Utilisation des réseaux sociaux dans la lutte contre la fraude à la carte bancaire sur Internet

Françoise Fogelman Soulié, Amine Mekki
Savaneary Sean, Philippe Stepniewski

KXEN, 25 Quai Gallieni, 92 158 Suresnes cedex
http://www.kxen.com
francoise.soulie@outlook.com
amine.mekki@kxen.com
savaneary.sean@kxen.com
philippe.stepniewski@kxen.com

Résumé. Du fait de sa croissance très rapide, le commerce électronique devient une cible majeure pour les fraudeurs. La fraude à la carte bancaire sur Internet est le fait de réseaux internationaux du crime organisé. La lutte contre cette fraude est donc un objectif majeur pour assurer la sécurité des moyens de paiement. Les systèmes classiques de détection de la fraude sont basés depuis les années 80 sur des techniques de data mining. Aujourd'hui, les techniques d'analyse des réseaux sociaux sont très largement décrites dans la littérature. Nous présentons ici une méthodologie permettant d'utiliser ces techniques dans la lutte contre la fraude, à la fois pour la détection et pour l'investigation. Nous présentons les résultats obtenus dans le cadre du projet collaboratif eFraudBox, mené avec le GIE Cartes Bancaires.

1 Introduction

Le commerce électronique est en très forte croissance en France : d'après la Fédération e-commerce et vente à distance (FEVAD), les ventes sur Internet ont ainsi atteint 45 milliards d'euros en 2012 (Fevad, 2013), en hausse de 19% par rapport à 2011. Le principal moyen de paiement est la carte bancaire qui d'après (Fevad, 2012) a été utilisée dans 79% des paiements. Cette croissance du e-commerce est très forte également partout en Europe où le chiffre d'affaires total en 2012 devrait atteindre 254 milliards d'euros.

Alors que le taux de fraude de proximité (sur terminal physique marchand) ou les retraits (sur automate) restent assez stables en France (grâce à la puce présente sur les cartes des porteurs), le taux de fraude des paiements à distance est en forte croissance. La figure1 représente ces évolutions (d'après les statistiques du rapport annuel de l'Observatoire de la sécurité des cartes de paiements (OSCP, 2011b)). On voit ainsi sur la figure que les taux de fraude des paiements de proximité et des retraits restent assez stables (aux alentours de 0,2%), alors que le taux de fraude des paiements à distance est en forte croissance (de 0,24% en 2007 à 0,32% en 2011, soit une augmentation de 36%).

Les montants sont croissants, mais beaucoup plus rapidement pour les paiements à distance : en effet le montant des ventes à distance a représenté 73% du montant total des paiements en 2011 contre 52% en 2007, et la part des paiements sur Internet dans les paiements à distance est passée à 80% en 2011 pour seulement 53% en 2007 (au détriment donc du paiement par courrier ou téléphone). La figure 1 montre ces évolutions : la fraude sur les paiements à distance représente ainsi 130 millions d'euros en 2011 (dont 104 millions sur Internet), alors qu'elle ne représentait «que» 50 millions (dont 26 sur Internet) en 2007.

La fraude à la carte bancaire sur Internet est donc un phénomène massif, en forte croissance, aux mains du crime organisé en large partie, que les banquiers et les marchands doivent absolument prévenir. Les marchands (qui subissent la perte en cas de fraude) ont ainsi mis en place progressivement des solutions de sécurisation des transactions, comme 3D-Secure par exemple (OSCP, 2011a), ou font appel à des sociétés spécialisées (par exemple (FiaNet, 2010)).

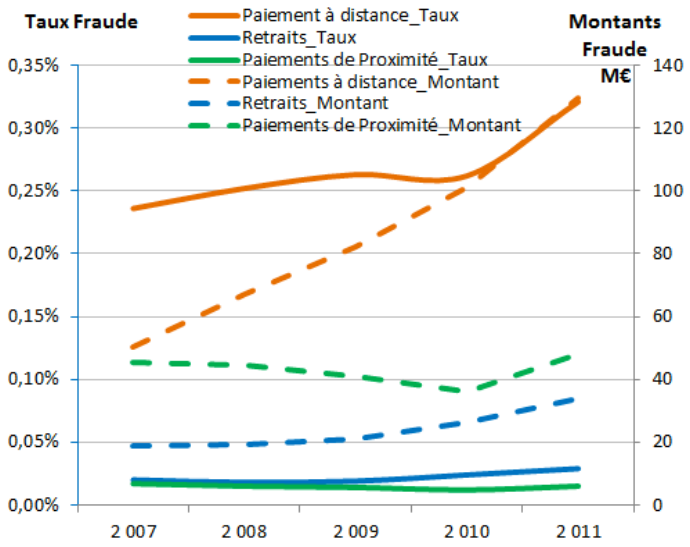

FIG. 1 – *La fraude aux moyens de paiement en France (OSCP, 2011b)*

Le GIE Cartes Bancaires est missionné par la plupart des établissements financiers français pour assurer l'inter-bancarité des cartes de paiement CB (Visa, MasterCard). En effet GIE CB a mis en place un système partagé par toutes les banques pour le paiement et le retrait par carte : celui-ci permet au titulaire d'une carte bancaire de l'utiliser partout. Il met en relation le titulaire de la carte (le porteur), le commerçant, la banque du porteur (l'émetteur) et la banque du commerçant (l'acquéreur).

Le GIE CB a géré en 2011 plus de 9 milliards de transactions réalisées par 60 millions de cartes CB, pour un montant total de 482,6 milliards d'euros (GIE-CB, 2011) en croissance de 7% par rapport à 2010. Dans le cadre de sa mission, le GIE CB participe activement à la lutte contre la fraude, investiguant sur réquisition judiciaire et allant en justice si nécessaire (33 nouveaux dossiers ont ainsi été ouverts au pénal en 2011 d'après GIE-CB, 2011). Le GIE CB

a ainsi souhaité s'associer à un projet financé par l'ANR, eFraudBox, visant à développer des boîtes à outils pour la détection et l'investigation de la fraude. Nous ne décrirons pas l'ensemble de ce projet (voir pour cela (Consortium e-Fraud Box, 2011)). Nous nous concentrons, dans cet article, sur les travaux menés par KXEN, partenaire du projet, autour des techniques d'analyse de réseaux sociaux.

Comme les chiffres donnés plus haut l'ont montré, les difficultés majeures rencontrées dans nos travaux ont tenu à la *volumétrie massive des données* (plusieurs dizaines de millions de transactions sur Internet par mois), avec des *taux de fraude faibles* (0,29% en 2011, d'après (Consortium e-Fraud Box, 2011)).

Le présent article est organisé comme suit : nous commençons par décrire ce que nous entendons par réseau social (section 2) ; puis nous définirons plus précisément la fraude sur internet (section 3) ; ensuite nous présenterons les données (section 4), les techniques pour la détection de la fraude (section 5), et pour l'investigation de la fraude (section 6) ; enfin, nous montrerons comment l'outil KXEN [1] nous a permis de mettre en œuvre ces développements (section 7), avant de conclure (section 8).

Le travail présenté dans cet article est entièrement basé sur les données et les besoins exprimés dans le projet eFraudBox. Les résultats obtenus par KXEN après la première année de travail ont été décrits dans (Fogelman Soulié et al., 2011). Nous présentons ici l'ensemble des résultats obtenus par KXEN à la clôture du projet, après 3 ans de travail.

