

Zスコアを用いたオブジェクト間関係の効果的可視化法

伏見 卓恭[†] 斉藤 和巳[†] 郷古 浩道^{††}

[†] 静岡県立大学経営情報イノベーション研究科 〒422-8526 静岡県静岡市駿河区谷田 52-1

^{††} 株式会社豊田中央研究所 〒480-1192 愛知県長久手市横道 41-1

E-mail: [†]{j11507,k-saito}@u-shizuoka-ken.ac.jp, ^{††}e1386@mosk.tytlabs.co.jp

あらまし 企業間あるいは業種間の取引関係や産業間の投入量など、オブジェクト間あるいはカテゴリ間の相互関係を効果的に可視化することは、全体構造や法則性を把握するのに有用な方法のひとつと考えられている。一方、重み付きグラフの可視化法として球面可視化法 (SE-PI-W 法) があるが、重みの規模の格差が大きい場合、規模の大きな重みが支配的となり、解釈可能な可視化結果を得ることが困難な場合がある。本稿では、オブジェクト間のインタラクションを数え上げカテゴリに関する Mixing Matrix を構築し、インタラクション数が多項分布に従うとし、ランダム性を仮定したモデルによる期待値および標準偏差により Z スコアを計算する。オブジェクト間の依存度合の重みとして Z スコアを用いた可視化法を提案する。企業間取引ネットワークおよび人工データを用いた実験より、規模の大きさにより隠れてしまう取引関係の本質構造をとらえられる可視化結果が得られることを示す。

キーワード 可視化, 企業間取引ネットワーク, Z スコア, Mixing Matrix

Proposing an Effective Visualization Method for Relationships between Objects using Z-scores

Takayasu FUSHIMI[†], Kazumi SAITO[†], and Hiromichi GOKO^{††}

[†] Graduate School of Management and Information of Innovation, University of Shizuoka

52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan

^{††} TOYOTA CENTRAL R&D LABS., INC.

41-1, Yokomichi, Nagakute, Aichi 480-1192, Japan

E-mail: [†]{j11507,k-saito}@u-shizuoka-ken.ac.jp, ^{††}e1386@mosk.tytlabs.co.jp

Abstract For aiming at uncovering whole structures or rules of mutual connections, it is considered to be one of the useful approaches to visualize effectively relationships between objects or categories such as business relations between enterprises or industries. We treat these multiple or weighted relationships as weighted graphs. On the other hand, it may be difficult to obtain the interpretable visualization results for an existing weighted graphs visualization method (SE-PI-W method) when disparities of scales of multiplicities or weights are large because large-scale weights may become dominant. In this paper, we propose an effective visualization method for these weighted graphs. Our method first builds the Mixing Matrix from the number of interactions between objects, second calculates Z scores by expected values and standard deviations under assuming the multinomial distribution for the values of the Mixing Matrix. Our method using the Z scores as weights between objects and visualize these weighted graphs by the SE-PI-W method. From our experimental evaluations using a inter-firm networks and artificial data, we show that the visualization results which can uncover the essential structures of business connections of hiding with the size of scales.

Key words visualization, inter-firm network, Z score, mixing matrix

1. はじめに

現実社会のあらゆる場面において、企業間あるいは業種間の

取引関係や産業間の投入量などオブジェクトは相互に影響を及ぼし合っている [1]。このようなオブジェクト間あるいはカテゴリ間の相互関係を効果的に可視化することは、オブジェクト群

の全体構造や法則性を把握するのに有用な方法のひとつと考えられる。

オブジェクト群を2部グラフとしてとらえ、2部グラフを効果的に可視化する手法として球面可視化法があげられる[2]。2つのオブジェクト集合を同心円状に配置し、隣接関係の類似するオブジェクト同士を円上の近傍に配置する手法である。また一般的に、オブジェクト間の関係は均一ではなく、関係の強さを表す重み、あるいは、相互関係の量を表す重みが与えられている場合が多い。具体的には、企業間取引ネットワークでは、企業の取引数はスケールフリー性を有することが知られており、取引数には大きな格差があるといえる[1]。このような重み付きのオブジェクト間関係を効果的に可視化する手法として、上記の球面可視化法を拡張したSE-PI-W法があげられる[3]。上記の標準的な球面可視化法と同様に、2つのオブジェクト集合を同心円状に配置し、重み付きの隣接関係の類似するオブジェクト同士を近傍に配置する手法である。この手法では、単純に重み付きの隣接行列を用いるため、重みの規模の格差が大きい場合、規模の大きな重みが支配的となり、解釈可能な可視化結果を得ることが困難な場合がある。

