

# Les parcours clients

Application à la distribution de biens d'équipement

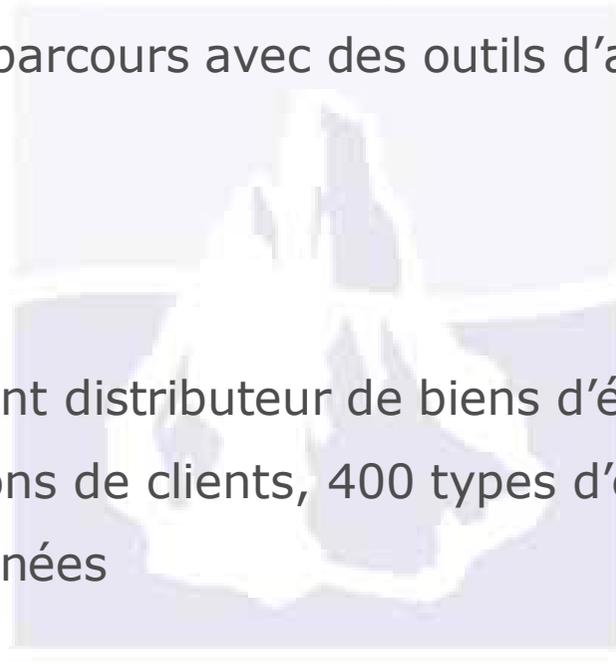
# Contexte et objectifs

## > Objectifs :

- Extraire d'une base de données clients les parcours les plus représentatifs des évolutions comportementales observées
- Représenter ces parcours avec des outils d'aide à l'interprétation simples et intuitifs

## > Matériel

- Source: Base client distributeur de biens d'équipements
- Données: 3 millions de clients, 400 types d'évènements
- Période: 2 à 5 années



# Problématique

## ➤ Premier niveau de connaissance:

### Le comportement instantané d'un client :

- Définition de l'état comportemental d'un client : une quarantaine de comportements différents identifiés (regroupés en 15 familles)

## ➤ Deuxième niveau de connaissance:

### L'évolution du comportement du client dans le temps:

- Mise en évidence pour chaque client d'une séquence d'états, décrivant les comportements successifs du client sur une période de 2 à 5 ans

## ➤ Difficulté à ce stade:

- Un nombre très élevé de séquences possibles et des séquences quasiment toutes différentes pour chaque client
- Comment identifier les principaux chemins suivis par les clients?





# Extraction & classification des chemins

# La première difficulté: l'absence d'outils

- **Des outils spécialisés pour les navigations d'internautes ne donnant pas complètement satisfaction**
  - Log analyser (Numsight)
  - ...
- **Pas de logiciel conçu pour traiter les spécificités de ce type de données temporelles**



# La solution adoptée

- **Développement d'un programme de classification et d'aides à l'interprétation dérivés de la classification ascendante et s'appuyant sur un calcul de distance de type CEM (développement interne SAS)**

- **Travaux de recherche**

- DEA: Algorithme CEM pour le calcul de distance de séquences temporelles (A. Zéboulon)
- Projets de recherche LIPN / NumSight : Extraction de trajectoires d'internautes
  - Khalid BENABDESLEM, Thèse : " Approches Connexionnistes pour la classification et la visualisation des séquences évolutives : Application aux données issues d'usages de l'Internet "  
Soutenue le : 17 Décembre 2003.
  - Farida ZEHRAOUI, Thèse : "Systèmes d'apprentissage connexionnistes et raisonnement à partir de cas pour la classification et le classement de séquences"  
Soutenue le : 12 octobre 2004 (Très Honorable+Félicitations)

# Publications

## > 2003

- ZEBOULON A., BENNANI Y., BENABDESLEM K. (2003) : "Classification et reconnaissance de séquences de navigations par SOM et modèle de mélanges ", RT N°3, Contrat de recherche Maxeem-NumSight, juin.
- BENNANI Y., BENABDESLEM K. (2003) : "Visualisation des traces de navigation par apprentissage non-supervisé ", RT N°2, Contrat de recherche Maxeem-NumSight, mars.
- BENNANI Y., BENABDESLEM K. (2003) : "Codage et compression de données issues d'Internet ", RT N°1, Contrat de recherche Maxeem-NumSight, janvier.

## > 2002

- JANVIER E., BENABDESLEM K., BENNANI Y. (2002), "Method for processing Internet Request files" Brevet International, [N° de publication : WO 02/059781, date de publication : 01/08/2002]
- BENABDESLEM K., BENNANI Y., JANVIER E. (2002) : " Visualization and Analysis of Web Navigation Data", International Conference on Artificial Neural Networks ICANN '02, 27-30 Août, pp. 486-491, Madrid, Espagne.