Nous pensons que beaucoup des techniques développées ici peuvent être utilisées de façon opérationnelle dans la lutte contre la fraude, mais aussi dans un cadre beaucoup plus large d'applications de sécurité.

2 Qu'est-ce qu'un réseau social ?

Les techniques d'analyse des réseaux sociaux (Social Network Analysis ou SNA) sont issues de la théorie des graphes ((Erdös et Rényi, 1959)) et de la sociologie ((Wasserman et Faust, 1994)). Plus récemment, de très nombreux travaux sont apparus dans la communauté du data mining (Barabasi, 2002; Newman, 2003; Watts, 2003; Aggarwal et Wang, 2010).

Dans la suite de cet article, un *réseau social* est défini comme un graphe, c'est-à-dire un couple $G = (N, A)$, où N est un ensemble d'entités (les nœuds) et A un ensemble d'arêtes liant ces entités (voir figure 2 à gauche). Les entités peuvent être de toutes sortes : clients, produits, cartes de crédit, marchands, ... et les liens représentent les relations entre ces entités : amitié sur un site social, appels téléphoniques pour un opérateur téléphonique, liens entre url... On pourra se référer pour plus de détails à (Chapus et al., 2011).

Contrairement à un simple graphe, un réseau social porte en général des attributs sur les nœuds (nom, adresse, possession d'un produit...), voire sur les liens (nombre d'appels téléphoniques entre les deux nœuds par exemple). Les propriétés des réseaux sociaux sont très différentes de celles d'un simple graphe : petit monde, homophilie, contagion et structuration en communautés sont des propriétés classiquement observées sur les réseaux sociaux et qu'on ne rencontre pas dans un graphe, aléatoire par exemple. Nous ne détaillerons pas ces différents points, voir (Watts, 2003).

1. KXEN est un éditeur data mining dont le logiciel KXEN InfiniteInsight^{TM}est considéré par Gartner comme l'un des produits leaders du marché : "KXEN is an emerging predictive analytics vendor which emphasizes rapid creation of insight across large datasets" (Herschel, 2007). Voir `http://www.kxen.com`.

FIG. 2 – *Un réseau social (à gauche) et le cercle du nœud A (à droite).*

Le *voisinage* – ou cercle – d'un nœud est l'ensemble des nœuds auxquels il est relié. Son *degré* est le nombre de ces nœuds voisins – ou amis – (figure 2 à droite).

La *communauté* d'un nœud est un ensemble de nœuds tel qu'il y ait plus de liens entre les nœuds de la communauté qu'avec des nœuds en dehors de la communauté. Il existe de très nombreuses méthodes de décomposition d'un réseau en communautés ((Fortunato, 2009)) : c'est un problème difficile, et donc posant des problèmes de calcul pour les réseaux de grande taille. Il existe cependant des techniques rapides pour obtenir ces décompositions dans le cas de l'optimisation de la modularité ((Blondel et al., 2008)). Il est alors possible de décomposer le réseau, donc de segmenter les nœuds, en des temps raisonnables, même si la taille du réseau dépasse quelques millions de nœuds. C'est cette dernière méthode que nous utiliserons dans toute la suite de cet article.

Pour l'analyse de la fraude, nous utiliserons un type de réseau particulier : le *réseau social bipartite*. Il s'agit d'un réseau dans lequel les nœuds sont de deux types différents (ici, les cartes et les marchands) et où il ne peut y avoir de lien qu'entre nœuds de types différents : il y a un lien entre un nœud carte et un nœud marchand si la carte a effectué une transaction chez le marchand sur la période considérée (figure 3 en haut). On peut ensuite projeter le réseau bipartite en deux réseaux «simples», où les nœuds sont de l'un des deux types seulement et où un lien relie deux nœuds d'un type si et seulement si ils sont tous les deux reliés à au moins k nœuds de l'autre type (k est un paramètre de seuillage qu'on pourra choisir : le poids du lien représentera par exemple ce nombre de voisins communs ; il doit donc être supérieur ou égal à k). On a donc deux paramètres k_c et k_m pour les deux réseaux (cartes et marchands).

La projection est une étape qui peut être assez longue, notamment si certains nœuds sont très connectés : par exemple, les marchands à très forte activité, comme la SNCF, Amazon..., sont visités par de très nombreuses cartes, et auront donc beaucoup de cartes communes avec les autres marchands ; ils seront donc connectés à beaucoup de marchands. Les nœuds très connectés sont appelés «*méga-hubs*», il est souvent préférable de les éliminer du réseau avant d'effectuer les projections. On peut pour cela, utiliser différentes méthodes :

- *seuil absolu* : on exclue tous les nœuds ayant un degré supérieur à un certain seuil, à déterminer au mieux : il doit permettre de garder le plus possible de nœuds tout en éliminant les nœuds indésirables ;
- *seuil relatif* : on exclue tous les nœuds ayant un degré largement supérieur à la moyenne des degrés. On dira par exemple que le nœud A est un méga-hub si son degré satisfait :

$$\deg(A) \geq \overline{\deg} + k\,\sigma_{\deg} \tag{1}$$

où $\overline{\deg}$ (resp. $\sigma_{\deg}$) est la moyenne (resp. l'écart type) des degrés des nœuds du réseau. C'est cette procédure que nous utiliserons dans la suite.

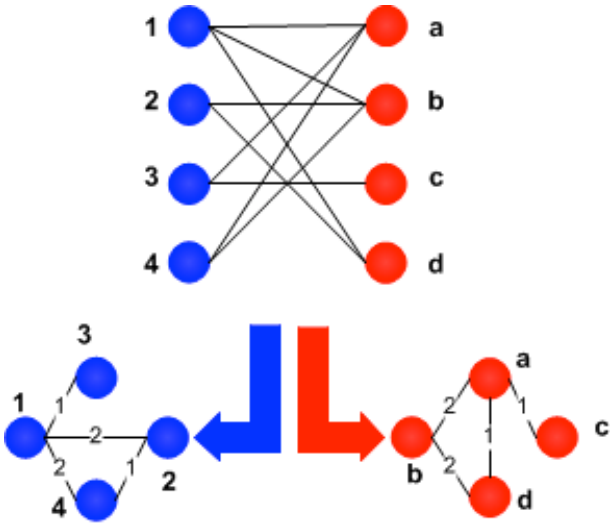

FIG. 3 – *Un réseau bipartite (en haut) et les réseaux projetés (en bas).*

3 La fraude à la carte bancaire sur Internet

Les transactions sur Internet sont traitées à travers le réseau GIE CB qui gère le processus d'autorisation pour les réseaux Visa et Mastercard. Deux processus d'analyse de la fraude sont mis en œuvre :

- La **détection de la fraude** : le but est de mettre une carte en alerte si on pense qu'elle a fait au moins une transaction frauduleuse. Le porteur de la carte est alors contacté par la banque : si le porteur confirme la fraude, la carte est mise en opposition (ce qui bloque toute fraude ultérieure) et la liste des éventuelles transactions frauduleuses antérieures est établie avec le porteur. On ne peut en général pas détecter la première transaction frauduleuse, mais plus on détecte tôt, plus le coût de la fraude sera réduit : le *montant de fraude évitée* est donc un indicateur important.
- L'**investigation de la fraude** : un lot de cartes est constitué à partir de remontées d'alertes manuelles ou automatiques (essentiellement des cartes ayant fraudé) et une équipe d'investigation étudie le lot pour essayer de comprendre le mécanisme de la fraude. On cherche en particulier à déterminer le *point de compromission*, c'est-à-dire le site où les numéros de cartes ont été dérobés.

