

# Data Mining

Amazon product co-purchasing network

Prof: M. Melançon & M. Pinaud

Réalisé par: Abdallah Ndoye

# Plan

- Problème & Motivations
- Les données Amazon
- Étude des données
- Réponses & Conclusion

# Data Mining & Secteurs d'activités

Analyse habitude consommateur : 26.8

Banque : 19.2

Santé ressource humaine : 13.1

Détection de fraude : 12.7

Finance : 11.3

Télécommunication : 10.8

# Problème & Motivations

- Peut on connaître les habitudes et préférences des clients?
- Peut on dégager des profils acheteurs?
- Comment établir des techniques marketing grâce à ces données clients?

**Comment trouver nous nos  
réponses?**

# Les Données Amazon

Données des clients Amazon

Relations entre les produit co-acheté

# Les Données Amazon

Nous avons un ensemble de 262111 sommets ainsi que 1234877 arêtes

Un graphe orienté

Nombre de noeuds de la PGCFC : 92%

Nombre d'arêtes de la PGCFC : 91%

Nombre de triplets de nœuds connectés : 717719

Coefficient de clustering moyen : 0.4240

Il s'agit d'un graphe compacte, cela implique que les produits sont très étroitement liés les uns par rapport aux autres.

# Étude des données

Construction du graphe

Comptage des motifs

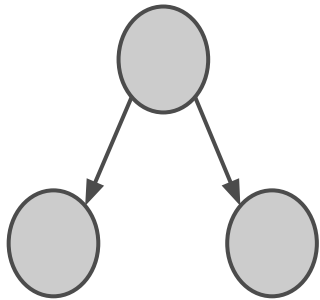
Explication et interprétation des motifs



# Comptage des motifs

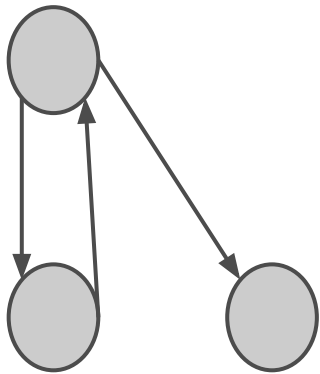
MotifId	Nodes	Edges	Count
1	3	2	131613
2	3	2	78578
3	3	3	104071
4	3	2	217566
5	3	3	16090
6	3	4	20400
7	3	4	32962
8	3	4	3685
9	3	3	2
10	3	3	135904
11	3	4	21579
12	3	5	28397
13	3	6	19319

