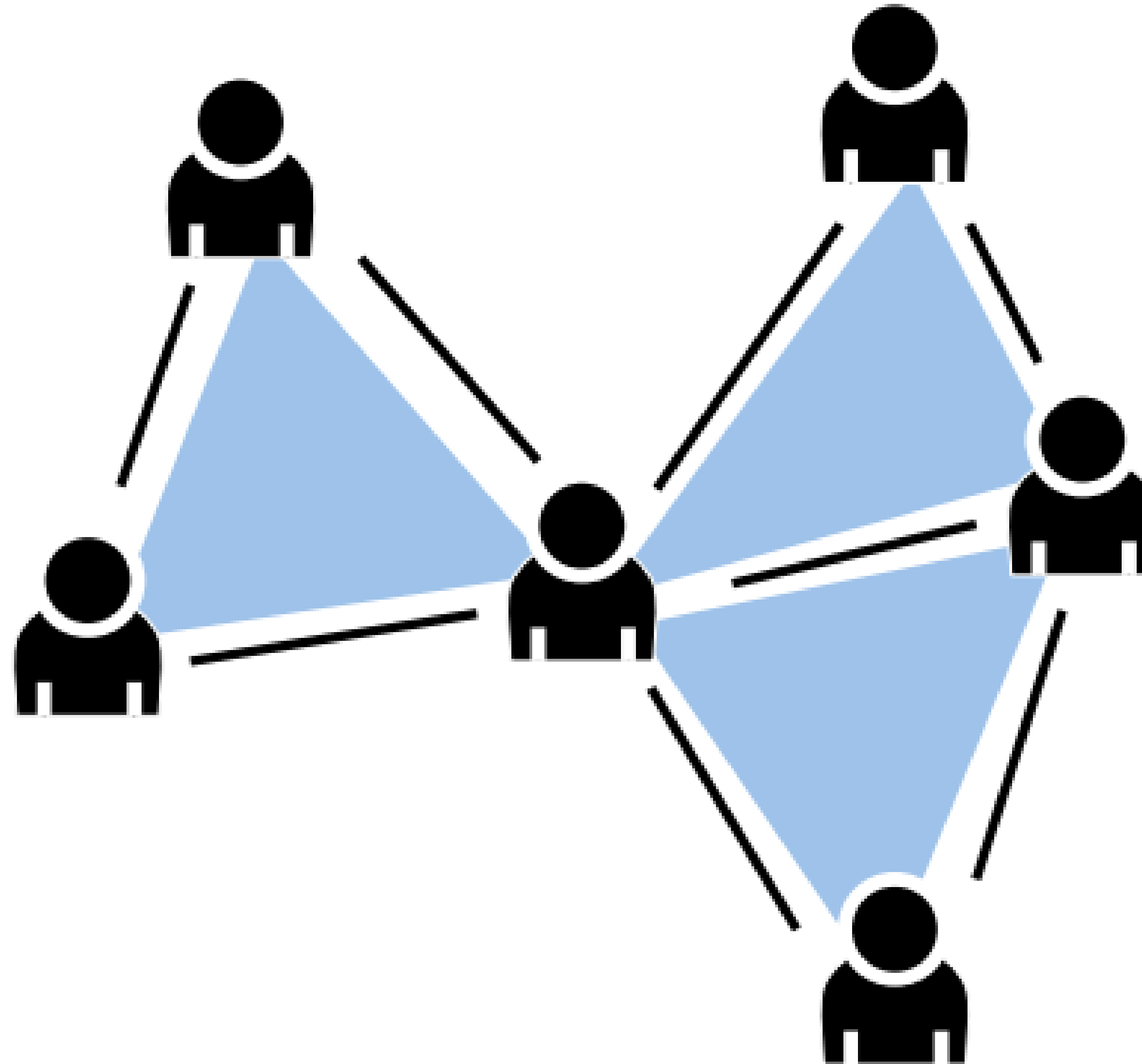
Algorithmes pour les graphes

Page Rank

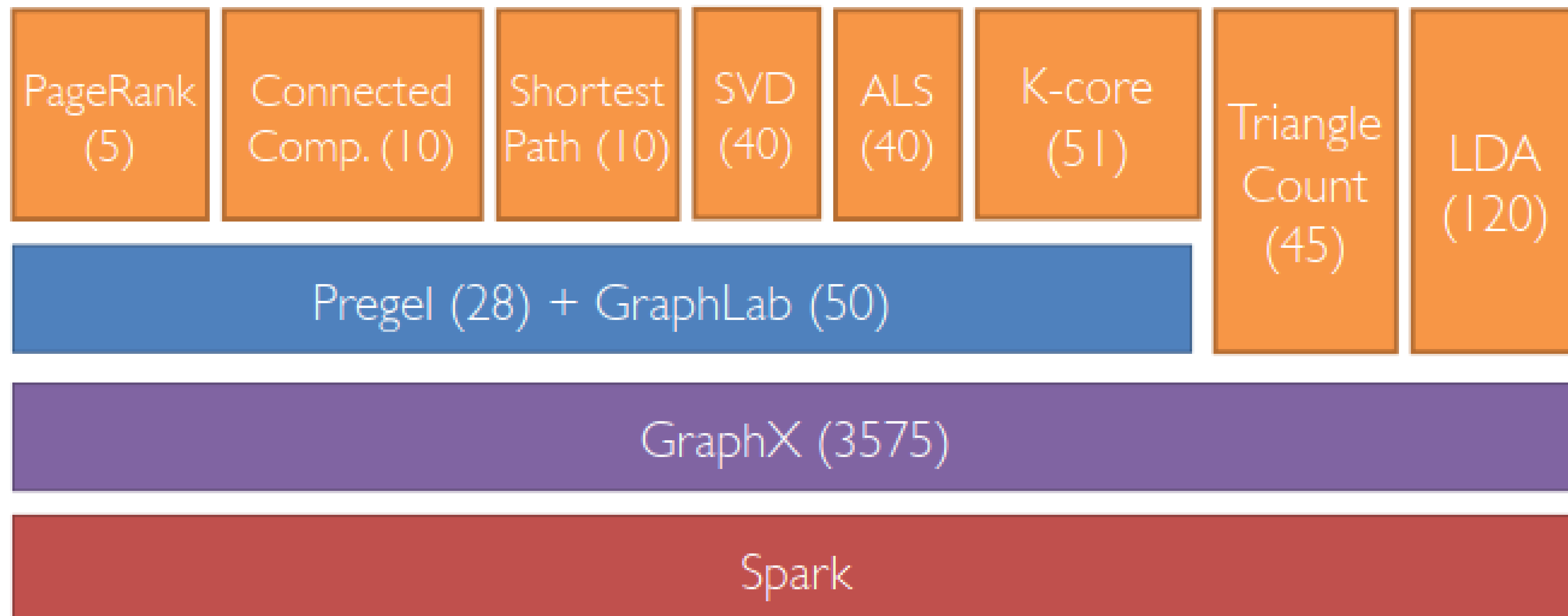


Algorithmes pour les graphes