Les techniques de réseaux sociaux peuvent être utilisées pour les deux problèmes, mais de façons assez différentes, ce que nous détaillons dans les sections suivantes.

4 Les données

Les données disponibles pour les transactions ne sont pas très riches : on ne dispose pas, en particulier, d'informations concernant le porteur de la carte (nom, adresse, âge, sexe, ...) ou

le produit acheté (type de produit, nombre de produits, ...), informations que seuls la banque (données cartes) ou le marchand (données produit) possèdent. Par contre, on dispose de l'intégralité des transactions réalisées par les porteurs sur Internet, ce qui ne serait pas le cas pour une banque (qui ne «verrait» que les transactions faites par les porteurs ayant une carte de la banque) ou pour un marchand (qui ne «verrait» que les transactions faites par les porteurs achetant chez lui).

Les données disponibles pour chaque transaction sont les suivantes :

- **Informations sur la carte** : numéro de carte, date d'expiration, banque émetteur, ...
- **Informations sur le marchand** : identifiant, SIRET, pays, activité du marchand, banque du marchand (acquéreur) et pays de la banque, terminal utilisé... ;
- **Informations sur la transaction** : date de la transaction (locale et GMT), montant (en monnaie locale et en euros).

Une fois qu'une carte est mise en opposition, on dispose en plus de la date de mise en opposition, du motif et, pour chaque transaction déclarée frauduleuse, d'une étiquette l'indiquant (notons que cette étiquette est renseignée de façon rétrospective, quelques fois plusieurs mois après que la fraude ait eu lieu).

À partir de ces transactions, nous calculerons de très nombreux agrégats, caractérisant l'historique d'activité de chaque carte/marchand :

- **Agrégats cartes** : à la date T, sur une fenêtre glissante (de longueurs variées : jour, semaine, mois) se terminant à T, on calcule des indicateurs pour chaque carte comme le nombre et le nombre moyen de transactions, le montant total et le montant moyen des transactions, l'écart entre le nombre, le montant de la date T et le nombre, le montant moyen dans la fenêtre ;
- **Agrégats marchands** : à la date T, sur une fenêtre glissante (de longueurs variées : jour, semaine, mois) se terminant à T, on calcule des indicateurs pour chaque marchand comme le nombre et le nombre moyen de transactions, le montant total et le montant moyen des transactions, le nombre de transactions frauduleuses et le montant total de la fraude, l'écart entre le nombre, le montant de la date T et le nombre, le montant moyen dans la fenêtre.

On obtient ainsi 703 agrégats comme indiqué dans la table ci-dessous :

Variables	**Nombre**
Variables GIE	37
Agrégats Carte	300
Agrégats Marchand	366
Total	**703**

TAB. 1 – *Variables disponibles.*

Chaque jour, pour chacune du million de transactions journalières environ, on calcule l'ensemble de ces variables : on obtient donc des volumes de données très importants.

5 La détection de la fraude

Le processus de détection de la fraude vise à lever une alerte sur les cartes :

– **en temps réel** : chaque transaction passe dans le moteur de détection qui lui attribue un score, d'autant plus élevé que la transaction a plus de risques d'être frauduleuse. Ce score peut alors être utilisé pour bloquer la transaction, par exemple en comparant à un seuil s, à déterminer selon les objectifs (par exemple, bloquer au plus 0,5% des transactions) :

$$\text{Si } score(t) \geq s, \text{ alors la transaction } t \text{ est frauduleuse} \tag{2}$$

– **en temps différé** : l'ensemble des transactions de la veille (par exemple) sont passées dans le moteur de détection et reçoivent un score de risque comme précédemment.
À partir du score des transactions, on calcule un score carte, par exemple comme la somme des scores des transactions de la carte. Les cartes ayant le plus fort score sont alors *mises en alerte*, et signalées aux banques des porteurs pour qu'elles vérifient auprès des porteurs s'il y a eu fraude.

Dans toute la suite, nous traiterons uniquement le cas de la détection en temps différé. Le problème de la détection en temps réel est très similaire, mais impose des contraintes additionnelles de temps de calcul.

5.1 Techniques classiques de détection de fraude

Les techniques de détection de fraude sont essentiellement issues du data mining : réseaux de neurones dans le système Falcon développé à partir de 1995 ((Hassibi, 2000)), puis techniques statistiques variées ((Hand et Weston, 2008) ou (Bolton et Hand, 2002)). Récemment, l'analyse des réseaux sociaux (ou SNA : Social Network Analysis), apparue initialement en sociologie ((Wasserman et Faust, 1994)), ou sur le Web ((Kleinberg, 1997)) a commencé à être utilisée dans le domaine de la sécurité : lutte contre le terrorisme ((Memon et Hicks, 2008), (Ressler, 2006)), fraude dans les enchères ((Pandit et al., 2007)) par exemple.

Nous utiliserons dans la suite comme mesures de performance du système de détection la couverture (ou rappel) et la pertinence (ou précision). La *couverture* mesure le taux de cas de fraude identifiés et la *pertinence* le taux d'alertes réellement frauduleuses. On définit d'abord la *matrice de confusion* (tableau 2) : elle représente, pour les cas positifs (fraude) et négatifs (non fraude) réels et prévus, les nombres VP de vrais positifs, FP (faux positifs), FN (faux négatifs), VN (vrais négatifs), A (mis en alerte), Non A (non mis en alerte), F (fraude) et Non F (non fraude).

		Prévu P	*Prévu* N	
Réel	P	VP	FN	F
	N	FP	VN	Non F
		A	non A	

TAB. 2 – *Matrice de confusion.*

La couverture et la pertinence (appelés classiquement rappel et précision dans la littérature sur la recherche d'informations) sont alors définies de la façon suivante :

$$\text{Couv}_s = \frac{\text{VP}}{\text{F}} = \frac{\text{VP}}{(\text{VP} + \text{FN})} \qquad \text{Pert}_s = \frac{\text{VP}}{\text{A}} = \frac{\text{VP}}{(\text{VP} + \text{FP})} \tag{3}$$

Remarquons qu'on s'attend en général à un taux d'alerte du même ordre que le taux de fraude. Pour maximiser la couverture, il faut minimiser les faux négatifs (ces fraudeurs ne seront pas investigués) et pour maximiser la pertinence, il faut peu de faux positifs (ces dossiers seraient investigués pour rien).

On cherchera donc à obtenir un système de détection ayant bonne couverture (pour ne pas laisser passer trop de cas de fraude) et bonne pertinence (pour ne lever des alertes qu'à bon escient).

5.2 Notre approche pour la détection de fraude

Nous avons décrit dans (Fogelman Soulié et al., 2011) notre approche pour construire les modèles de détection de la fraude : nous utilisons le logiciel InfiniteInsight^{TM} de KXEN : les modèles sont construits avec le composant InfiniteInsight^{TM} Modeler qui exploite la théorie de minimisation structurelle du risque de (Vapnik, 1995, 2006).

L'implémentation de cette théorie par KXEN réalise une optimisation de l'AUC (Fogelman Soulié et Marcadé, 2008) par une méthode de régularisation *ridge*, ce qui est particulièrement utile dans le cas de la fraude, puisqu'on peut traiter des cas de classes très déséquilibrées.

