

Prédiction des liens dans les réseaux sociaux dans le cadre de la théorie des fonctions de croyance

S. Mallek¹

I. Boukhris¹

Z. Elouedi¹

E. Lefèvre²

¹ Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC, Tunisie

² Univ. Lille Nord de France, UArtois EA 3926 LGI2A, France

sabrinemallek@yahoo.fr, imen.boukhris@hotmail.com, zied.elouedi@gmx.fr

eric.lefevre@univ-artois.fr

Résumé :

Le problème de prédiction de liens constitue un domaine de recherche important traité dans l'analyse des réseaux sociaux. Il consiste à deviner les liens qui apparaîtront dans le réseau en se basant sur son état actuel. Plusieurs méthodes ont été proposées pour faire face à ce problème mais la plupart d'entre elles le traitent dans un cadre certain. En effet, les données des réseaux sociaux sont souvent incomplètes et bruitées, par conséquent il est nécessaire de gérer l'incertitude lors de la prédiction. Nous examinons dans ce papier, le problème de prédiction de liens dans un cadre incertain en utilisant la théorie des fonctions de croyance. D'abord, nous présentons un nouveau modèle graphique pour les réseaux sociaux qui encapsule les incertitudes dans la structure des liens. Ensuite, nous proposons une nouvelle approche pour la prédiction des futurs liens via la fusion d'information des nœuds voisins.

Mots-clés :

Prédiction de liens, analyse des réseaux sociaux, théorie des fonctions de croyance, réseau social incertain.

Abstract:

The link prediction problem is an important research area handled in social network analysis. It consists of inferring the potential links to be formed in the future given a current snapshot of the network. Several methods have been proposed to address this problem but most of them consider it under a certain framework. Yet, social networks data are often incomplete and noisy, thus it is necessary to manage uncertainty in the prediction task. We review in this paper, the link prediction problem under uncertainty using the belief function theory. First, we present a new graph based model for social networks that encapsulates the uncertainties in the link structure. Then, we propose a new approach for link prediction through information fusion of the neighboring nodes.

Keywords:

Link prediction, social network analysis, belief function theory, uncertain social network

1 INTRODUCTION

Les réseaux sociaux sont de très grands systèmes permettant de représenter les interactions sociales entre des millions d'individus.

Ils sont souvent modélisés par des graphes, où les nœuds correspondent à des personnes et les arêtes représentent une certaine forme d'interaction entre elles. L'analyse des réseaux sociaux a surgi comme un outil important pour surveiller et analyser ces réseaux. C'est une collection de méthodes spécialement conçues pour examiner les aspects relationnels des structures sociales. L'objectif principal est de déterminer les conditions selon lesquelles la structuration des liens sociaux se posent et de découvrir leurs conséquences.

En effet, les réseaux sociaux sont des structures très dynamiques où de nouveaux nœuds et liens sont formés en permanence. L'un des problèmes intéressants traités dans l'analyse des réseaux sociaux, permettant la compréhension de l'évolution de ces derniers concerne le problème de prédiction de liens. Il consiste à prédire les associations futures entre une paire de nœuds sachant qu'il n'y a pas de lien entre eux dans l'état actuel du graphe. Cependant, la plupart des techniques existantes traitant ce problème sont conçues pour faire face aux réseaux sociaux dans un cadre certain, elles supposent des liens avec des relations binaires, soit 1 (existe) ou 0 (n'existe). Néanmoins, la structure de ces réseaux repose sur la nature précise des données. Comme indiqué dans [6, 9], les données issues des réseaux sociaux sont sujettes à des erreurs d'observations, et sont souvent incomplètes et bruitées, c-à-d des nœuds et/ou liens sont manquants dans les données, ce qui affecte les résultats d'analyse. Ainsi, deux problèmes

apparaissent : le premier est de considérer tous les liens et les nœuds et de risquer la possibilité d'incorporer à tort de faux liens et nœuds dans le réseau. Le second consiste à éliminer tous les liens et nœuds incertains mais dans ce cas nous nous privons d'informations potentiellement pertinentes [9]. Pour cette raison, nous proposons d'intégrer l'incertitude dans la structure du graphe du réseau social.

La plupart des études existantes considèrent des réseaux pondérés avec des valeurs entières. Cependant, une autre façon de représenter un réseau incertain consiste à pondérer les liens avec des valeurs dans $[0, 1]$ pour coder les degrés d'incertitude [7]. En effet, plusieurs réseaux sociaux du monde réel sont caractérisés par des degrés d'incertitude et plus particulièrement les réseaux sociaux de grande échelle, comme souligné dans [6]. Pour cette raison, nous utilisons la théorie des fonctions de croyance [2, 11] comme cadre général pour le raisonnement incertain. Nous utilisons ses outils pour le traitement de données imprécises et la modélisation de l'ignorance totale et partielle pour quantifier les degrés d'incertitude au niveau des arêtes du graphe du réseau social.