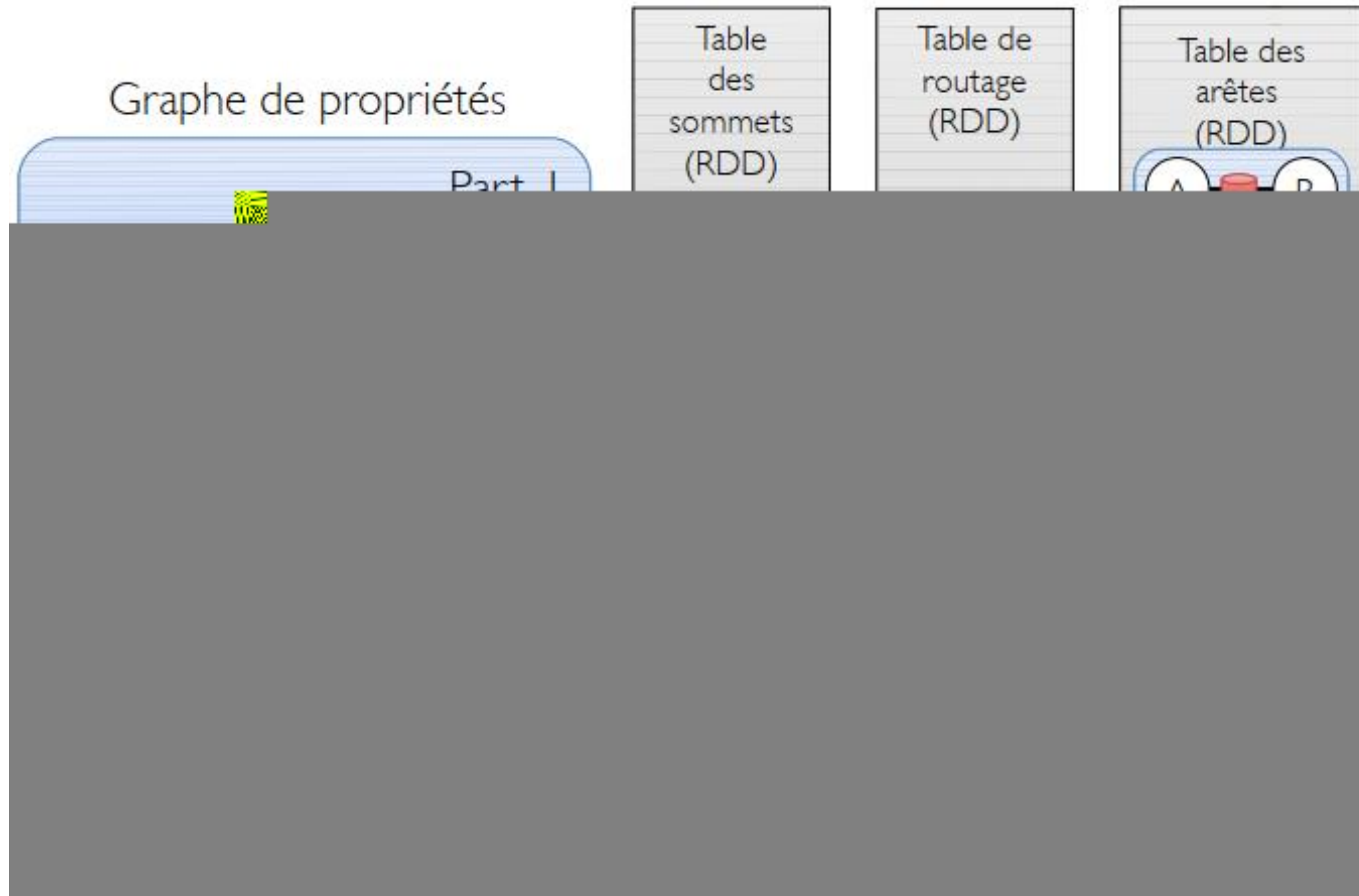
Triangle Counting



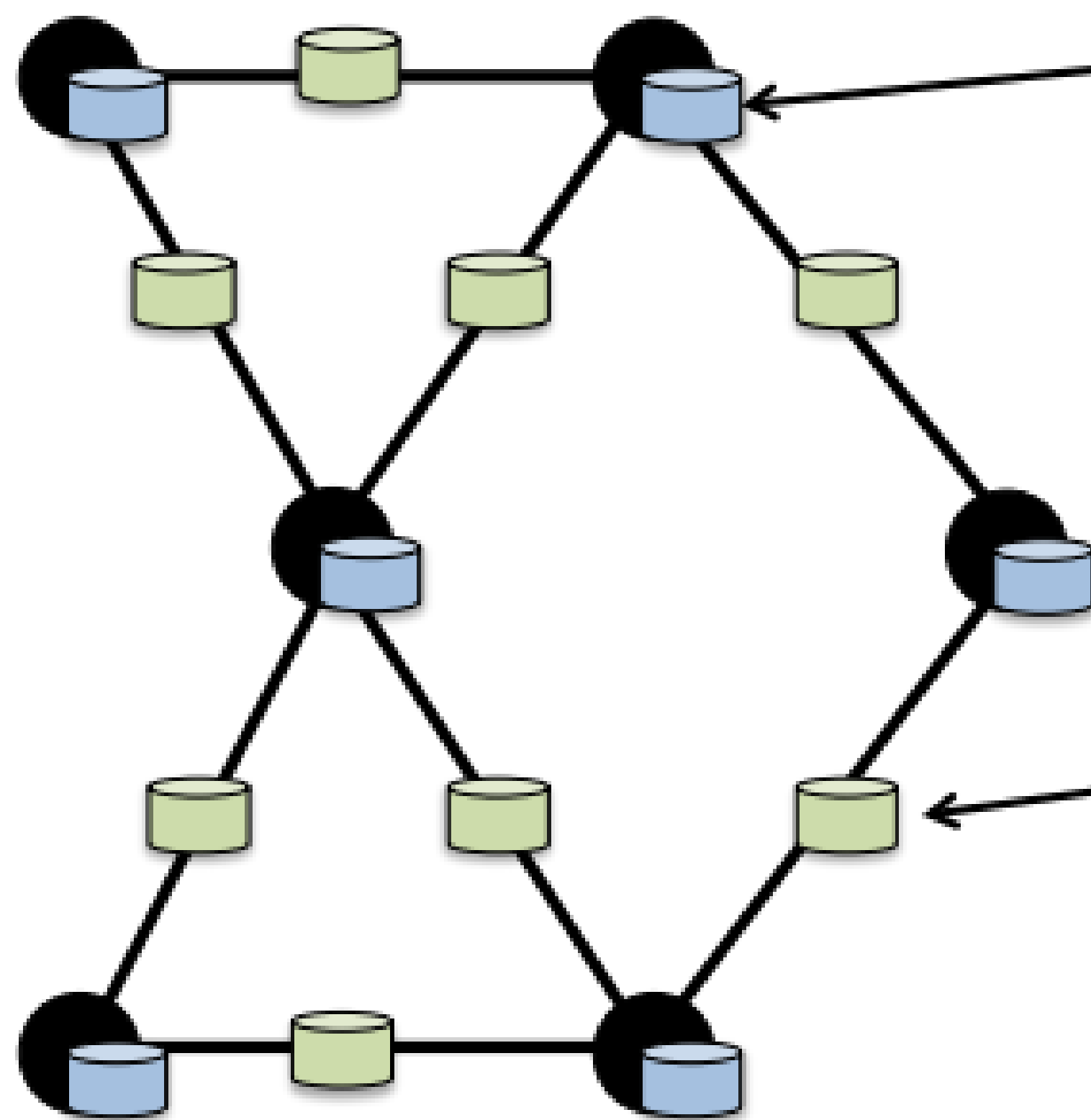
Spark GraphX



Graphe distribué



Graphe distribué



Vertex Property:

- User Profile
- Current PageRank Value

Edge Property:

- Weights
- Relationships
- Timestamps