修士論文　要旨

プライバシを保護したリンク解析に関する研究

数理情報学専攻　48096231　森井 正覚
指導教員　中川 裕志 教授

1　はじめに

情報技術の普及に伴い，Web 訪問履歴や購買履歴といった個人情報を扱うサービスが盛んに提供されている．複数のサービス提供者の保持する詳細な個人情報を統合して用いるデータマイニングは，実社会における情報活用に大きく貢献すると期待される．そのような状況下で，データのプライバシを保護しながらデータマイニングを実現する研究 (プライバシ保護データマイニング) が盛んに行われている．多くのデータマイニング手法に対して，データのプライバシを保護する手法が提案されているが，本稿では，リンク解析におけるプライバシ保護について扱う．

リンク解析とは，エンティティとそれらの関係を表すリンクによって表現されたグラフ構造から有用な情報を抽出する手法である．既存のリンク解析は，解析者にエンティティのリンク構造全体が見えているということを前提としている．しかしながら，人間関係や企業取引などの実世界でのリンク情報が公であることは稀である．Sakuma らは，そのような秘密のリンク関係をもつエンティティのグラフにおけるプライバシモデルを定義し，それらのモデルに基づくプライバシを保護したリンク解析を提案した [2]．彼らの提案したリンク解析は自身に関係するリンク情報しか知り得ないモデルにおけるリンク解析である．それゆえ，彼らの研究ではグラフ上のエンティティ単体を一つのパーティと見なしている．我々の研究では，グラフ上の複数のエンティティに関するデータベースをパーティが秘密に保持することを想定する．パーティがあるグラフに関する部分的なリンク情報を保持しており，そのリンク情報を他パーティに対しては知らせたくない状況を考える．我々は，複数のパーティが秘密に保持するグラフを統合したグラフに対してのリンク解析を提案する．

もし秘密のリンク情報を統合したネットワークを対象として，その秘匿性を損なうことなく安全にリンク解析を適用できれば，現実の多様なネットワークからの情報抽出を可能にする．本稿では，その統合したグラフに対する安全なリンク解析アルゴリズムを提案する．

2　リンク解析

ノード集合 $V = \{1, \ldots, n\}$，リンク集合 $E = \{e_{ij}\}$，および重み行列 $W = (w_{ij})$ からなる非負の重み付き有向グラフ $G = (V, E, W)$ を考える．エンティティはノードとして抽象化される．ノード ij 間にリンクが存在しなければ，$w_{ij} = 0$ とする．ノード i の度数は $d_i = \sum_{j \in V} w_{ij}$ と定義される．$D = \mathrm{diag}(d_1, ..., d_n)$ を度数行列と呼ぶ．ノード ij 間にリンクが存在したら (i, j) 成分が 1，そうでなければ 0 であるような行列を隣接行列 $A = (a_{ij})$ とする．

Spectral Ranking:　リンク解析は，与えられたグラフのリンク構造の特徴を考慮して各ノードに何らかのスコアを与えるアルゴリズムである．グラフ上のマルコフランダムウォークにおける定常分布密度によりノードのスコアを計算する方法を spectral ranking と呼ぶ．ノード i からノード j に, 確率 p_{ij} で遷移するマルコフ連鎖を考える．ただし状態遷移確率行列 $P = (p_{ij})$ を $P = D^{-1}W$ として定義する．定常分布 $\boldsymbol{x} = (x_1, \ldots, x_n)^T$ は遷移後もその分布を変えないことから $\boldsymbol{x}^T = \boldsymbol{x}^T P$ を満たす．ただし $\sum_i x_i = 1$ である．この定常分布は，P^T の最大の固有値 $(= 1)$ と対になる固有ベクトル (主固有ベクトル) に対応することが知られている．主固有ベクトルの計算にはべき乗法がしばしば用いられる．初期値として $\sum_i x_i^{(0)} = 1$ なるベクトルを与え，以下の更新式を繰り返す:

$$(\boldsymbol{x}^{(t)})^T \leftarrow (\overline{\boldsymbol{x}}^{(t-1)})^T P, \quad \overline{\boldsymbol{x}}^{(t)} \leftarrow \frac{\boldsymbol{x}^{(t)}}{\| \boldsymbol{x}^{(t)} \|}. \qquad (1)$$