De plus, nous développons une nouvelle approche pour la prédiction de liens qui prend en compte l'aspect incertain décrivant les réseaux sociaux. Notre approche est différente des méthodes existantes car elle opère seulement avec les outils des fonctions de croyance. Les degrés d'incertitude sur l'existence des liens sont étendus et combinés en tant que sources d'information indépendantes. Une procédure de correspondance et de fusion est ensuite appliquée afin de prédire l'existence d'un nouveau lien dans le réseau entre une paire de nœuds.

La suite du papier est structurée de la façon suivante : la Section 2 donne un aperçu du problème de prédiction de liens. Après avoir rappelé les concepts de base de la théorie des fonctions de croyance au niveau de la Section 3, nous introduisons notre nouveau modèle graphique pour la représentation d'un réseau social

incertain à la Section 4. La Section 5 expose l'approche proposée pour la prédiction des liens dans un cadre incertain. La Section 6 illustre la nouvelle méthode proposée.

2 LA PRÉDICTION DE LIENS

La prédiction de liens est appliquée dans une grande variété de domaines tels que l'analyse des liens, la bioinformatique, la recherche d'information [8], etc. Par exemple, on pourrait prévoir des futures amitiés lors de l'analyse des réseaux sociaux ou prédire les futurs co-auteurs dans un réseau de collaboration [13]. Formellement, la tâche de prédiction de liens peut être formulée comme suit [4] :

Etant donné un réseau social $G(V, E)$ où V est l'ensemble des nœuds qui peuvent être de différents types (individus, organisations, entreprises, etc) et E est l'ensemble des arêtes les reliant à travers un type d'interdépendance (amitié, échange financier, proximité physique, etc). Une arête entre une paire de nœuds $(v_i, v_j) \in V$ représente une association qui a eu lieu à un moment donné t . La tâche est de prédire l'ensemble des liens potentiels qui peuvent être formés à l'instant $t + 1$.

La plupart des méthodes de prédiction de liens de l'état de l'art reposent sur deux groupes d'information du réseau qui peuvent être classés en information locale (basée sur les nœuds voisins) et information globale (basée sur les chemins dans le graphe).

Les approches reposant sur l'information locale utilisent les similarités locales qui caractérisent les nœuds dans le réseau. Ces dernières peuvent être les attributs essentiels à savoir le genre, l'âge, les intérêts, ou des indices structurels basés uniquement sur la structure du réseau, par exemple les voisins communs entre deux nœuds. Cependant, les attributs des nœuds ne sont pas généralement disponibles ou sont cachés [12]. Pour cette raison, la majorité des approches utilisent seulement les mesures reposant sur les similarités structurelles. Les ap-

proches globales utilisent des mesures basées sur l'ensemble des chemins entre les nœuds dans le réseau afin de déterminer ceux qui sont plus proches. L'intuition est que plus les nœuds sont proches dans le réseau, plus ils ont tendance à être liés ou à s'influencer dans l'avenir. L'avantage principal de ces deux types de mesures est qu'elles sont génériques et par conséquent peuvent être appliquées sur des graphes de différents domaines [8]. Ainsi, nous rappelons dans la suite de cette section certaines mesures de l'état de l'art basées sur les informations locales et globales.

2.1 LES APPROCHES BASÉES SUR L'INFORMATION LOCALE

Ces approches reposent sur des mesures basées sur les nœuds voisins. Parmi les mesures que l'on retrouve couramment dans la littérature, on peut citer la mesure des voisins communs, notée $CN(v_i, v_j)$, qui indique le nombre de voisins communs d'une paire de nœuds (v_i, v_j) dans le réseau social. Étant donné l'ensemble des voisins d'un nœud v_i dans le graphe, noté $\tau(v_i)$, la mesure des voisins communs est calculée comme suit : $CN(v_i, v_j) = |\tau(v_i) \cap \tau(v_j)|$. Newman [13] a utilisé cette mesure dans le contexte des réseaux de collaboration, en supposant une corrélation entre le nombre de voisins communs, et la probabilité que deux auteurs collaboreront dans un travail futur. Le coefficient de Jaccard prend en compte tous les voisins d'une paire de nœuds (v_i, v_j) . Il est calculé comme suit : $JC(v_i, v_j) = \frac{|\tau(v_i) \cap \tau(v_j)|}{|\tau(v_i) \cup \tau(v_j)|}$. La mesure de Adamic/Adar notée $AA(v_i, v_j)$ pondère chaque voisin commun v_k par $\frac{1}{\log|\tau(v_k)|}$ pour mesurer sa contribution, elle est définie par : $AA(v_i, v_j) = \sum_{v_k \in (\tau(v_i) \cap \tau(v_j))} \frac{1}{\log|\tau(v_k)|}$.

