

1 Objectifs

2 Génération aléatoire de graphe

Objectifs

- Modélisation des données à l'aide d'un graphe
 - Nœuds : utilisateurs
 - Arrêtes : contact, influence, etc.

Objectifs

- Modélisation des données à l'aide d'un graphe
 - Nœuds : utilisateurs
 - Arrêtes : contact, influence, etc.
- Intérêt pour la notion de communauté :
 - Suggestion de contenus adaptés
 - Proposition de groupe

Objectifs

- Modélisation des données à l'aide d'un graphe
 - Nœuds : utilisateurs
 - Arrêtes : contact, influence, etc.
- Intérêt pour la notion de communauté :
 - Suggestion de contenus adaptés
 - Proposition de groupe
- Calculs sur serveur mais aussi en local
 - Puissance de calcul raisonnable
 - Mémoire limitée

1 Objectifs

2 Génération aléatoire de graphe

Motivation

- Volume de données faible

Motivation

- Volume de données faible
- Méthodes robustes

Motivation

- Volume de données faible
- Méthodes robustes
- Tests de scénarios

Modèle p_1

Idée

- Principe
 - Modèle paramétrique
 - Paramètres physiques/concrets (popularité, etc.)
 - Apprentissage des paramètres sur un réseau réel (maximum de vraisemblance)

Modèle p_1

Idée

- Principe
 - Modèle paramétrique
 - Paramètres physiques/concrets (popularité, etc.)
 - Apprentissage des paramètres sur un réseau réel (maximum de vraisemblance)
- Paramètres
 - Expansion global du graphe θ
 - Expansion d'un nœud α_i
 - Popularité d'un nœud β_i
 - Intérêt mutuel $\rho_{ij} = \rho_{ji}$ (ex : Twitter)

Modèle p_1

Vraisemblance




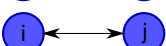
	:	λ_{ij} (normalisation)
	:	$\lambda_{ij} + \alpha_i + \beta_j + \theta$
	:	$\lambda_{ij} + \alpha_j + \beta_i + \theta$
	:	$\lambda_{ij} + \alpha_i + \alpha_j + \beta_i + \beta_j + 2\theta + \rho_{ij}$

TABLE: Log-vraisemblance en fonction des paramètres.

Modèle p_1

Exemple

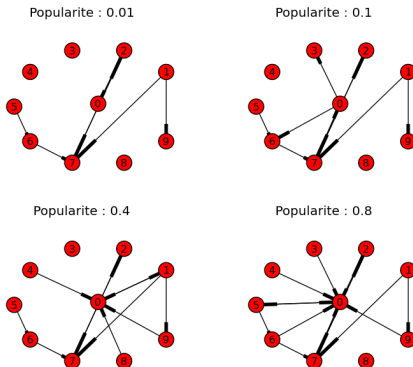


FIGURE: Modifications du graphe pour différentes valeurs de β_0 .

Modèle p_1

Discussion

Avantages du modèle

- Modèle simple
- Calcul aisé dans cas favorables
- Paramètres interprétables
- Nombreuses variantes :
 - Dynamique (chaîne de Markov)
 - Bayésienne

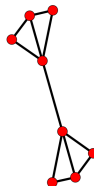
Modèle p_1

Discussion

Évaluation du modèle : quelle métrique utiliser ?

- Métriques usuelles peu pertinentes (précision)
- Pertinence de la distance directement liée à l'exploitation souhaitée des données

Graphe d'origine



Graphe reconstruit