## > 2001

- BENABDESLEM K., BENNANI Y., JANVIER E. (2001), "Connectionist Approach for Website visitors Behaviors Mining", ACS/IEEE International Conference on Computer Systems and Applications, 26-29 June 2001, Beirut, Lebanon.
- BENNANI Y., BENABDESLEM K., JANVIER E. (2001), "Analyse des données comportementales d'internautes", 1ère Journée Thématique "Exploration de données issues d'Internet", JT-EDI'2001, Mars 2001, Villetaneuse.

# Résultat obtenu: un exemple de chemin

Base = Total client actif	Année 1				Année 2				Année 3				Année 4			
	T1	T2	T3	T4												
	10%	2%	3%	2%	2%	1%	1%	0%	1%	2%	3%	2%	2%	1%	1%	3%
		1%	2%	2%	5%	2%	3%	4%	6%	6%	8%	11%	7%	13%	6%	7%
	2%		11%	3%	9%	6%	11%	5%	5%	5%	6%	8%	8%	6%	9%	7%
			1%	1%				1%	1%	1%	1%	2%	1%	2%	1%	1%
	2%	10%	7%	4%	6%	9%	6%	13%	9%	12%	10%	13%	13%	12%	13%	15%
			3%	1%	1%	10%	11%	12%	11%	8%	11%	9%	7%	8%	10%	9%
	13%	5%	7%	5%	13%	8%	11%	8%	11%	7%	13%	12%	5%	5%	5%	10%
Comportements typés utilitaire maison	29%	1%	13%	4%	11%	8%	10%	13%	15%	13%	13%	9%	19%	13%	15%	10%
Comportements typés femme classique budget raisonnable	2%	2%	3%	2%	1%	1%	1%	0%	0%		1%	0%			0%	1%
	13%	13%	22%	18%	14%	11%	13%	9%	10%	14%	8%	9%	10%	10%	12%	10%
		2%	1%	4%	3%	6%	5%	5%	5%	6%	5%	5%	9%	7%	6%	6%
Comportements typés enfants	3%	1%	6%	3%	6%	5%	7%	5%	9%	8%	9%	6%	4%	4%	9%	4%
		8%	4%	24%	23%	14%	14%	9%	6%	7%	7%	5%	7%	7%	5%	9%
Comportements orientés discount			1%	2%	1%	2%	2%	4%	5%	4%	3%	4%	3%	3%	4%	2%
	27%	55%	18%	25%	4%	15%	4%	9%	5%	6%	3%	5%	6%	7%	3%	9%
inactifs	52%	43%	51%	42%	67%	63%	62%	52%	58%	57%	63%	58%	54%	57%	62%	54%



# La représentation des parcours

**Une nouvelle difficulté:**

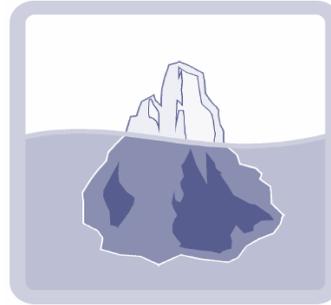
**Représenter et comparer de  
façon simple et intuitive des  
séquences d'évènements**





# Les applications

- **Téléphonie**
- **Grande distribution**
  - **VPC**
  - **Banque**
  - **...**

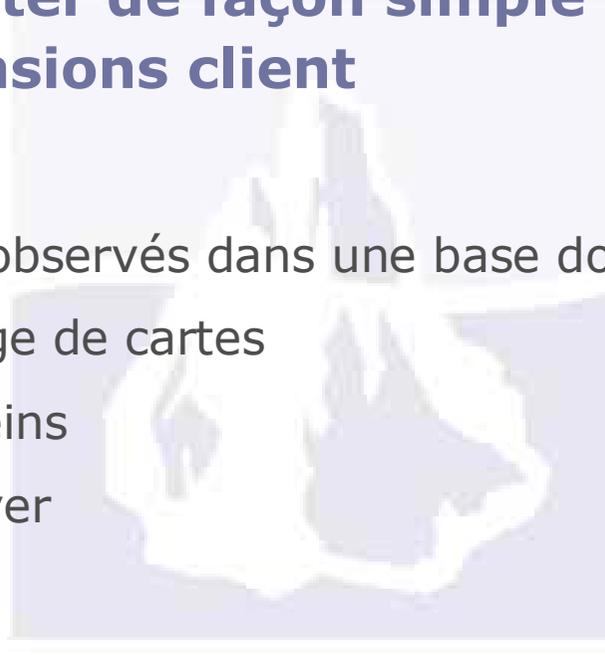


# Les cycles de vie clients

Application aux services bancaires

# Contexte et objectifs

- **Les cartes de crédit et de paiement**
- **Objectif - Représenter de façon simple une vision complète de toutes les dimensions client**
  - Comportements observés dans une base données
  - Détention et usage de cartes
  - Motivations et freins
  - Profil client et foyer