2.2 LES APPROCHES BASÉES SUR L'INFORMATION GLOBALE

Elles sont basées sur des mesures telles que la distance du chemin le plus court, le temps de déplacement moyen (ACT), l'indice SimRank, etc [8]. Par exemple, la distance du che-

min le plus court est basée sur le fait que plus la distance entre deux nœuds est courte, plus la probabilité que ces derniers soient connectés est élevée. L'indice SimRank repose sur l'hypothèse que deux nœuds sont liés s'ils sont liés à des nœuds similaires. L'ACT est basée sur les marches aléatoires dans le graphe, elle représente le nombre moyen d'étapes $m(v_i, v_j)$ faite par un marcheur aléatoire en commençant par le nœud v_i pour atteindre v_j . L'ACT entre v_i et v_j est $n(v_i, v_j) = m(v_i, v_j) + m(v_j, v_i)$ [12].

Les approches basées sur les chemins utilisent l'information topologique globale du réseau, elles opèrent généralement mieux que les méthodes basées sur les nœuds voisins mais elles présentent deux inconvénients : le premier est que le calcul d'un indice global est coûteux en termes de complexité temporelle, plus particulièrement pour les réseaux de grandes échelles. Le deuxième inconvénient est que l'information topologique globale n'est pas toujours accessible [12]. Par conséquent, nous proposons dans ce papier une approche pour la prédiction des liens basée sur l'intuition des méthodes utilisant les nœuds voisins dans un cadre incertain.

3 LA THEORIE DES FONCTIONS DE CROYANCE

La théorie des fonctions de croyance introduite par Dempster [2] et généralisée par Shafer [11], est une théorie utile pour la représentation et la gestion des données incertaines. Elle permet de faire face à l'incertitude et l'imprécision caractérisant les données et de les gérer de manière flexible. Soit $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$, appelé le cadre de discernement, un ensemble exhaustif et fini d'événements mutuellement exclusifs liés à un problème donné. L'ensemble, noté 2^Θ , comporte tous les sous-ensembles possibles formés des événements et unions d'événements de Θ . Il comprend aussi l'ensemble vide \emptyset qui correspond à la proposition impossible ou le conflit. Une distribution de masses de croyance (*bba* : basic belief assign-

ment) notée m , représente la part de croyance sur les sous-ensembles du cadre de discernement Θ . Elle est définie par :

$$\begin{aligned} m : 2^\Theta &\rightarrow [0, 1] \\ \sum_{A \subseteq \Theta} m(A) &= 1 \end{aligned} \quad (1)$$

Afin de combiner deux fonctions de masse m_1 et m_2 distinctes et fiables définies sur Θ , la règle de combinaison conjonctive, notée \odot , est utilisée. Elle est définie comme suit [16] :

$$m_1 \odot m_2(A) = \sum_{B, C \subseteq \Theta: B \cap C = A} m_1(B) \cdot m_2(C) \quad (2)$$

D'autre part, pour combiner deux fonctions de masse m_1 et m_2 définies sur deux cadres disjoints $\Theta = \{\theta_1, \dots, \theta_N\}$ et $\Omega = \{\omega_1, \dots, \omega_K\}$, l'extension vide est appliquée. Pour ce faire, les fonctions de masse doivent être étendues à l'espace produit $\Theta \times \Omega = \{(\theta_i, \omega_k), \forall k \in \{1, \dots, K\}, \forall i \in \{1, \dots, N\}\}$. L'opération d'extension vide, notée \uparrow , est définie par :

$$m^{\Theta \uparrow \Theta \times \Omega}(C) = \begin{cases} m^\Theta(A) & \text{si } C = A \times \Omega, \\ & A \subseteq \Theta, C \subseteq \Theta \times \Omega \\ 0, & \text{sinon} \end{cases} \quad (3)$$

Étant donnés deux cadres de discernement, nous pouvons appliquer un mapping multivalué afin de spécifier la relation entre eux [2]. En d'autres termes, un mapping multivalué noté τ associe à deux cadres de discernement disjoints Θ et Ω les sous-ensembles $S_i \subseteq \Omega$ qui peuvent éventuellement correspondre selon τ aux sous-ensembles $A_i \subseteq \Theta$:

$$m_\tau(A_i) = \sum_{\tau(S_i)=A_i} m(S_i) \quad (4)$$

Le Modèle des Croyances Transférables (TBM), proposé par Smets [15], est l'un des modèles d'interprétation des fonctions de croyance. La prise de décision est effectuée au niveau pignistique où les croyances sont transformées en utilisant des mesures de probabilités

pignistiques notées $BetP$ [14] :

$$BetP(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{(1 - m(\emptyset))}, \quad (5) \quad \forall A \in \Theta$$