Un modèle simple (*baseline*), utilisant uniquement les 37 variables initiales décrivant les transactions (tableau 1), a été construit avec InfiniteInsight^{TM} Modeler. Il a des performances très faibles : 8.2% en pertinence et 1.4% en couverture. Or les performances visées en opérationnel sont de l'ordre de 70% en pertinence et 30% en couverture. On doit donc mettre en œuvre des techniques complémentaires pour tenter d'atteindre ces performances.

Nous avons donc constitué une chaîne de traitement comprenant un certain nombre de composants qui peuvent être assemblés en tout ou partie :

- **Calcul d'agrégats** : nous calculons 300 agrégats cartes et 366 agrégats marchands, comme décrit dans la section 4 ;
- **Segmentation** : les cartes ont des comportements de fraude très différents, ce qui rend difficile la tâche d'un classifieur global. Nous avons donc réalisé une segmentation supervisée des cartes, avec le composant InfiniteInsight^{TM} Modeler en mode segmentation (qui utilise un algorithme k-means). La supervision est faite par l'information de fraude, de sorte que les cartes ayant un comportement de fraude similaire sont regroupées. On obtient ainsi 19 segments de tailles et de taux de fraude très différents comme le montre la figure suivante (figure 4).

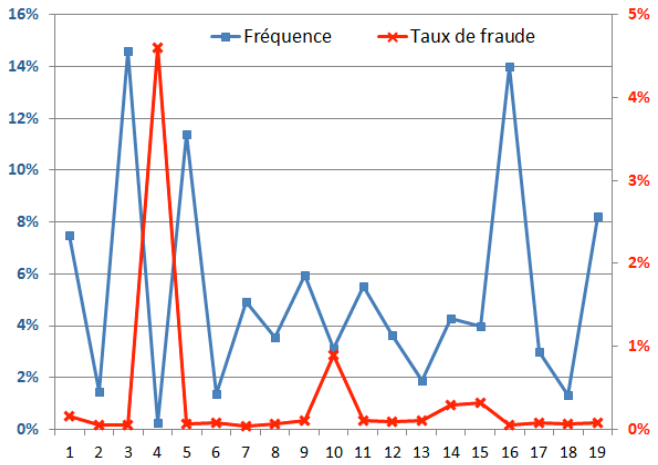

FIG. 4 – *Taille et taux de fraude des 19 segments Carte obtenus.*

– ***Segmented modeling*** : on construit ensuite un modèle par segment (avec InfiniteInsight™ Modeler en mode classification), en utilisant soit les variables initiales, soit les agrégats cartes et marchands ; soit au plus 703 variables (tableau 1).

Les performances obtenues sont décrites dans le tableau 5 : comme on le voit, le modèle simple baseline utilisant uniquement les 37 variables initiales a des performances très faibles ; les modèles basés sur la segmentation Cartes définie précédemment triplent les performances et si ces modèles utilisent de plus les agrégats cartes (300) et marchands (366), les performances sont encore augmentées, mais restent toutefois encore loin des performances visées.

5.3 Les réseaux sociaux pour la détection de fraude

Dans cette section, nous allons décrire comment utiliser une technique de réseau social bipartite comme présentée à la section 2 pour d'une part décrire le comportement d'achat des cartes (§5.3.2) et d'autre part améliorer les performances de détection (§5.3.3).

5.3.1 Construction des réseaux sociaux

À partir de l'ensemble des transactions d'une période P (une journée, un mois...), nous construisons un réseau bipartite (figure 3) avec des nœuds «cartes» et des nœuds «marchands» reliés s'il y a eu une transaction de la carte chez le marchand dans la période considérée. Nous projetons ensuite ce réseau bipartite en un réseau Cartes et un réseau Marchands, dans lequel deux cartes (resp. marchands) sont reliées si elles ont vu au moins les mêmes k_c marchands (resp. ils ont vu les mêmes k_m cartes) : k_c et k_m sont les supports des projections. Ce travail est fait par le module Social du logiciel InfiniteInsight™ de KXEN. On peut, si nécessaire, supprimer les méga-hubs comme indiqué dans l'équation 1.

Nous pouvons ensuite calculer les communautés dans ces deux réseaux, calculer différents attributs sociaux sur les nœuds comme le degré, la taille de la communauté du nœud, la moyenne des degrés de ses voisins... (voir plus de détails dans (Chapus et al., 2011)).

Il se trouve que la fraude est faite à partir d'un ensemble de numéros de cartes volés que le fraudeur va faire passer chez un petit nombre de marchands (ceux dont il sait revendre les produits) : ces cartes vont donc être reliées entre elles dans le réseau Cartes et les marchands concernés reliés dans le réseau Marchands. L'analyse de ces réseaux doit donc nous apporter des informations utiles sur les méthodes de fraude.

Nous présentons deux exemples d'application de ces techniques : la première, basée sur le co-clustering (§5.3.2), nous fournit des rapports sur les principaux comportements observés ; la seconde (§5.3.3) nous permet d'utiliser des variables sociales pour améliorer les performances.

5.3.2 Co-clustering

Nous allons présenter cette technique en utilisant les données d'une seule journée (le 1er mai 2011). Nous construisons donc comme décrit ci-dessus les réseaux Cartes et Marchands (en projetant avec un faible support $k_c = k_m = 2$, pour conserver le plus de nœuds possible).

L'extraction des communautés avec l'algorithme (Blondel et al., 2008) implémenté dans KXEN fournit 6 700 communautés Cartes et 134 communautés Marchands. Les distributions des tailles de ces communautés sont comme bien souvent en loi de puissance.

Nous calculons ensuite, pour chaque communauté Cartes c et chaque communauté Marchands m le *taux d'appartenance* de la communauté Cartes c à la communauté Marchands m, défini comme le rapport du nombre de cartes de la communauté c ayant effectué une transaction chez un marchand de la communauté m au nombre total de cartes de la communauté c.

$$\text{Taux\_App}(c, m) = \frac{\text{Nb de cartes de } c \text{ ayant acheté chez un marchand de } m}{\text{Nb total de cartes de } c} \tag{4}$$

La matrice de co-clustering (figure 5) dont les éléments sont ces taux d'appartenance est une matrice très creuse : une communauté Cartes fait en général ses transactions chez un petit nombre de communautés marchands, voire une seule. De même, une communauté Marchands ne voit en général que quelques communautés de cartes. Dans la figure (les communautés marchands, en ligne, sont ordonnées par taille croissante de haut en bas et les communautés cartes, en colonne, de même, de gauche à droite), on voit ainsi une très grosse communauté de marchands (ligne au bas de la figure) et une très petite (ligne au haut de la figure) visitées par de très nombreuses communautés de cartes (cellules marquées d'un tiret, taux d'appartenance nul, à rouge, taux très fort). Mais la plupart des lignes (et des colonnes) sont vides.

On peut alors utiliser la segmentation fournie par le co-clustering pour analyser les comportements de fraude. Le tableau 3 ci-dessous montre ainsi par exemple une communauté de 56 cartes dont 53 ont acheté dans une communauté de 16 marchands dédiée au voyage. Huit de ces cartes ont fraudé pour un montant total de 1 691,80€ dans la journée. On peut alors investiguer plus en détail le groupe de marchands concernés par ces fraudes.

Ce reporting peut être généré, à partir des communautés calculées par le logiciel KXEN, en utilisant tout logiciel de reporting du marché.