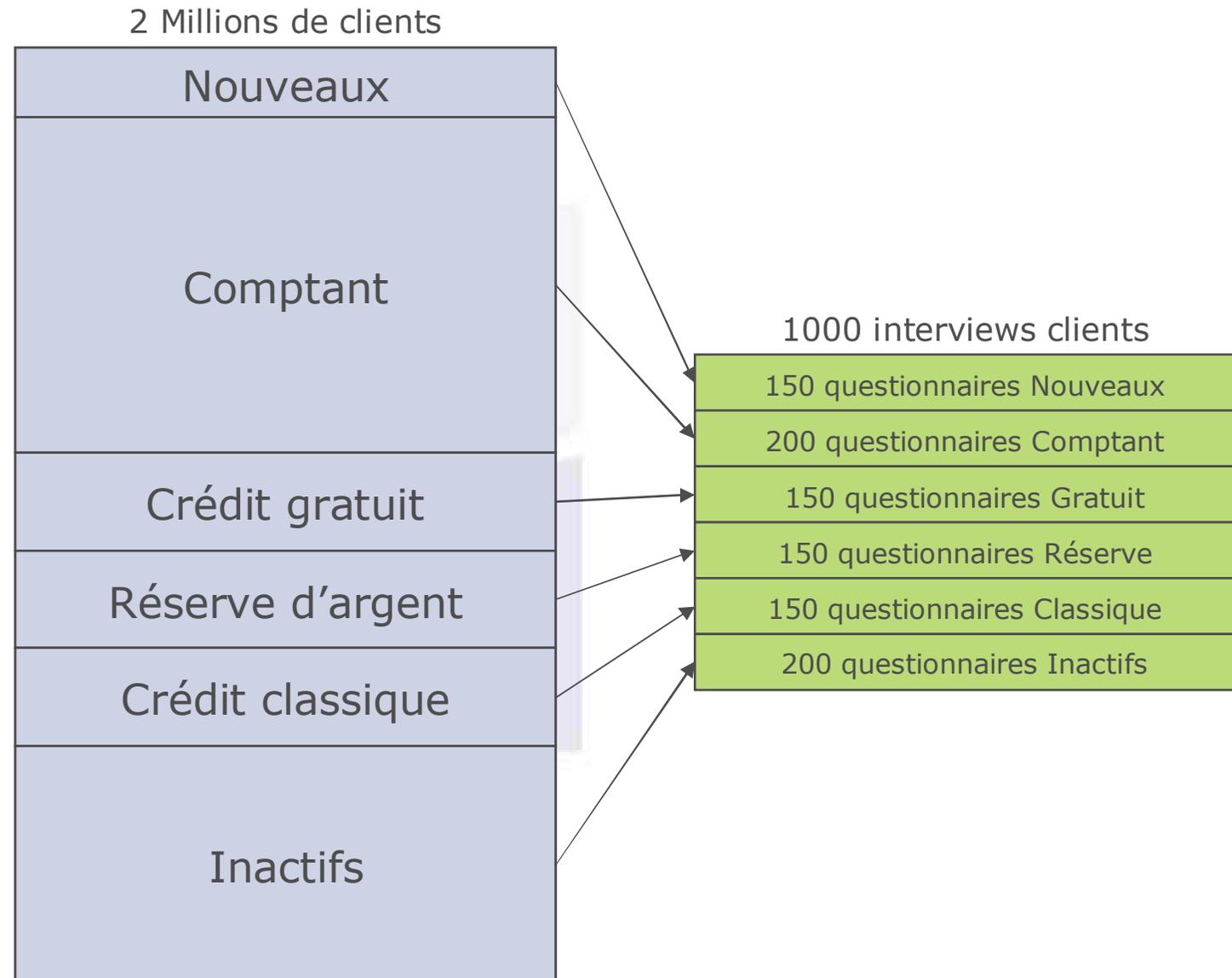


# Matériel

## > 2M Clients observés pendant 4 ans dans une base

- 300 types d'évènements,
- 6 comportements clients (obtenus après segmentation comportementale de la base clients)