4 RÉSEAU SOCIAL INCERTAIN

Un réseau social est généralement modélisé par un graphe $G = (V, E)$ où V est l'ensemble des nœuds et E est l'ensemble des arêtes. Cependant, une telle représentation ne prend pas en compte les imperfections qui résultent des données imprécises et incertaines ou dues à l'usage d'outils de construction non fiables. Comme indiqué dans [7], il est important d'intégrer l'incertitude dans les réseaux sociaux, en particulier ceux générés à partir de données textuelles. Les auteurs ont proposé de coder les degrés d'incertitude au niveau des liens en utilisant des probabilités. Cependant, cette structure de réseau social est seulement concevable dans le contexte de réseaux sociaux construits à partir de données textuelles. De plus, les auteurs dans [17] ont proposé un réseau social dont les nœuds, les liens et les messages sont pondérés par des fonctions de masses de croyance dans le but de déduire la nature d'un message qui circule dans le réseau. Néanmoins, notre objectif dans ce travail est de traiter l'incertitude rencontrée au niveau des liens. À cet égard, nous introduisons notre réseau social incertain qui est codé par des distributions de masses de croyance. En effet, nous attribuons à chaque lien $v_i v_j$ une fonction de masse de croyance définie sur le cadre de discernement $\Theta^{v_i v_j} = \{E_{v_i v_j}, \neg E_{v_i v_j}\}$ notée $m^{v_i v_j}$, où $E_{v_i v_j}$ signifie que le lien entre v_i et v_j existe et $\neg E_{v_i v_j}$ signifie qu'il est absent. Ainsi, un graphe modélisant un réseau social incertain est défini par $G(V, E)$ où : $V = \{v_1, \dots, v_{|V|}\}$ est l'ensemble des nœuds et E est l'ensemble des arêtes. Une paire $(v_i v_j, m^{v_i v_j})$ est assignée à chaque arête $v_i v_j \in E$ où $v_i, v_j \in V$, $v_i \neq v_j$, et $m^{v_i v_j}$ est une fonction de masse qui encode les degrés d'incertitude concernant l'existence

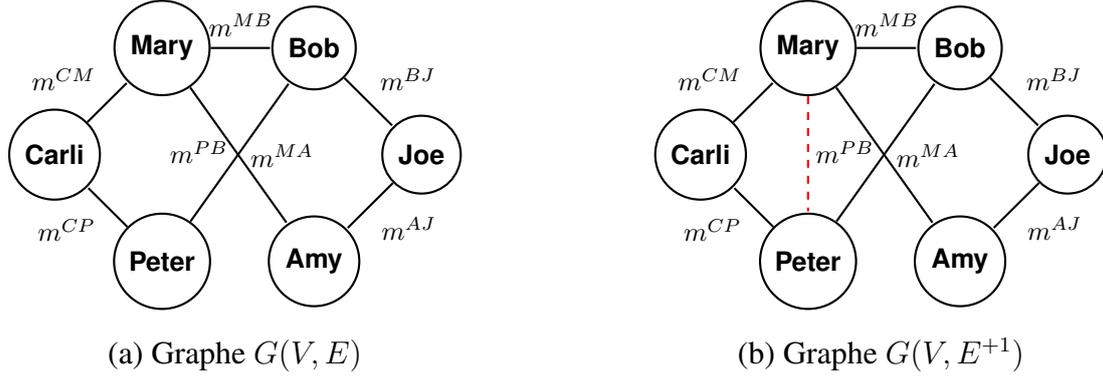


Figure 1 – Un réseau social avec des arêtes pondérées par des fonctions de masse à t (a) et $t + 1$ (b)

et l'absence d'un lien entre v_i et v_j . Un exemple d'une telle structure de graphe est donnée dans la Figure 1(a). Pour plus de clarté, un lien entre une paire de nœuds (v_i, v_j) est représenté si la probabilité pignistique $BetP^{v_i v_j}(E_{v_i v_j}) > 0.5$. En effet, $BetP^{v_i v_j}(E_{v_i v_j}) > 0.5$ signifie que la probabilité que le lien existe entre v_i et v_j est supérieure à 50%.