本稿では、上記の問題を回避するために、解釈可能な可視化結果を出力する方法を提案する。提案法は、最初にオブジェクト間のインタラクション数を数え上げ、オブジェクトの属するカテゴリに関するMixing Matrixを構築する。Mixing Matrixは、ネットワークにおいてどのカテゴリに属するオブジェクト間にリンク(インタラクション)が多いかを測るための行列であり、総リンク数で正規化してカテゴリの結合確率分布となる。このインタラクション数が多項分布に従うと仮定し、ランダム性を仮定したモデルによる期待値および標準偏差よりZスコアを計算する。そして、カテゴリ間の重みとしてZスコアを用いることにより、本質的かつ特徴的なカテゴリ間構造を鮮明にした可視化結果を出力する。複数の人工データおよび企業間取引ネットワークを用いた実験より、規模の大きさにより隠れてしまう取引関係の本質構造をとらえられる可視化結果が得られることを示す。

本稿の構成は以下のとおりである。2章で本稿において取り扱う問題を明確にし、3章で提案手法および要素技術であるMixing Matrix, Zスコア, 球面可視化法(SE-PI-W法)について説明する。4章で人工データおよび企業間取引ネットワークを用いて、提案手法および既存手法を比較、評価する。最後に本稿のまとめと今後の展望について述べる。

2. 問題設定

この節では、本稿で取り扱う問題を明確にする。本稿では、階層構造のあるオブジェクト集合に対して、各オブジェクトをそれらが属するカテゴリに統合し、特徴的なカテゴリ間関係を鮮明化することを目的とする。他階層のカテゴリとの関係に焦点を当て、他階層カテゴリとの隣接関係が類似するカテゴリを近傍に配置することで、カテゴリ間の関係を可視化する。

以下に簡単な例を示す。図1のような関係のカテゴリ集合が与えられたとする。この例では、第1階層のカテゴリが第2

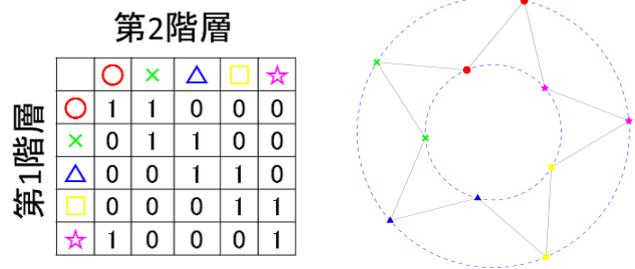


図1 カテゴリ間のインタラクション数の行列 図2 可視化結果

階層のカテゴリとxカテゴリと均等な重みで隣接している。すなわち、第1階層のカテゴリに属するオブジェクトは、第2階層のカテゴリに属するオブジェクトとも、xカテゴリに属するオブジェクトとも、同程度のインタラクションを持っていることを意味する。他のカテゴリも同様に、均等な重みで隣接している。このようなグラフは、図2のように可視化されるのが理想的である。具体的には、第1階層のカテゴリと第2階層のカテゴリ、xカテゴリは原点から見た角度が小さく、近傍に配置されていると言える。また第1階層のカテゴリは、共通隣接カテゴリを有する第1階層のxカテゴリ、カテゴリとも原点からみた角度が小さく近傍に配置される。以上のように、階層間の関係に焦点を当て、各カテゴリ間の重みを考慮したカテゴリの適切な配置場所を求める問題を扱う。