3　問題の定式化

本章では，各パーティが複数のエンティティのリンク情報を秘密に保持している状況で，それらのグラフを統合したグラフにおけるプライバシモデルを定義する．そして，そのモデルに基づいたリンク解析を定義する．各パーティ P_k $(k = 1, \ldots, l)$ が重みつき有向グラフ G^k $(k = 1, \ldots, l)$ をそれぞれ秘密に保持しているとする状況を考える．

定義 1. *(Additive integrated private-weighted graph ; Additive IPWG)* パーティ P_k $(k = 1, \ldots, l)$ はグラフ $G^k = (V, E^k, W^k)$ だけを知っているとする．ただし，$|V| = n$ とする．*additive integrated private-weighted graph* を $G = \left(V, \cup_{k=1}^{l} E^k, \sum_{k=1}^{l} W^k\right)$ と定義する．

続いて，IPWG 上でのリンク解析を定義する．$f : \mathbb{R}^{n \times n} \to \mathbb{R}^n$ をリンク解析のためのスコアリング関数とする．f は重み行列 $W \in \mathbb{R}^{n \times n}$ を入力として，スコアベクトル $\boldsymbol{x} \in \mathbb{R}^n$ を出力する．このとき secure integrated link analysis は以下のように定義される．

定義 2. *(Secure integrated link analysis)* $G = (V, E, W)$ を *IPWG* とする．*Secure integrated link analysis* の実行後，$f(W) \to \boldsymbol{x}$ は正しく評価され，各パーティは $\boldsymbol{x}$ を知るが，それ以外の知識は得ない．

4　暗号学的ツール

本章では，secure integrated spectral ranking (SISR) を実現するために必要ないくつかの暗号学的ツールを導入する．

準同型性公開鍵暗号:　公開鍵暗号系において，暗号化は公にされた公開鍵 pk を，解読にはメッセージの受信者のみが保持する公開鍵に対応した秘密鍵 sk を用いる．平文 m について，$c = \mathrm{Enc}_{\mathsf{pk}}(m; \rho)$ は m の確率暗号による暗号化を，$m = \mathrm{Dec}_{\mathsf{sk}}(c)$ はその解読をあらわす．ρ が $\mathbb{Z}_N (= \{0, 1, \ldots, N-1\})$ 上で一様ランダムに選ばれたならば，暗文 c も同様に $\mathbb{Z}_N$ で一様ランダムに分布する．加法的準同型公開鍵暗号は，秘密鍵の知識なしに，暗文同士の加算

$$\mathrm{Enc}_{\mathsf{pk}}(m_1 + m_2; \rho) = \mathrm{Enc}_{\mathsf{pk}}(m_1; \rho_1) \cdot \mathrm{Enc}_{\mathsf{pk}}(m_2; \rho_2) \quad (2)$$

が可能である．ここで，ρ_1 か ρ_2 の少なくともどちらか 1 つが $\mathbb{Z}_N$ 上で一様ランダムならば，ρ は同様に一様ランダムである．以降は，簡単のために乱数 ρ は表示しない．

ランダムシェア: 行列 $X \in \mathbb{Z}_N^{n_1 \times n_2}$ の各要素 $x_{ij} \in \mathbb{Z}_N$ (for all i,j) が，$\sum_{k=1}^{l} r_{ij}^k \bmod N = x_{ij}$, を満たすように $\mathbb{Z}_N$ から一様ランダムに選択された $r_{ij}^1, \ldots, r_{ij}^l$ に分割されているとする．パーティ $\mathrm{P}_1, \ldots, \mathrm{P}_l$ が X を知らずに $r_{ij}^1, \ldots, r_{ij}^l$ をそれぞれ保持しているとき, これを X のランダムシェアによる秘密共有，と呼ぶ.