Ainsi, comme présenté dans la Figure 1, au lieu d'avoir des liens pondérés par des valeurs qui peuvent être soit 1 soit 0 pour exprimer leur existence, nous attribuons une fonction de masse avec des valeurs dans $[0, 1]$ pour quantifier le degré d'incertitude concernant l'existence d'un lien.

5 PREDICTION DE LIENS DANS LE CADRE DES FONCTIONS DE CROYANCE

L'objectif est de déduire les futurs liens en se basant sur un état actuel d'un réseau social. Pour ce faire, nous nous appuyons sur des méthodes basées sur les voisins communs. En effet, c'est une façon très simple de résoudre le problème de prédiction. De plus, les algorithmes basés sur ces mesures se sont avérés efficaces dans des réseaux sociaux de plusieurs domaines [10, 13]. Les nœuds ayant plus de voisins communs sont plus susceptibles d'avoir une connexion future. De ce fait, nous examinons le problème de prédiction de liens dans un cadre incertain en utilisant la théorie des fonc-

tions de croyance. Les nœuds communs sont considérés comme des sources d'information indépendantes. Les degrés d'incertitude qui les décrivent sont transférés et fusionnés afin de déduire le degré d'existence d'un nouveau lien. Nous formulons le problème de prédiction de liens comme suit. Étant donné l'état actuel du graphe $G(V, E)$ à l'instant t , l'objectif est de prédire l'existence d'un lien $v_i v_j$ dans le nouvel ensemble d'arêtes E^{+1} entre une paire de nœuds v_i et $v_j \in V$ à l'instant $t + 1$, en prenant en considération les liens partagés entre v_i, v_j et leurs voisins directs à l'instant t .

Il faut noter que tous les liens partagés entre chaque voisin commun v_k et les nœuds v_i et v_j dont $m^{v_k v_i}\{\neg E_{v_k v_i}\} \neq 1$ et $m^{v_k v_j}\{\neg E_{v_k v_j}\} \neq 1$ sont considérés dans la tâche de prédiction, même si ceux-ci ne sont pas représentés schématiquement sur le graphe, c-à-d $BetP^{v_k v_i}(E_{v_k v_i}) \leq 0.5$ ou $BetP^{v_k v_j}(E_{v_k v_j}) \leq 0.5$. Ainsi, nous proposons un mécanisme global pour la prédiction d'un nouveau lien entre une paire de nœuds (v_i, v_j) composé des quatre étapes suivantes.

5.1 Étape 1 : Acquisition d'informations

Pour chaque voisin commun v_k , étendre les cadres de chaque lien $v_k v_i$ et $v_k v_j$ au cadre de discernement joint noté Θ^{N_k} où $\Theta^{N_k} = \Theta^{v_k v_i} \times \Theta^{v_k v_j}$ en utilisant l'équation 3. Cette étape est essentielle car elle permet de travailler sur un référentiel unifié. Les fonctions de masse sont ensuite combinées conjonctivement pour obte-

nir les masses des paires issues de la combinaison qui sont incluses dans Θ^{N_k} .

5.2 Étape 2 : Transfert d'informations

Afin de transférer les fonctions de masses de croyances obtenues sur $\Theta^{v_i v_j}$, nous appliquons une opération de mapping multivalué notée τ telle que $\tau : \Theta^{N_k} \rightarrow 2^{\Theta^{v_i v_j}}$. Elle rassemble les paires de combinaison comme suit :

- Les masses $m\{(E_{v_k v_i}, E_{v_k v_j})\}$, $m\{(E_{v_k v_i}, \Theta^{v_k v_j})\}$ et $m\{(\Theta^{v_k v_i}, E_{v_k v_j})\}$ sont transférées à $m\{E_{v_i v_j}\}$;
- Les masses $m\{(\neg E_{v_k v_i}, \neg E_{v_k v_j})\}$, $m\{(\neg E_{v_k v_i}, \Theta^{v_k v_j})\}$ et $m\{(\Theta^{v_k v_i}, \neg E_{v_k v_j})\}$ sont transférées à $m\{\neg E_{v_i v_j}\}$;
- Les masses $m\{(\neg E_{v_k v_i}, E_{v_k v_j})\}$, $m\{(E_{v_k v_i}, \neg E_{v_k v_j})\}$ et $m\{(\Theta^{v_k v_i}, \Theta^{v_k v_j})\}$ sont transférées à $m\{\Theta^{v_i v_j}\}$.

En appliquant ce mapping, nous considérons chaque noeud comme une source d'information distincte, la part de croyance décrivant la présence d'au moins l'un des deux liens $v_k v_i$ et $v_k v_j$ est affectée à l'élément "existe" de $v_i v_j$, et réciproquement pour l'événement "non existe". D'autre part, la présence d'un lien voisin et l'absence de l'autre est affecté à l'ignorance ($\Theta^{v_i v_j}$).