3. 提案手法

この節では、2.節で述べた問題に対する提案手法および、その要素技術について説明する。提案手法は、オブジェクト集合V, オブジェクト間のつながりを表すリンク集合Eが与えられた際、以下の手順によりカテゴリの布置座標ベクトル群Xを決定する。

- (1) リンク集合EからMixing Matrix c_{ij} を構築;
- (2) Mixing Matrixの要素が多項分布に従うと仮定し、Zスコア z_{ij} を計算;
- (3) Zスコアのハイパボリックタンジェント値 w_{ij} を計算;
- (4) (3)で計算した重みを用いて、球面可視化法により布置座標Xを計算;

以下の小節で、各ステップについて説明する。

3.1 Mixing Matrix

Mixing Matrixは、カテゴリに属するオブジェクト間のリンク数を用いて、どのカテゴリに属するオブジェクト間にリンクが多く存在するかの確率を要素とする行列である[4]。表1は、企業をオブジェクト、企業間の取引をリンクとする業種間のMixing Matrixの例である。

この行列 $C = \{c_{ij}\}$ をMixing Matrixと呼び、カテゴリ数がKの場合、 $K \times K$ の行列となる。表1の各要素 c_{ij} は、カテゴリiに属するオブジェクトとカテゴリjに属するオブジェクト間に存在するリンク数の割合である。例えば、“製造業”という業種カテゴリに属する企業オブジェクト同士が繋がっている

表 1 Mixing Matrix の例

c_{ij}	製造業	小売業	サービス業	卸売業	a_i
製造業	0.08	0.1075	0.0075	0.055	0.25
小売業	0.0704	0.0946	0.0066	0.0484	0.22
サービス業	0.16	0.215	0.015	0.11	0.5
卸売業	0.0096	0.0129	0.0009	0.0066	0.03
b_i	0.32	0.43	0.03	0.22	1

割合は 0.08 であり、業種“製造業”の企業と“卸売業”の企業間にあるリンク割合は 0.055 であると読み取れる。この値により、カテゴリ間の癒着や依存度などの関係の強さがわかる。また、行と列それぞれの周辺確率分布を $a_i = \sum_{j=1}^K c_{ij}, b_j = \sum_{i=1}^K c_{ij}$ とする。

3.2 Z スコア

Mixing Matrix の各要素の値 c_{ij} に対して Z スコアを考える。総リンク数 $L = |E|$ としたとき、 Lc_{ij} の期待値を $Le_{ij} = La_i b_j$ で計算する。すなわち、オブジェクト間の関係がランダムに決まると仮定したモデルにおける期待値である。 $K \times K$ の要素が多項分布に従うと仮定し、ランダムなモデルと比較してどの程度有意に存在するかを表す Z スコアを期待値および標準偏差から計算する。

$$z_{ij} = \frac{Lc_{ij} - Le_{ij}}{\sqrt{Le_{ij}(1 - e_{ij})}} \quad (1)$$

Z スコアが正で大きいほど、カテゴリ i のオブジェクトとカテゴリ j のオブジェクト間にリンクが統計的に有意に存在するといえる。すなわちな特徴的なリンク構造となる。逆に Z スコアが負で絶対値が大きいほど、有意に存在しないといえる。Z スコアを用いることにより、出現頻度の少ないリンクであっても、特徴的な関係は大きな値となり、規模の格差により隠れてしまう本質構造を考慮できる。

3.3 ハイパボリックタンジェント関数

Z スコアの値は、ランダムなモデルの期待値と比較してどの程度特徴的に存在するのかを表す。そのため、実データを対象とすると、非常に大きな値を示す場合がある。Z スコア値 z_{ij} の順序関係を維持し大きな値に対処するために、常套手段として用いられるシグモイド関数が考えられるが、Z スコアの値は負の値もとるため、一般的なハイパボリックタンジェント関数を用いる。

$$w_{ij} = \tanh \alpha z_{ij} \quad (2)$$

w_{ij} は、 z_{ij} が大きな負の場合に $w_{ij} \simeq -1$ 、 z_{ij} が大きな正の場合に $w_{ij} \simeq 1$ 、 $z_{ij} \simeq 0$ なら $w_{ij} \simeq 0$ となる。なお、本稿では簡単のため $\alpha = 1$ とする。