# Les triplets de noeuds



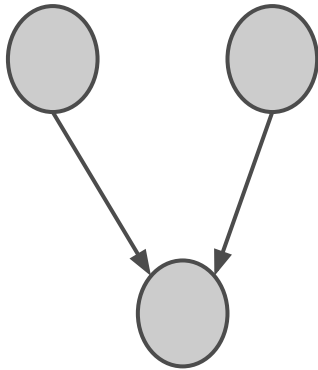
Motif n°1  
131613 occurrences

# Les triplets de noeuds



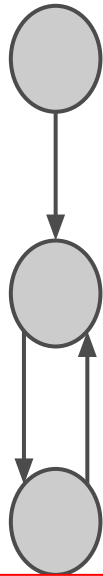
Motif n°3  
104071 occurrences

# Les triplets de noeuds



Motif n°4  
217566 occurrences

# Les triplets de noeuds



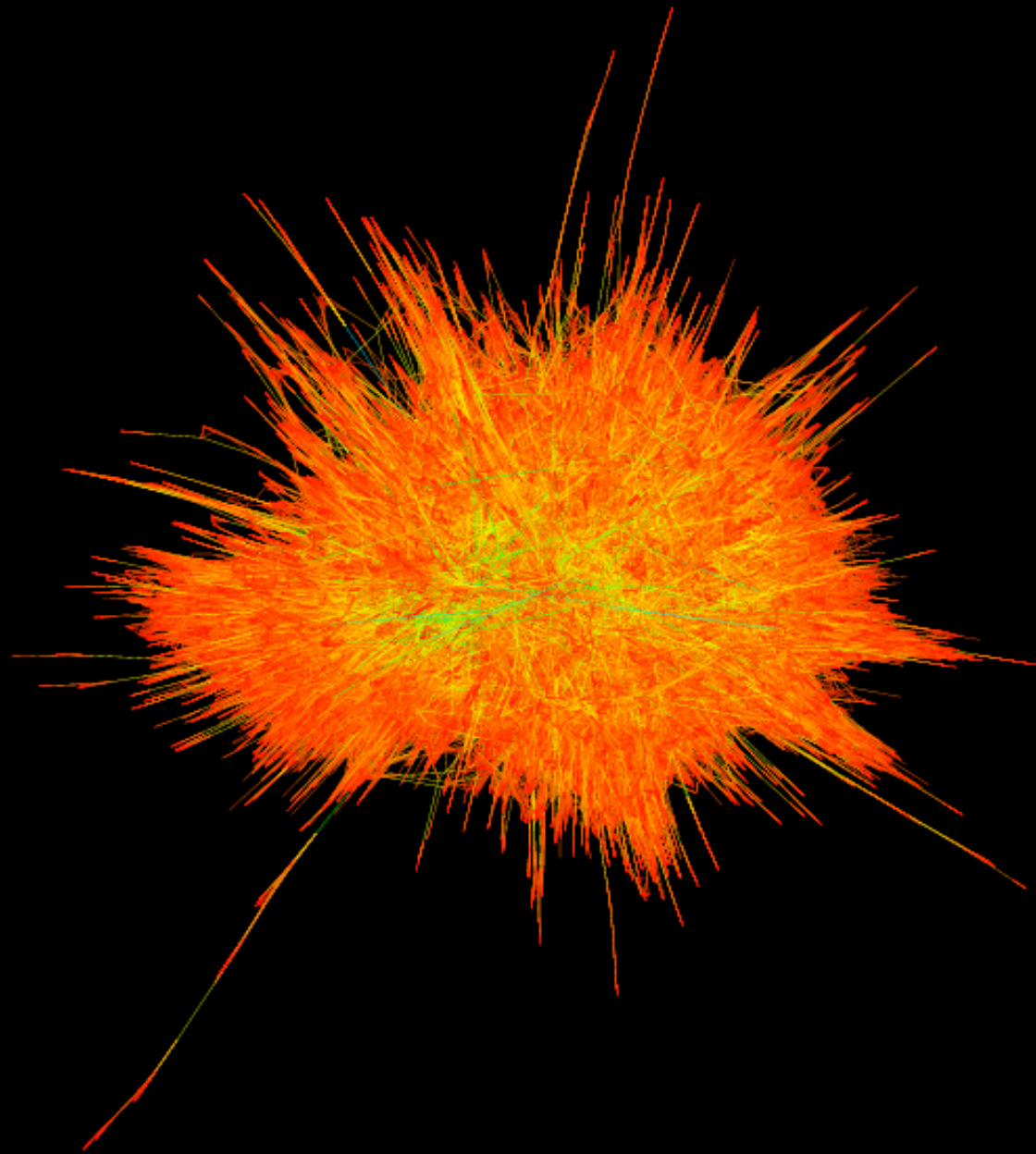
Motif n°10  
135904 occurrences

# Réponses & Conclusion

Peut on connaître les habitudes et préférences des clients?

Peut on dégager des profils acheteurs?

Comment établir des techniques marketing grâce à ces données clients?



SNAP@amazon0302. 262111 nodes, 899792 edges.

# Source

Leskovec, Adamic L.A, and Huberman, B. A.. *The dynamics of viral marketing*. ACM Trans. 2007

SNAPStanford Network Analysis Platform <http://snap.stanford.edu>

<http://www.kdnuggets.com/polls/2010/analytics-data-mining-industries-applications.html>

<http://www.nytimes.com/2010/11/29/business/economy/29shop.html>

<http://www.cise.ufl.edu/research/sparse/matrices/SNAP/amazon0302.html>

Abhishek Srivastava, C.S.D, Columbia University *Motif Analysis in the Amazon Product Co-Purchasing Network*. COMS6998 - Network Theory 2010