Secure Scalar Product Protocol: パーティ P_1 と P_2 がそれぞれベクトル $\boldsymbol{x} = (x_1, \ldots, x_{n_1})^T$ と $\boldsymbol{y} = (y_1, \ldots, y_{n_1})^T$ を保持しているとする．それらの内積のランダムシェアを安全に計算するために，Goethals らによって提案された secure scalar product protocol [1] を用いる．以下では，secure scalar product protocol を $\mathcal{P}_{\mathrm{SSP}}$ と略記し，アルゴリズムの説明に用いる．

Weighted Average Random Share Protocol: パーティ P_k $(k = 1, \ldots, l)$ が自然数 x^k, a^k のペア (x^k, a^k) を保持しているとする．各パーティは互いに協力し，$\sum_{k=1}^{l} x^k / \sum_{k=1}^{l} a^k$ をランダムシェアによって秘密共有する以下の機能を持つプロトコルが必要である．

$$((x^1, a^1), (x^2, a^2), \ldots, (x^l, a^l)) \mapsto (r^1, r^2, \ldots, r^l), \qquad (3)$$

ここで，r^k $(k = 1, \ldots, l)$ は $\sum_{k=1}^{l} r^k = \sum_{k=1}^{l} x^k / \sum_{k=1}^{l} a^k$ を満たすような一様ランダムな数である．

5 Secure Integrated Spectral Ranking

本章では，前章に示した定義に基づく SISR を提案する．SISR の全体的な手順を Procedure 1 に示す.

確率遷移行列のランダムシェア: 問題の定式化で述べたように，リンクとリンク間の重みは公にすべき情報ではない．Spectral ranking には IPWG から計算される確率遷移行列 P が必要であるが，各パーティは P について何も情報を得ない状況で，SISR を実現したい．そのために，P をランダムシェアで秘密共有する．ランダムシェアを計算するために，$\mathcal{P}_{\mathrm{WARS}}$ を用いる．具体的には，パーティ P_k の入力を $(w_{ij}^k, \sum_{j=1}^{n} w_{ij}^k)$ として，$\mathcal{P}_{\mathrm{WARS}}$ を実行し，その出力を P^k の (i,j) 成分とすればよい．もし，$\mathcal{P}_{\mathrm{WARS}}$ の出力が整数でなければ，全パーティが十分大きい自然数をかけることによって整数に拡大すればよい.

べき乗法: IPWG である $G = (V, E, W)$ に対して，$B = LP$ を l パーティでランダムシェアによる秘密共有をする．ここで，$B = \sum_{k=1}^{l} B^k$ であり，パーティ P_k は，B^k のみを知っている．また，パーティ P_k のみが知っているベクトル $\boldsymbol{q}^{(t),k}$ に対して，$\boldsymbol{q}^{(t)} = \sum_{k=1}^{l} \boldsymbol{q}^{(t),k}$ とおく．正規化ステップを省略したべき乗法の更新式は以下のようになる.

$$\begin{aligned} B^T \boldsymbol{q}^{(t)} &= \sum_{k=1}^{l} \left((B^k)^T (\boldsymbol{q}^{(t),1} + \ldots + \boldsymbol{q}^{(t),l}) \right) \\ &= (B^1)^T \boldsymbol{q}^{(t),1} + \ldots + (B^1)^T \boldsymbol{q}^{(t),l} + \ldots \\ &\quad + (B^l)^T \boldsymbol{q}^{(t),1} + \ldots + (B^l)^T \boldsymbol{q}^{(t),l}. \end{aligned} \qquad (4)$$

各パーティ P_k は，$(B^k)^T \boldsymbol{q}^{(t),k}$ を独自に計算できるので，式 (4) の計算は，$(B^k)^T \boldsymbol{q}^{(t),k'} (\forall k \neq k')$ の安全な計算ができればよいことになる．これは，$\mathcal{P}_{\mathrm{SSP}}$ を用いることにより計算が可能である．各パーティは，$(B^k)^T$ の行と $\boldsymbol{q}^{(t),k'}$ の内積をランダムシェアによって秘密共有する．$\mathcal{P}_{\mathrm{SSP}}$ を用いた後，パーティ P_k と $\mathrm{P}_{k'}$ がランダムシェア $\boldsymbol{r}_{k'}^{(t),k}$ と $\boldsymbol{s}_{k}^{(t),k'}$ をそれぞれ保持しているとする．パーティ P_k は以下のようにして $\boldsymbol{q}^{(t),k}$ を更新できる.