5.3 Étape 3 : Fusion des croyances

Afin d'obtenir $m^{v_i v_j}$, les fonctions de masse des n voisins communs obtenues après transfert sont combinées en utilisant la règle conjonctive (Equation 2) telles que :

$$m^{v_i v_j} = m_{v_1}^{v_i v_j} \odot m_{v_2}^{v_i v_j} \odot \dots \odot m_{v_n}^{v_i v_j} \quad (6)$$

$m_{v_n}^{v_i v_j}$ est la fonction de masse sachant le voisin commun v_n . Cette étape est fondamentale car elle permet de fusionner les informations fournies par les noeuds voisins et de traiter les liens partagés en tant que sources d'information indépendantes.

5.4 Étape 4 : Prise de décision

Finalement, nous calculons la probabilité pignistique $BetP^{v_i v_j}(E_{v_i v_j})$ en utilisant l'Équation 5 afin de prendre une décision sur l'existence du lien dans le graphe. En effet, si $BetP^{v_i v_j}(E_{v_i v_j}) > 0.5$, alors le lien entre v_i et v_j est susceptible d'exister avec une probabilité $> 50\%$ à $t + 1$, sinon il serait absent. En d'autres termes, la valeur de $BetP$ indique si le lien est plus probable ou pas.

Il est à noter que notre méthode est incrémentale. En effet, à chaque fois qu'un nouveau voisin commun est ajouté, la nouvelle information est facilement combinée avec l'état précédent du graphe après le transfert des masses et l'application des étapes 1 et 2 décrites dans la Section 5. Les résultats sont ensuite combinés avec les informations déjà traitées pour les anciens voisins communs sans passer par les étapes 1 et 2 pour ces derniers.

6 ILLUSTRATION

Supposons que le graphe de la Figure 1(a) décrit un réseau d'amitiés entre des personnes. L'incertitude peut être due aux informations partielles et incomplètes fournies par l'informateur. Par exemple, "J'ai vu la Personne A parler à B". Cependant, cela ne signifie pas que A et B sont connectés. D'autre part, l'incertitude peut survenir à cause de l'imperfection de l'information, c-à-d l'informateur n'est pas sûr de l'identité de la personne (A ou C). Pour illustrer notre approche de prédiction de liens, nous étudions l'existence d'un nouveau lien à l'instant $t + 1$ entre la paire de noeuds (*Mary*, *Peter*). Pour plus de clarté, nous proposons de représenter les noeuds par la première lettre des labels, par exemple le noeud *Mary* est représenté par *M*. Nous considérons donc tous leurs voisins communs V_k pour lesquels les masses $m^{MV_k}\{(\neg E_{MV_k})\}$ et $m^{PV_k}\{(\neg E_{PV_k})\}$ sont $\neq 1$. Supposons que nous avons des fonctions de masse allouées comme décrit dans le Tableau 1, où $\Theta^{link} = \{E_{link}, \neg E_{link}\}$.

L'ensemble des voisins communs est $N = \{C, B, A\}$. Par conséquent, les liens partagés avec chaque voisin commun sont : MC, PC, MB, PB, MA, PA .

Tableau 1 – Les fonctions de masse allouées aux liens du réseau social de la Figure 1(a)

Link	$m^{link}\{(E_{link})\}$	$m^{link}\{\neg E_{link}\}$	$m^{link}(\Theta^{link})$
MC	0.6	0.2	0.2
PC	0.5	0.2	0.3
MB	0.4	0	0.6
MA	0.35	0.25	0.4
PB	0.6	0.3	0.1
BJ	0.4	0.2	0.4
AJ	0.5	0.2	0.3
PA	0.3	0.45	0.25

Les étapes 1 et 2, présentées dans la Section 5, sont appliquées à tous les voisins communs. Nous considérons le nœud C en premier. L'espace produit $\Theta^{N_C} = \Theta^{PC} \times \Theta^{MC}$ contient les couples $\{(E_{PC}, E_{MC}), (E_{PC}, \neg E_{MC}), (\neg E_{PC}, E_{MC}), (\neg E_{PC}, \neg E_{MC})\}$.

Étape 1 En premier lieu, l'extension vide est calculée (Tableau 2). Ensuite, les masses sur Θ^{N_C} sont combinées en utilisant la règle de combinaison conjonctive (Equation 2) afin d'obtenir $m_{\odot}^{N_C} = m^{PC \uparrow N_C} \odot m^{MC \uparrow N_C}$. Le tableau 3 présente les résultats.