5.3.3 Utilisation dans les modèles de détection

Nous utilisons la procédure de construction du réseau bipartite et de projection en réseaux Cartes et marchands décrite dans la section 5.3.1, sur les données de transactions d'un mois.

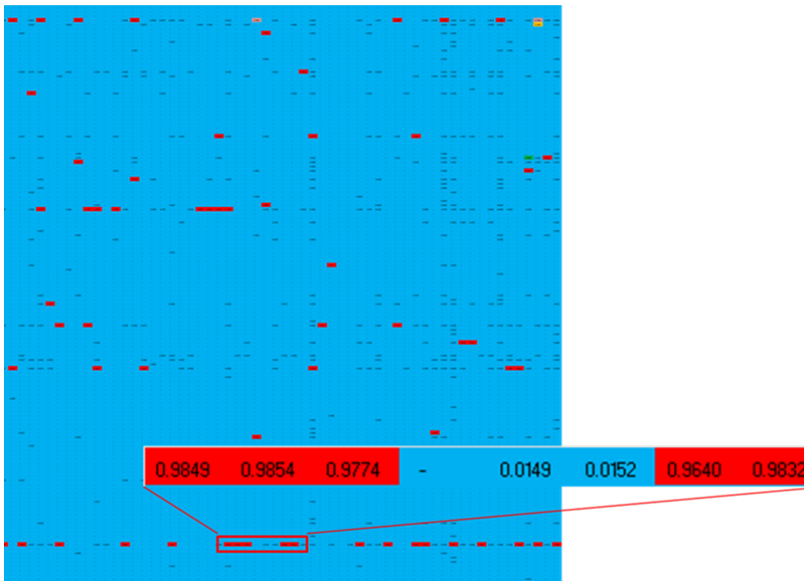

FIG. 5 – *Matrice de co-clustering des cartes et marchands du 1er mai 2011.*

Comme précédemment, nous calculons les communautés et définissons ensuite, pour chaque nœud (carte ou marchand), un certain nombre de variables exploitant la structure des réseaux, comme par exemple : le degré du nœud, l'index et la taille de sa communauté ; pour les marchands, le montant moyen et le nombre de cas de fraude réussis dans sa communauté/ son premier cercle, le nombre moyen de cartes distinctes fraudées dans sa communauté/son premier cercle, le nombre moyen de transactions acceptées/refusées dans sa communauté/son premier cercle... On obtient ainsi 195 variables sociales cartes et 99 variables sociales marchands, qui viennent s'ajouter aux variables déjà définies (table 4).

On construit maintenant, comme défini dans la section 5.2, un modèle segmenté, en construisant un modèle sur chacun des 19 segments et exploitant l'ensemble des 997 variables (agrégats et agrégats sociaux). Comme le montre le tableau ci-dessous (table 5), les performances sont nettement améliorées, pour le modèle *baseline* comme pour le modèle segmenté : on obtient, pour le modèle segmenté une augmentation de 23% en couverture et 11% en pertinence (soit un facteur de plus de 6 et 5 par rapport au modèle *baseline*). Nous atteignons maintenant à peu près les objectifs de performance visés, avec 997 variables.

L'analyse des 50 variables les plus significatives pour quelques modèles de segments (1, 3, 4, 10 et 16 de gauche à droite sur la figure 6), montre que les variables les plus contributives diffèrent selon les segments. Ce sont les agrégats marchands (en bleu foncé) ou les agrégats cartes (bleu clair), les variables sociales (en vert clair pour le réseau Cartes et vert foncé pour le réseau marchands), les variables initiales (en rouge) apparaissant plus ou moins selon les segments. Les variables sociales apparaissent massivement, expliquant l'augmentation des performances.

Nous avons donc montré que l'utilisation des réseaux sociaux et des variables sociales extraites permet d'améliorer de façon significative les performances des modèles de détection.

Communauté Cartes			**Communauté Marchands**			**Taux appartenance**
Numéro	Taille	Tx Cartes Fraudées	Numéro	Taille	Tx Cartes Fraudées	94,6%
5193	56	14,3%	38	16	0,32%	
			Marchands de la communauté			
Fraude			Transport			
Nb Cartes Fraudées : 8			Restauration-Hôtel			
Montant Fraude : 1 691,80€			Location de voiture, ...			

TAB. 3 – *Exemple d'un co-cluster.*

Variables	**Nombre**
Initiales (transactions)	37
Agrégats Carte	300
Agrégats Marchand	366
Variables Sociales Carte	195
Variables Sociales Marchand	99
Total	**997**

TAB. 4 – *Ensemble des variables disponibles.*

Nous allons maintenant présenter l'utilisation des réseaux sociaux pour l'investigation de la fraude.

6 L'investigation de la fraude

Le but de l'investigation est très différent de celui de la détection : il s'agit ici, pour un enquêteur qui reçoit un *lot* de cartes (dont beaucoup ont été fraudées), d'identifier le mode de fonctionnement des fraudeurs, et notamment de déterminer le *point de compromission* (c'est-à-dire le site où les numéros de cartes ont été dérobés), voire les *points de test* (les fraudeurs ayant volé un numéro de carte vont la tester, en général avec un petit montant, sur un site où la sécurité est faible). Ensuite, le fraudeur va faire des transactions frauduleuses, souvent sur plusieurs sites.

Le problème de l'investigation nécessite donc d'identifier ces sites d'une part, et d'autre part, de comprendre la succession des actions au cours du temps. Une fois que l'enquêteur a fini son analyse, il en transmet les résultats aux équipes de police ou gendarmerie qui auront la responsabilité d'arrêter le fraudeur (si celui-ci est en France).

Modèle	Couverture	Pertinence
Baseline	1,40%	8,18%
Baseline + Agg	9,13%	19,00%
Baseline + Agg + Agg Sociaux	9,09%	40,58%
Seg 19	5,09%	28,21%
Seg 19 + Ag.	7,38%	28,82%
Seg 19 + Agg + Agg Sociaux	16,46%	60,89%

TAB. 5 – *Ensemble des variables disponibles.*

L'investigation est donc aujourd'hui une tâche très manuelle, qui peut prendre beaucoup de temps. Automatiser en partie cette tâche permettrait aux enquêteurs d'être plus réactifs et de se concentrer sur les cas vraiment complexes.

Nous allons présenter deux approches utilisant les réseaux sociaux que nous avons développées pour l'investigation : la visualisation des sites marchands impliqués dans la fraude et la détermination de la dynamique de la fraude.

Dans tout ce qui suit, on dispose de données caractérisant un lot de cartes : c'est-à-dire un ensemble de numéros de cartes bancaires et une période de temps à laquelle on s'intéresse.

6.1 Visualisation

On va ici s'intéresser à l'ensemble des cartes ayant effectué au moins une transaction frauduleuse sur la période considérée (par exemple un mois). On recherche ensuite l'ensemble des transactions (frauduleuses et non frauduleuses) effectuées par ces cartes sur une période englobant la période considérée et on constitue alors le lot des marchands où ces cartes ont effectué des transactions. Le lot de cartes est donc constitué ici de toutes les cartes ayant fraudé pendant la période. Ce lot est de petite taille (quelques centaines de cartes au plus), et donc les données obtenues sont de petite taille.