3.4 球面可視化法

以下、オブジェクトを統合したカテゴリをノードとし、2 階層 (2 部グラフで表現) の場合の球面可視化法について説明する。第 1 階層のノード集合 V_1 、ノード数 $N_1 = |V_1|$ 、第 2 階層のノード集合 V_2 、ノード数 $N_2 = |V_2|$ とする。2 つの部分集合 V_1 と V_2 は、2 次元ユークリッド空間内の 2 同心円上にそれぞれのノードを配置して表される。部分集合 V_1 に属すノードは

内側の円 (θ_1) 上に描かれ、部分集合 V_2 に属すノードは外側の円 (θ_2) 上に描かれる。 θ_1 は半径 $r_1 = 1$ 、 θ_2 は半径 $r_2 = 2$ とする。この可視化法は、同じようなノードと隣接するノード同士を近くに、異なるノードと隣接するノードを遠くに配置するという方法になっている。

球面可視化法は、重み行列 $\mathbf{W} = \{w_{m,n}\}$ に対して多次元尺度法などで知られるヤング・ハウスホルダー変換 (中心化) を施す [5]。左右より中心化された隣接行列は $\tilde{\mathbf{W}} = \mathbf{H}_{N_1} \mathbf{W} \mathbf{H}_{N_2}$ となる。ここで \mathbf{H}_{N_1} は、 $N_1 \times N_1$ の中心化行列である。球面可視化法では、行列 $\tilde{\mathbf{W}}$ の要素を $\{\tilde{w}_{m,n}\}$ とし、以下の目的関数を最大化するようにノード座標行列 $\mathbf{X} = [x_1, \dots, x_{N_1}]^T$ および $\mathbf{Y} = [y_1, \dots, y_{N_2}]^T$ を適切に初期化した座標行列を基に反復的に求める。

$$J(\mathbf{X}, \mathbf{Y}) = \sum_{m=1}^{N_1} \sum_{n=1}^{N_2} \tilde{w}_{m,n} \frac{x_m^T y_n}{r_1 r_2} + \frac{1}{2} \sum_{m=1}^{N_1} \lambda_m (r_1^2 - x_m^T x_m) + \frac{1}{2} \sum_{n=1}^{N_2} \mu_n (r_2^2 - y_n^T y_n). \quad (3)$$

式 (3) において $\frac{x_m^T y_n}{r_1 r_2} = \cos \theta_{m,n}$ であり、隣接するノード同士がその重みに応じて原点から見て同じ方向に配置されることにより、 $J(\mathbf{X}, \mathbf{Y})$ は最大化される。また、ベクトル群 \mathbf{Y} を固定すれば、ベクトル x_m の最適配置は以下のように求まる。

$$x_m = \frac{r_1}{\|\tilde{x}_m\|} \tilde{x}_m, \quad \tilde{x}_m = \sum_{n=1}^{N_2} \tilde{w}_{m,n} y_n \quad (4)$$

\tilde{x}_m は、ノード m が重み $\tilde{w}_{m,n}$ で隣接するノード n の現在の座標ベクトルの合成ベクトルに他ならない。そして半径 r_1 上にくるように正規化している。よって、対角行列 Λ_{N_1} を用いて式 (4) を行列表記すれば以下となる。

$$\mathbf{X} = \Lambda_{N_1} \tilde{\mathbf{W}} \mathbf{Y} = \Lambda_{N_1} \mathbf{H}_{N_1} \mathbf{W} \mathbf{H}_{N_2} \mathbf{Y}. \quad (5)$$

同様に、ベクトル群 \mathbf{X} を固定すれば、ベクトル y_n の最適配置は以下のように求まる。

$$y_n = \frac{r_2}{\|\tilde{y}_n\|} \tilde{y}_n, \quad \tilde{y}_n = \sum_{m=1}^{N_1} \tilde{w}_{m,n} x_m \quad (6)$$