Tableau 2 – Extension vide entre m^{PC} et m^{MC}

$m^{PC \uparrow N_C}$	$\{E_{PC}\} \times \Theta^{MC}$	0.5
	$\{\neg E_{PC}\} \times \Theta^{MC}$	0.2
	$\Theta^{PC} \times \Theta^{MC}$	0.3
$m^{MC \uparrow N_C}$	$\Theta^{PC} \times \{E_{MC}\}$	0.6
	$\Theta^{PC} \times \{\neg E_{MC}\}$	0.2
	$\Theta^{PC} \times \Theta^{MC}$	0.2

Tableau 3 – Combinaison conjonctive $m_{\odot}^{N_C}$

$m_{\odot}^{N_C}$	$\{E_{PC}\}$	$\{\neg E_{PC}\}$	Θ^{PC}
$\{E_{MC}\}$	0.3	0.12	0.18
$\{\neg E_{MC}\}$	0.1	0.04	0.06
Θ^{MC}	0.1	0.04	0.06

Étape 2 L'étape suivante consiste à transférer les masses obtenues en utilisant la fonction τ (Equations 4). Ainsi, nous obtenons :

$$m_C^{MP}(\{E_{MP}\}) = 0.58, m_C^{MP}(\{\neg E_{MP}\}) = 0.14 \text{ et } m_C^{MP}(\Theta^{MP}) = 0.28.$$

Le même processus est appliqué au voisin B . L'espace produit : $\Theta^{N_B} = \Theta^{PB} \times \Theta^{MB} = \{(E_{PB}, E_{MB}), (E_{PB}, \neg E_{MB}), (\neg E_{PB}, E_{MB}), (\neg E_{PB}, \neg E_{MB})\}$.

Étape 1 L'extension vide entre m^{PB} et m^{MB} est décrite dans le Tableau 4. L'application de la règle de combinaison conjonctive donne les résultats indiqués dans le Tableau 5.

Tableau 4 – Extension vide entre m^{PB} et m^{MB}

$m^{PB \uparrow N_B}$	$\{E_{PB}\} \times \Theta^{MB}$	0.6
	$\{\neg E_{PB}\} \times \Theta^{MB}$	0.3
	$\Theta^{PB} \times \Theta^{MB}$	0.1
$m^{MB \uparrow N_B}$	$\Theta^{PB} \times \{E_{MB}\}$	0.4
	$\Theta^{PB} \times \Theta^{MB}$	0.6

Tableau 5 – Combinaison conjonctive $m_{\odot}^{N_B}$

$m_{\odot}^{N_B}$	$\{E_{PB}\}$	$\{\neg E_{PB}\}$	Θ^{PB}
$\{E_{MB}\}$	0.24	0.12	0.04
Θ^{MB}	0.36	0.18	0.06

Étape 2 En appliquant la fonction τ (Equations 4), nous obtenons : $m_B^{MP}(\{E_{MP}\}) = 0.64, m_B^{MP}(\{\neg E_{MP}\}) = 0.18$ et $m_B^{MP}(\Theta^{MP}) = 0.18$.

Le dernier voisin est le nœud A . L'espace produit : $\Theta^{N_A} = \Theta^{PA} \times \Theta^{MA} = \{(E_{PA}, E_{MA}), (E_{PA}, \neg E_{MA}), (\neg E_{PA}, E_{MA}), (\neg E_{PA}, \neg E_{MA})\}$.

Tableau 6 – Extension vide entre m^{PA} et m^{MA}

$m^{PA \uparrow N_A}$	$\{E_{PA}\} \times \Theta^{MA}$	0.3
	$\{\neg E_{PA}\} \times \Theta^{MA}$	0.45
	$\Theta^{PA} \times \Theta^{MA}$	0.25
$m^{MA \uparrow N_A}$	$\Theta^{PA} \times \{E_{MA}\}$	0.35
	$\Theta^{PA} \times \{\neg E_{MA}\}$	0.25
	$\Theta^{PA} \times \Theta^{MA}$	0.4

Étape 1 L'extension vide de Θ^{PA} et Θ^{MA} à Θ^{N_A} est donnée dans le Tableau 6. La combinaison conjonctive est décrite dans le Tableau 7.

Tableau 7 – Combinaison conjonctive m_{\odot}^{NA}

m_{\odot}^{NA}	$\{E_{PA}\}$	$\{\neg E_{PA}\}$	Θ^{PA}
$\{E_{MA}\}$	0.105	0.1575	0.0875
$\{\neg E_{MA}\}$	0.075	0.1125	0.0625
Θ^{MA}	0.12	0.18	0.1

Étape 2 Le mapping multivalué τ (Equations 4) donne les résultats suivants : $m_A^{MP}(\{E_{MP}\}) = 0.3125$, $m_A^{MP}(\{\neg E_{MP}\}) = 0.355$ et $m_A^{MP}(\Theta^{MP}) = 0.3325$.