On construit ensuite, comme décrit dans la section 5.3.1, le réseau bipartite des transactions entre les cartes du lot et les marchands qu'elles ont visités sur la période ; puis on projette pour obtenir un réseau Marchands. Comme précédemment, on calcule les communautés du réseau Marchands et pour chaque marchand, ses agrégats comme indiqué dans la section 4 : nombre et montant des transactions sur la période, nombre et montant des transactions frauduleuses sur la période, taux de fraude, nombre de cartes et nombre de cartes frauduleuses vues par le marchand.

L'algorithme d'extraction des communautés (Blondel et al., 2008) produit une décomposition hiérarchique : au niveau 0, les marchands sont regroupés en communautés. Ensuite on construit le réseau de communautés dont les nœuds sont les communautés (du niveau 0) et les liens entre deux communautés sont pondérés par la somme des poids des liens entre nœuds de ces communautés. On peut alors extraire les communautés (dites de niveau 1) de ce nouveau réseau. Et ainsi de suite aux niveaux successifs. Nous avons effectué cette décomposition hiérarchique avec le module Social du logiciel InfiniteInsight^{TM} de KXEN et calculé les agrégats

FIG. 6 – *Variables les plus significatives pour les modèles de 5 segments avec 997 variables.*

et variables sociales comme dans la section 5.3.3 précédente. Nous exportons ensuite les réseaux de communautés aux différents niveaux dans le logiciel de visualisation de graphe Gephi (Gephi est un logiciel *open source* disponible sur le site http://gephi.org). Dans Gephi, nous pouvons alors visualiser le réseau Marchands en exploitant différentes fonctions du logiciel : couleurs et tailles des nœuds notamment. En navigant dans la hiérarchie des réseaux, nous pouvons explorer un niveau pour détecter un phénomène intéressant, puis descendre au niveau inférieur pour obtenir les détails (*drill-down*).

Nous pouvons tout d'abord choisir une couleur du nœud (communauté ou marchand) pour représenter le taux de fraude (rouge : taux de fraude élevé, vert : taux faible) et la taille en fonction du nombre de marchands dans la communauté. La visualisation dans Gephi fait par exemple apparaître (figure 7) un ensemble de communautés dont le taux de fraude varie au cours du temps (janvier à avril 2009), avec l'apparition de forts taux de fraude dans certaines communautés jusque-là assez peu fraudées. Notons qu'un travail d'analyse plus fin sur l'évolution temporelle reste à faire pour suivre les communautés au cours du temps ((Seifi et Guillaume, 2012)).

On peut ensuite zoomer sur une communauté particulière et visualiser les marchands de cette communauté en choisissant un code couleur pour représenter le montant de la fraude du marchand (de vert, pas de fraude, à rouge fort montant de fraude). La figure 8 (à gauche) montre ainsi une communauté de marchands ayant tous eu des montants de fraude très élevés et fortement connectés. Si on choisit de baser la coloration sur le taux de transactions frauduleuses, on voit (figure 8 à droite) que cette même communauté contient en fait un ensemble de

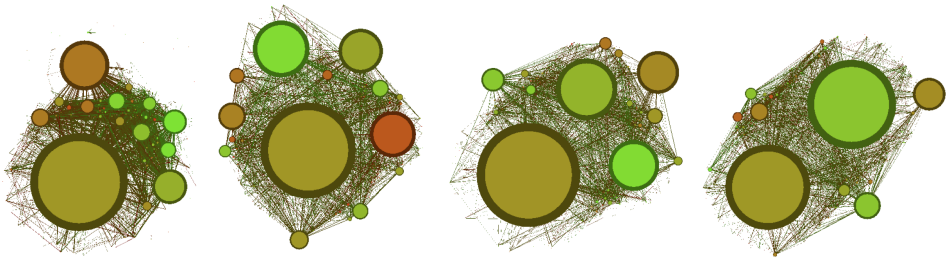

FIG. 7 – *Communautés du réseau marchands au cours du temps (Gephi).*

5 sites qui concentrent les plus forts taux de transactions frauduleuses (les identifiants des marchands sont cryptés) ; les autres sites ayant des taux faibles (mais des montants élevés, parce qu'ils font beaucoup de transactions). Remarquons de plus que les marchands à fort taux de fraude sont très connectés (et notamment 3 d'entre eux), c'est-à-dire qu'ils ont vu beaucoup des mêmes cartes. On devrait donc ici lancer une exploration plus approfondie sur ces 5 sites pour voir si les marchands impliqués ne seraient pas complices de la fraude ou bien auraient des failles graves de sécurité de leurs sites.

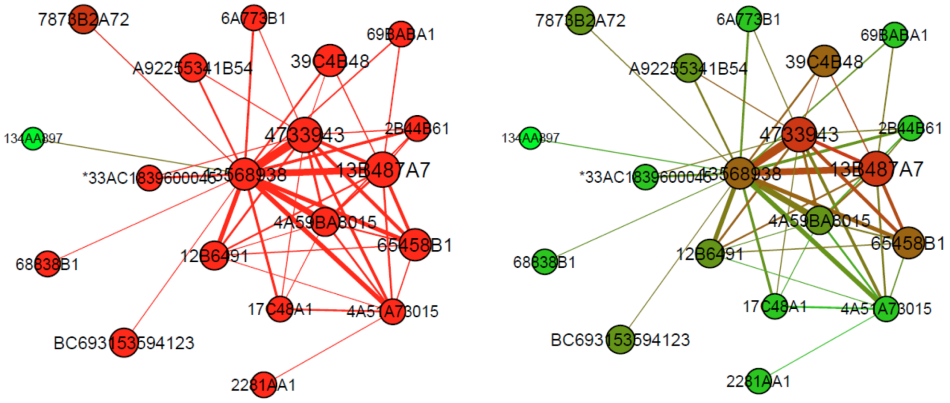

FIG. 8 – *Marchands d'une communauté colorés avec le montant de la fraude, à gauche et le taux de fraude à droite (Gephi).*

Nous pouvons donc mettre en place un processus d'investigation nous permettant de naviguer dans la décomposition hiérarchique en communautés : on identifie une communauté d'intérêt (fort taux ou fort montant de fraude), puis on descend au niveau des marchands, on filtre éventuellement sur le poids des liens entre marchands (en ne retenant que les marchands ayant vu beaucoup de cartes communes) et on fait ainsi apparaître des groupes de marchands très connectés ayant vu beaucoup des mêmes cartes fraudées. Ces groupes de marchands peuvent alors faire l'objet d'une procédure d'enquête de police approfondie.

La visualisation du réseau social offre une façon très intuitive de comprendre les relations entre les marchands impliqués dans la fraude : la démarche est très exploratoire et permet à l'enquêteur d'analyser les cartes signalées ou de repérer simplement les situations qui auraient pu ne pas encore être signalées. Nous allons maintenant présenter une autre méthode permettant de comprendre la dynamique de la fraude.

6.2 Dynamique de la fraude

Nous allons maintenant chercher, pour un lot de cartes signalées, à identifier la séquence temporelle des actions ayant conduit à la fraude : nous chercherons en particulier à déterminer les points de compromission et les points de test.

À partir du lot de cartes, on recherche l'ensemble des transactions (frauduleuses et non frauduleuses) effectuées par ces cartes sur la période considérée.