よって、対角行列 Λ_{N_2} を用いれば以下を得る。

$$\mathbf{Y} = \Lambda_{N_2} \tilde{\mathbf{W}}^T \mathbf{X} = \Lambda_{N_2} \mathbf{H}_{N_2} \mathbf{W}^T \mathbf{H}_{N_1} \mathbf{X}. \quad (7)$$

球面可視化法のアルゴリズムを以下に示す。

- (1) ベクトル群 X と Y を初期化する;
- (2) ベクトル群 Y を固定し、ベクトル X_m を求める;
- (3) ベクトル群 X を固定し、ベクトル Y_n を求める;
- (4) 目的関数 $J(X, Y)$ の変化が十分小さければ終了する;
- (5) (2) へ戻る;

このアルゴリズムは HITS アルゴリズム [6] と類似した構造を

持つことが分かる。ただし、ベクトル群に対して2重の中心化を施す点、および、正規化の施し方の点に特徴を持つ。提案アルゴリズムの1反復は、2部グラフのリンク数に比例した計算量となる。よって、ネットワーク可視化の代表手法の一つパネモデル法[7]などの非線形最適が必要な可視化法と比較して、高速化な方法である。

さらに、 D 階層のデータに適用する場合は、半径をそれぞれ $1 \leq d \leq D$ とした D 個の同心円上にノードを配置する。その場合偶数階層のノードは直近の奇数階層のノードとの関係に基づき、奇数階層のノードは直近の偶数階層のノードとの関係に基づき、配置座標が決定される。

4. 評価実験

提案手法に対して、人工データおよび実データを用いて評価する。評価に際して提案法の優位性を確認するために、重みそのまま使用する単純法、 Z スコアを使用する Z スコア法と比較する。

4.1 使用データ

提案法の妥当性を検証するために、2種の人工データ、および、2種の実データを用いる。

1つ目の人工データは、以下のような重みで隣接するノード群(カテゴリ群)からなるデータである。

表2 ノード間の重み(人工データ1)

第1階層 \ 第2階層	第2階層				
		x			
	10000	1	0.0	0.0	0.0
x	0.0	1	1	0.0	0.0
	0.0	0.0	1	1	0.0
	0.0	0.0	0.0	1	1
	1	0.0	0.0	0.0	1

2つ目の人工データは、以下のような重みで隣接するノード群からなるデータである。

表3 ノード間の重み(人工データ2)

第1階層 \ 第2階層	第2階層				
		x			
	1	10000	0.0	0.0	0.0
x	0.0	1	1	0.0	0.0
	0.0	0.0	1	5000	0.0
	0.0	0.0	0.0	1	1
	1	0.0	0.0	0.0	1

1つ目の実データは、帝国データバンク産業調査部産業分析課に提供いただいた企業間取引関係のデータである。愛知県のある自動車産業の企業を起点として3ステップの取引関係にある企業を集めたものである。第0階層のノードとして愛知県のある自動車産業の企業(シードノード)、第1階層のノードはシードノードと取引関係にある企業群のうち重要業種に属する

企業を選択し、業種ごとに統合したもの、第2階層のノードは第1階層の企業の取引関係にある企業群を業種ごとに統合したもの、第3階層のノードは第2階層の企業と取引関係にある企業群のうち、主要な企業群を業種ごとに統合したものである。それぞれの階層で業種中分類に基づき業種ごとに統合した業種間取引ネットワークである。

2つ目の実データは、1つ目の実データと同様に、愛知県のある自動車産業の企業を起点として3ステップの取引関係にある企業を集めた企業間取引関係のデータである。企業群を業種中分類で業種ごとにまとめた、業種間取引ネットワークである。但し、電気機械器具製造業のみ小分類(電球・照明器具製造(x)、蓄電池製造()、その他())ごとにまとめた業種間取引ネットワークである。

