

## Contexte

- Ensemble d'utilisateurs
- Caractéristiques
- Calcul potentiel

## Objectifs

- Regrouper les utilisateurs par caractéristiques communes
- Répartir le calcul

## Contraintes

- Informations privées
- Limitation de la communication

1 Modélisation

2 Résolution

# Notations

- Réseau des utilisateurs  $G = (V, E)$ ,  $V = \{1, \dots, N\}$
- Caractéristiques  $\mathbf{X} = (X_i)_{1 \leq i \leq N}$   
ex : SMS, a écouté [XXX], a partagé [YYY]
- Distance entre utilisateurs :

$$D : \begin{cases} V \times V & \rightarrow \mathbb{R}_+ \\ (x, y) & \mapsto D(x, y) \end{cases}$$

$$D(x, y) = 0 \iff x = y$$

$$D(x, y) = \|x - y\|_p$$

# Formulation

- Objectif : partitionner par affinités
- Problème à résoudre :

$$\min_{\mathcal{P}} W(\mathcal{P}) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq N} D(X_i, X_j) \Phi_{\mathcal{P}}(i, j),$$

1 Modélisation

2 Résolution

# Contraintes

- Contraintes supplémentaires :
  - Distribution du calcul
  - Limitation du transfert de données
- Problème d'optimisation distribuée

## Énoncé

Nouvelle formulation du problème :

$$\min_{\mathcal{P}} W(\mathcal{P}) = \frac{1}{n} \sum_{1 \leq i \leq N} f_i(\mathbf{X}),$$

où

$$f_i(\mathbf{X}) = \frac{1}{n} \sum_{1 \leq j \leq N} D(X_i, X_j) \Phi_{\mathcal{P}}(i, j)$$

- Principe
  - Utilisateur  $i$  effectue seul le calcul de  $f_i$
  - Communication possible avec le voisinage
- Intérêt
  - Problème décomposé suivant chaque utilisateur
  - Nombreuses méthodes si problème convexe
  - Possibilité d'optimiser en ligne
- Challenges
  - $f_i$  non calculable par  $i$
  - Estimation de  $f_i$