On caractérise ensuite les marchands qui ont été visités par un pourcentage important de cartes du lot (*point commun*). On définit trois rôles possibles dans la fraude :

- **Point de fraude** : les cartes du lot sont utilisées frauduleusement chez ce marchand (le taux de fraude du marchand est élevé).
- **Point de Compromission** : les numéros de cartes ont été volés chez ce marchand (le marchand a été vu avant la fraude, il y a peu ou pas de fraude chez lui).
- **Point de test** : les cartes ont été testées chez ce marchand, pour vérifier que la carte est utilisable (le marchand a été vu avant la fraude, il y a peu ou pas de fraude chez lui, les montants des transactions sont faibles).

On cherche ensuite les séquences temporelles de points communs les plus fréquentes, en utilisant le module de règles d'association avec contrainte temporelle du logiciel InfiniteInsight™ de KXEN.

Par exemple, avec des séquences de longueur 3, on obtient des règles du genre :

$$A\&B \Rightarrow C \tag{5}$$

Cette séquence indique qu'un nombre important de cartes ont vu le marchand C *après* avoir vu les marchands A et B. Ce nombre de cartes est déterminé par le support de la règle défini par :

$$\text{supp}(A\&B \Rightarrow C) = \frac{\text{nb de cartes ayant acheté chez A, B et C}}{\text{nb total de cartes}} \tag{6}$$

On retient toutes les séquences de support supérieur à un seuil (par exemple 3, pour garder le plus de séquences possible).

Les lots de cartes sont en général de petite taille (une centaine de cartes au plus), et donc le nombre de transactions effectuées sur la période est en général limité (quelques milliers).

Une fois cette caractérisation effectuée, on représente les règles calculées comme un réseau social : un nœud est un marchand, un lien est une règle, le poids du lien est le support de la règle. On peut alors visualiser ce réseau dans Gephi. On obtient une représentation (figure 9) montrant les marchands : points de compromission éventuels (en violet), points de test (en orange) et points de fraude (en rouge). La figure indique la séquence temporelle des actions (de gauche à droite), le label des liens indiquant le support de la règle (figure 9 à gauche) et la durée entre la visite d'un marchand et le suivant (figure 9 à droite). On voit par exemple que sur les cartes ayant visité le marchand 9913698866 (les identifiants des marchands sont cryptés),

19% ont été testées 1 mois après chez le marchand 6B6B51476, puis ont été fraudées 1 jour après chez le marchand 13B6733, alors que 19% ont été fraudées directement, 1 mois après, chez ce même marchand. Toutes les autres cartes ont été fraudées directement, plusieurs mois après leur visite au point de compromission.

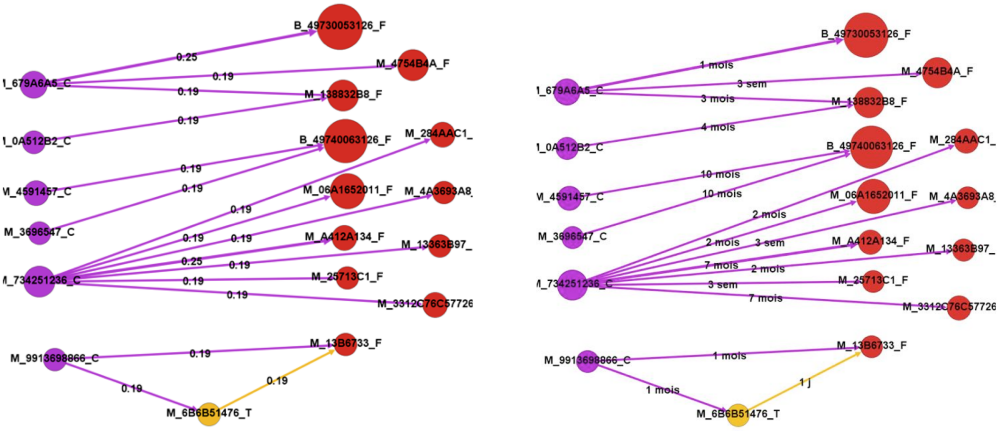

FIG. 9 – *Graphe des séquences temporelles ayant conduit à la fraude (Gephi).*

L'analyse du nombre de cartes ayant vu le marchand avant d'être fraudées au cours du temps chez un marchand point de compromission potentiel (figure 10) permet ensuite de définir la période de compromission.

L'ensemble de ces résultats donne à l'enquêteur les moyens de déterminer très rapidement les informations dont il a besoin pour lancer d'éventuelles actions de police.

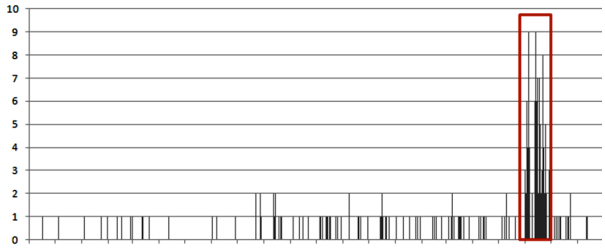

FIG. 10 – *Nombre de futures cartes fraudées vues par le marchand au cours du temps.*

Les scénarios d'investigation que nous avons présentés démontrent l'intérêt de l'approche par analyse de réseaux sociaux : la visualisation des réseaux Marchands ou du réseau des règles d'association temporelles permet d'identifier de façon très intuitive les comportements de fraude. L'enquêteur peut travailler en mode semi-automatique ou, au contraire, explorer de façon approfondie les réseaux de marchands de façon à extraire les groupes de marchands suspects.

7 L'apport de KXEN

KXEN commercialise une solution de data mining InfiniteInsight^{TM} exploitant certains résultats des théories de l'apprentissage statistique développés par (Vapnik, 1995) : KXEN a intégré les principes de la minimisation structurelle du risque (SRM : Structural Risk Minimization) pour automatiser le codage des données et la production de modèles robustes (http://www.kxen.com).

Cette solution permet de mettre en œuvre des «usines à modèles» ((Fogelman Soulié et Marcadé, 2008)) en automatisant le codage des variables et la production du modèle tout en contrôlant la robustesse de la solution. Le logiciel accepte les gros volumes de données et comprend les deux modules qui ont été utilisés dans les travaux présentés ici (figure 11) :

- **Modeler** : ce module permet de construire des modèles de classification/régression ; de segmentation ou des règles d'association.
- **Social** : ce module permet de construire des réseaux sociaux, y compris les réseaux bipartites, de calculer les communautés (avec un algorithme issu de (Blondel et al., 2008)) et d'extraire les variables sociales.

L'automatisation des traitements dans ce logiciel nous a permis dans le travail présenté ici de réaliser de très nombreux modèles rapidement pour les comparer et en tester les performances.