Après avoir appliqué les étapes 1 et 2 à tous les voisins communs, les croyances obtenues sont fusionnées à l'Étape 3.

Étape 3 Dans cette étape, les fonctions de masse m_C^{MP} , m_B^{MP} et m_A^{MP} sont combinées en utilisant la règle conjonctive. Ainsi, $m_{\odot}^{MP}(\{E_{MP}\}) = 0.438$, $m_{\odot}^{MP}(\{\neg E_{MP}\}) = 0.087$, $m_{\odot}^{MP}(\Theta^{MP}) = 0.017$ et $m_{\odot}^{MP}(\emptyset) = 0.458$.

Étape 4 Pour prendre une décision sur l'existence d'un lien entre (M, P) , nous calculons la probabilité pignistique (Equation 5) de $\{E_{MP}\}$ et $\{\neg E_{MP}\}$: $BetP^{MP}(E_{MP}) = 0.6755$ et $BetP^{MP}(\neg E_{MP}) = 0.3245$. Ainsi, il y a 67% de chance qu'un lien existe entre P et M . Un lien connectant Peter et Mary est donc inséré dans le graphe $G(V, E^{+1})$ à $t + 1$ (Figure 1(b)).

7 CONCLUSION

Dans ce papier, nous avons étudié le problème de prédiction de liens dans le cadre de la théorie des fonctions de croyance. Nous avons proposé un modèle graphique pour les réseaux sociaux qui encapsule l'incertitude au niveau des liens. Nous avons aussi développé une nouvelle méthode pour la prédiction de liens sociaux en considérant l'incertitude dans le réseau. La théorie des fonctions de croyance est utilisée pour représenter les croyances sur l'existence des liens en les pondérant par des fonctions de masse. Les informations obtenues par les nœuds voisins sont transférées et combinées pour prédire l'existence d'une nouvelle association entre les nœuds non liés. L'extension au cas des nœuds et liens incertains sera considérée

dans les travaux futurs pour étudier l'impact sur la tâche de prédiction.

Références

- [1] A. Popescul, L. H. Ungar. Statistical relational learning for link prediction. In *IJCAI03 Workshop on Learning Statistical Models from Relational Data*, 2003.
- [2] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38 :325–339, 1967.
- [3] C. J. Rhodes, P. Jones. Inferring missing links in partially observed social networks. *JORS*, 60(10) :1373–1383, 2009.
- [4] D. Liben-Nowell, J. Kleinberg. The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Technol.*, 58(7) :1019–1031, 2007.
- [5] D. S. Goldberg, F. P. Roth. Assessing experimentally derived interactions in a small world. In *Proceedings of the National Academy of Sciences*, 100(8) :4372–4376, 2003.
- [6] E. Adar, C. Ré. Managing uncertainty in social networks. *Data Engineering Bulletin*, 30(2) :23–31, 2007.
- [7] F. Johansson, P. Svenson. Constructing and analyzing uncertain social networks from unstructured textual data. In *Mining Social Networks and Security Informatics*, Lecture Notes in Social Networks, pages 41–61. Springer, 2014.
- [8] M. A. Hasan, M. J. Zaki. A survey of link prediction in social networks. In *Social Network Data Analytics*, pages 243–275. Springer, 2011.
- [9] G. Kossinets. Effects of missing data in social networks. *Social Networks*, 28 :247–268, 2003.
- [10] G. Kossinets and D. Watts. Empirical analysis of an evolving social network. *Science*, 311(5757) :88–90, 2006.
- [11] G. Shafer. *A Mathematical Theory of Evidence*. 1976.
- [12] L. Lu, T. Zhou. Link prediction in complex networks : A survey. *Physica A*, 390(6) :1150–1170, 2011.
- [13] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Phys. Rev. E*, 2001.
- [14] P. Smets. The transferable belief model for quantified belief representation. In *Quantified Representation of Uncertainty and Imprecision*, 1 :267–301. Handbook of Defeasible Reasoning and Uncertainty Management Systems, 1988.
- [15] P. Smets, R. Kennes. The transferable belief model. *Artif. Intell.*, 66(2) :191–234, 1994.
- [16] P. Smets. Application of the transferable belief model to diagnostic problems. *International Journal of Intelligent Systems*, 13(2-3) :127–157, 1998.
- [17] S. Ben Dhaou, M. Kharoune, A. Martin, B. Ben Yaghlane. Belief approach for social networks. In *Belief Functions : Theory and Applications*, 8764 :115–123, 2014.