$$\boldsymbol{q}^{(t+1),k} \leftarrow (B^k)^T \boldsymbol{q}^{(t),k} + \sum_{k' \neq k} \left(\boldsymbol{r}_{k'}^{(t),k} + \boldsymbol{s}_{k'}^{(t),k} \right), \qquad (5)$$

ここで，$\sum_{k=1}^{l} \boldsymbol{q}^{(t+1),k}$ は，$B^T \boldsymbol{q}^{(t)}$ に等しい．このように，全パーティは式を秘密に更新できる.

Procedure 1 Secure Integrated Spectral Ranking

Require: 公的な入力: $K \in \mathbb{Z}_N, L \in \mathbb{Z}_N$ *s.t.* $Lp_{ij}^k \in \mathbb{Z}_N$ for all i, j, k, IPWG の種類.

Require: パーティ P_k $(k = 1, \ldots, l)$ の秘密の入力: W^k, A^k.

Require: 鍵の設定: すべてのパーティが共同で鍵集合 $\mathcal{K} = \{\mathsf{pk}, \mathsf{sk}^i, \ldots, \mathsf{sk}^j\}$ を生成し，pk はすべてのパーティが所持し，sk^i はパーティ P_i だけが所持するように配布する.

1. (B のランダムシェア) 各パーティは $\mathcal{P}_{\mathrm{WARS}}$ を用いて $B = LP$ のランダムシェアを得る．パーティ P_k はランダムシェア B^k を保持している．ここで，$B = \sum_{k=1}^{l} B^k$ である.

2. (初期化) パーティ P_k はすべての i について，以下のように設定する.
$q_i^{(0),k} \leftarrow K_i^k$ *s.t.* $\sum_{k=1}^{l} \sum_{i=1}^{n} K_i^k = K$, $t \leftarrow 1$

3. (べき乗法) 各パーティ P_k は以下の計算を収束するまで繰り返す.
(a) パーティ P_k は $B^k \boldsymbol{q}^{(t-1),k}$ を計算する.
(b) すべての i と $k' \neq k$ なるすべての k' について，パーティ P_k は $\mathcal{P}_{\mathrm{SSP}}$ を用いてランダムシェア $r_{i,k'}^{(t-1),k}, s_{i,k}^{(t-1),k'}$ を得る．ここで，$r_{i,k'}^{(t-1),k} + s_{i,k}^{(t-1),k'}$ は，B^k の i 番目の行と $\boldsymbol{q}^{(t-1),k'}$ の内積である.
(c) すべての i について，パーティ P_k は $\boldsymbol{q}^{(t),k} \leftarrow B^k \boldsymbol{q}^{(t-1),k} + \sum_{k' \neq k} \left(\boldsymbol{r}_{k'}^{(t-1),k} + \boldsymbol{s}_{k'}^{(t-1),k} \right)$ を計算する.
(d) パーティ P_i と ランダムに選ばれたパーティ P_j $(j \in_r \{1, \ldots, l\} \backslash \{i\})$ は収束を判定するプロトコルを実行する．収束していなければ，"未収束" と全てのパーティに知らせる．もしそのようなメッセージがなければ，ステップ 4 へ進み，そうでなければ，ステップ 3(a) へ戻る.

4. (復号) パーティ P_k は，復号を実行し $\boldsymbol{q}^{(t),k}$ を得，それを全てのパーティに知らせる．それゆえ，出力は $\boldsymbol{\pi}^{(t)} = \sum_{k=1}^{l} \boldsymbol{q}^{(t),k} / KL^{t-1} = \boldsymbol{q}^{(t)} / KL^{t-1}$ となる.

参考文献

[1] B. Goethals, S. Laur, H. Lipmaa, and T. Mielikäinen. On private scalar product computation for privacy-preserving data mining. *Information Security and Cryptology-ICISC 2004*, pp. 104–120, 2005.

[2] J. Sakuma and S. Kobayashi. Link analysis for private weighted graphs. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pp. 235–242. ACM, 2009.