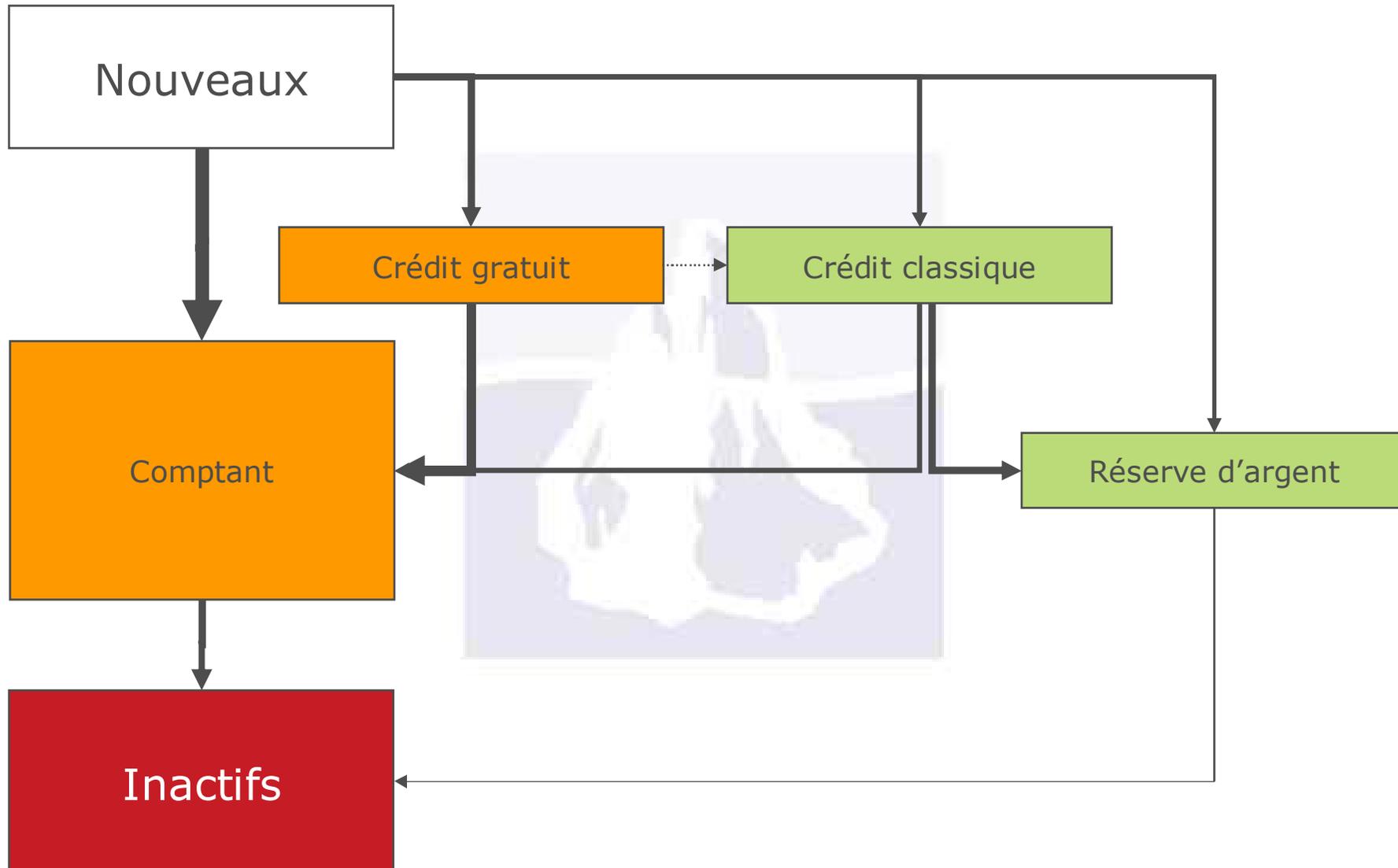
## > 1000 interviews clients stratifiées par segment comportemental





# Les comportements clients

# Parcours clients





# Les profils clients issus de l'interview

# Pourquoi un profil déclaratif client

Comprendre pour chacun  
des 6 comportements clients :

- L'usage de TOUS leurs moyens de paiement
  - Leurs motivations
    - Qui ils sont