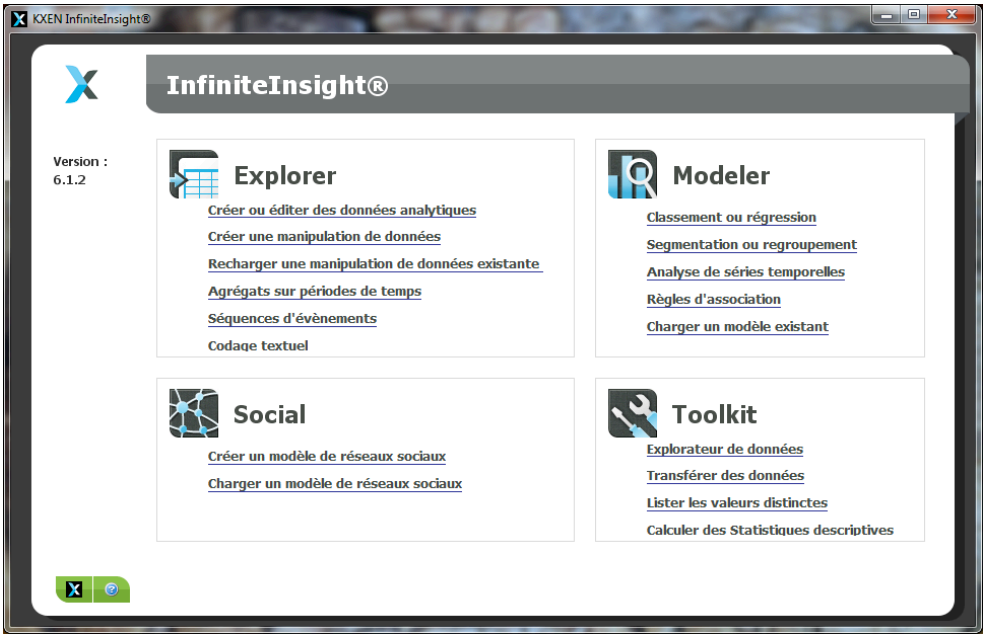

FIG. 11 – *Le logiciel InfiniteInsight^{TM} de KXEN.*

8 Conclusion

Nous avons présenté ici quelques utilisations de l'analyse des réseaux sociaux pour la détection et l'investigation de la fraude. Nous avons ainsi introduit deux grands modes d'utilisation :

- **En mode prédictif**, les transactions permettent de construire deux réseaux sociaux, Cartes et Marchands, dont on peut extraire des variables sociales. Ces variables viennent enrichir les divers agrégats calculés et apportent une forte augmentation des performances des modèles de détection (doublement de la pertinence) ;
- **En mode exploratoire**, on peut réaliser un co-clustering des deux réseaux Cartes et Marchands qui fait apparaître des groupes de cartes et marchands «travaillant ensemble» ; pour l'investigation de la fraude, le réseau Marchands construit à partir d'un lot de cartes signalées fait apparaître des groupes de marchands impliqués dans la fraude sur les mêmes cartes ; par ailleurs, des séquences temporelles visualisées sous forme de graphe mettent en évidence les séquences d'actions menant à la fraude : compromission des numéros de cartes, tests des cartes, puis fraude.

L'ensemble de ces traitements est en cours de test opérationnel chez le GIE CB.

Remerciements

Ce travail a été financé par un contrat ANR-CSOSG pour le projet eFraudBox.

Références

Aggarwal, C. et H. Wang (2010). *Managing and Mining Graph Data*. Springer.

Barabasi, A.-L. (2002). *Linked*. Plume, Penguin Group.

Blondel, V. D., J.-L. Guillaume, L. R., et L. E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. P10008*.

Bolton, R. J. et D. J. Hand (2002). Statistical fraud detection. a review. *Statistical Science 17*(3), 235–255.

Chapus, B., F. Fogelman Soulié, E. Marcadé, et J. Sauvage (2011). Mining on social networks. in statistical learning and data science. In M. Gettler Summa, L. Bottou, B. Goldfarb, et F. Murtagh (Eds.), *Computer Science and Data Analysis Series*. CRC Press, Chapman & Hall.

Consortium e-Fraud Box (2011). e-fraud box : Détection et investigation de la fraude à la carte bancaire sur internet.

Erdös, P. et A. Rényi (1959). On random graphs. *Publicationes Mathematicae 6*, 290–297.

Fevad (2012). Chiffres clés 2012.

Fevad (2013). Bilan e-commerce 2012.

FiaNet (2010). Livre blanc 2010 : La fraude à la carte bancaire sur internet.

Fogelman Soulié, F. et E. Marcadé (2008). Industrial mining of massive data sets. In F. Fogelman Soulié, D. Perrotta, J. Pikorski, et R. Steinberger (Eds.), *Mining massive Data Sets*

for Security. Advances in data mining, search, social networks and text mining and their applications to security, NATO ASI Series, pp. 44–61. IOS Press.

Fogelman Soulié, F., A. Mekki, et S. Sean (2011). Using social networks for on-line credit card fraud analysis. In A. A. Ekrem Duman (Ed.), *Use of Risk Analysis in Computer-Aided Persuasion*, Volume 88 of *NATO Science for Peace and Security Series E*, pp. 60–72. IOS Press.

Fortunato, S. (2009). Community detection in graphs. *Physics Reports 486*, 75–174.

Hand, D. J. et D. J. Weston (2008). Statistical techniques for fraud detection, prevention and assessment. In F. Fogelman Soulié, D. Perrotta, J. Pikorski, et R. Steinberger (Eds.), *Mining massive Data Sets for Security. Advances in data mining, search, social networks and text mining and their applications to security.*, NATO ASI Series, pp. 257–270. IOS Press.

Hassibi, K. (2000). Detecting payment card fraud with neural networks. In *Business applications of neural networks : the state-of-the-art of real-world applications*, Volume 13 of *Progress in Neural Processing*, Chapter 9, pp. 141–158. World Scientific Publishing.

Herschel, G. (2007). Magic quadrant for customer data mining 2q07.

Kleinberg, J. M. (1997). Authoritative sources in a hyperlinked environment. In *Proc. 9th ACM-SIAM Symposium on Discrete Algorithms*. Extended version in Journal of the ACM 46 (1999). Also appears as IBM Research Report RJ 10076.

Memon, N. et D. L. Hicks (2008). Detecting core members in terrorist net-works : a case study. In F. Fogelman Soulié, D. Perrotta, J. Pikorski, et R. Steinberger (Eds.), *Mining massive Data Sets for Security. Advances in data mining, search, social networks and text mining and their applications to security.*, NATO ASI Series, pp. 345–356. IOS Press.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM REVIEW 45*, 167–256.

OSCP (2011a). Commerçants, comment renforcer la sécurité des paiements sur internet ? Observatoire de la sécurité des cartes de paiements.

OSCP (2011b). Rapport annuel. Observatoire de la sécurité des cartes de paiements.

Pandit, S., D. H. Chau, S. Wang, et C. Faloutsos (2007). Netprobe : a fast and scalable system for fraud detection in online auction networks. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, New York, NY, USA, pp. 201–210. ACM.

Ressler, S. (2006). Social network analysis as an approach to combat terrorism : Past, present and future research. *New York 2*(2), 1–10.

Seifi, M. et J.-L. Guillaume (2012). Community cores in evolving networks. In *Proceedings of the Mining Social Network Dynamic 2012 Workshop (MSND). In conjunction with the international conference World Wide Web WWW 2012*, Lyon, France, pp. 1173–1180.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag.

Vapnik, V. (2006). *Estimation of Dependencies Based on Empirical Data (reprint of 1982 Edition)*. Springer-Verlag.

Wasserman, S. et K. Faust (1994). *Social network analysis : methods and applications*. Cambridge University Press.

Watts, D. (2003). *Six Degrees, the science of a connected age*. W.W. Norton & Company.

Summary

Electronic commerce, in rapid growth, becomes a major target for fraudsters. Credit card fraud on the Internet is mostly done by international criminal networks. Classical fraud detection systems are based, since the 80s, on data mining techniques. Nowdays, social network analysis techniques are very well described in the litterature. We present in this paper the results obtained in during the French project eFraudBox, in collaboration with the GIE Cartes Bancaires.