# La construction des états et cycles de vie client

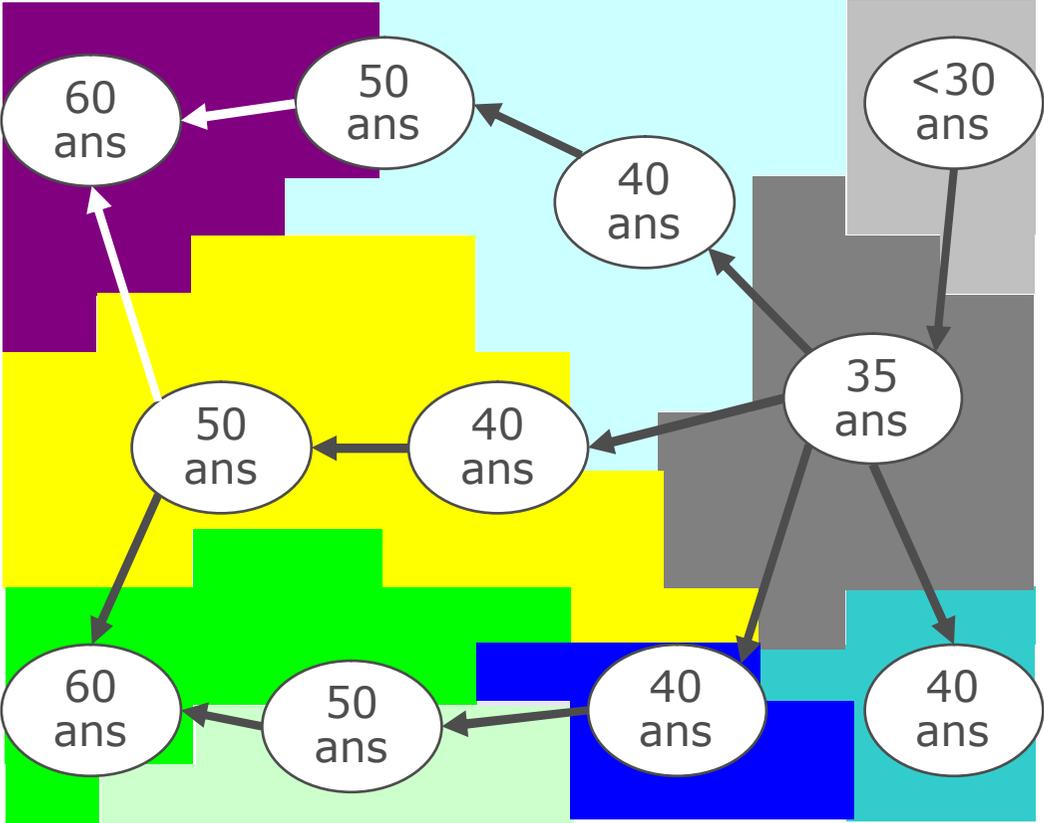
# Difficultés

- **Des données de deux origines, beaucoup de variables (Base client + Interview)**
- **Des variables hétérogènes**
  - Continues et discrètes
  - Des comportement observés
  - Des données déclaratives (Produits détenus et utilisés, Besoins, ...)
- **Beaucoup de redondance d'information**
- **La nécessité d'aboutir à une segmentation**
  - Simple / tenant compte des contraintes opérationnelles de l'entreprise cliente
  - Mais tenant compte de la richesse des données

# La solution adoptée

- **Une approche basée sur les cartes topologiques**
- **Des ajustements méthodologiques rendus accessibles grâce à un partenariat avec le LIPN depuis fin 2004**
  - GUERIF S., BENNANI Y., JANVIER E. (2005) : « $\mu$ -SOM : Weighting features during clustering», WSOM'05, International Workshop On Self-Organizing Maps, September 5-8, Paris.
  - GUERIF S., BENNANI Y. (2005) : « Apprentissage numérique non supervisé à partir de données évolutives », Rapport de recherche de deuxième année, Juin 2005.
  - Financement de thèse S. Guérif [Apprentissage numérique non supervisé, Extraction de Connaissances (ECD), Pondération et Sélection de Variables, Cartes topologiques de Kohonen (Self-Organizing Maps - SOM), Données évolutives (en terme de distributions et de disponibilité d'indicateurs)]

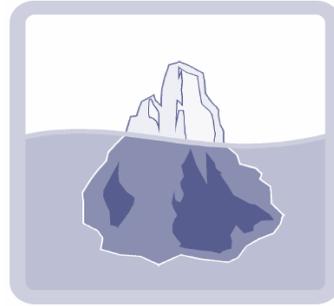
# Représentation topologique de la clientèle de la banque ...





# Les applications

- **Restauration**
- **Produits de grande consommation**
- **Banque, Assurance**
  - **Téléphonie**
- **Distribution, VPC**
  - ...



**Merci pour votre attention**

...