4.2 実験結果

4.2.1 人工データに対する処理結果

人工データ1に対する結果を図4に示す。図4(a)より、重みを単純に用いる方法(単純法)では、重みの大きな赤いノード間が強調され、その他のノード間の関係は全て反対側へ重なってしまうことがわかる。つまり、ノード以外のノード間関係は不鮮明である。重みの規模の格差が大きすぎて、適切に反映できていないことがわかる。図4(b)より、重みの Z スコアを用いる方法(Z スコア法)では、重みの絶対量が大きな赤いノード間は、原点からみて同一方向に(なす角度が比較的小さい位置に)配置されており適切に可視化されていることがわかる。他のノードに関していうと、重みの絶対量が少ないながらも隣接しているノード同士は、原点から見て比較的同一方向に配置されている。さらに提案法を用いた場合でも、 Z スコア法と同様に適切な可視化結果が得られている(図4(c))。

人工データ2に対する結果を図5に示す。図5(a)より単純法では、重みの大きな第1階層の赤いノードと第2階層の緑のxノードとの間、その他のノード間の関係は全て反対側へ重なってしまうことがわかる。重みの規模の格差が大きすぎて、適切に反映できていないことがわかる。図5(b)より、 Z スコア法でも同様に重みの規模の格差が大きすぎて、適切に反映できていないことがわかる。 Z スコアを用いても、 Z スコアの規模の格差が大きすぎるため、単純法と同様の結果が得られたと考えられる。一方、図5(c)より提案法では、上述した重み絶対量が最大のノードペアは、原点から見て同一方向に配置されており、その他のノードペアも同様に、隣接関係が鮮明に表れている。

以上の人工データを用いた評価実験より、提案法は重みの絶対量が大きなペア関係を維持しながら、規模の格差により隠れてしまう本質構造を鮮明化できることが示唆された。

4.2.2 企業間取引データに対する処理結果

企業間取引データの可視化結果における業種ノードの色対応表を図3に示す。

企業間取引データ1に対する結果を図6に示す。図6(a)より単純法では、輸送機械製造や卸売、一般機械器具製造、金属製品製造、電気機械器具製造などの企業数が多い業種ノードの

表 4 Z スコア行列 (企業間取引データ 1)

発注元	発注先	輸送機械製造	卸売 (1)	一般機械器具製造	金属製品製造	電気機械器具製造	鉄・非鉄金属製造	ゴム製品製造	化学工業	繊維工業	皮革・同製品製造	石油石炭製品製造	精密・医療機械製造	その他
輸送機械製造	輸送機械製造	7.04E+1	-3.40E+0	1.60E+1	2.15E+1	-1.19E+1	-2.03E+0	6.02E+0	-3.98E+0	1.30E+0	1.64E+0	-2.06E+0	-8.75E-1	-4.83E-1
卸売 (1)	卸売 (1)	-3.85E+0	1.18E+1	1.58E+1	2.64E+0	2.43E+0	3.11E+1	1.72E+0	1.39E+1	2.37E+1	1.56E+1	7.43E+0	2.85E+0	7.25E+0
一般機械器具製造	一般機械器具製造	4.54E+0	-6.81E+0	6.30E+1	1.89E+1	4.04E+0	6.85E+0	-8.51E-2	-5.55E+0	-2.60E+0	-1.68E+0	-1.26E+0	-7.70E-1	-2.64E+0
金属製品製造	金属製品製造	1.75E+1	-2.00E+0	1.04E+1	2.88E+1	-5.08E+0	-1.10E+0	3.33E+0	-1.61E+0	-1.17E+0	-7.53E-1	-1.30E+0	-6.54E-1	-7.75E-1
電気機械器具製造	電気機械器具製造	-4.18E-1	-1.66E+0	1.53E+1	1.37E+1	6.65E+1	-3.86E+0	9.40E-1	-2.94E+0	-2.06E+0	2.80E+0	-8.08E-1	1.51E+0	-2.44E+0
鉄・非鉄金属製造	鉄・非鉄金属製造	9.37E-1	3.28E+0	2.01E+1	1.24E+1	6.79E+0	2.35E+1	3.51E-1	-2.87E+0	-1.67E+0	-1.57E-1	7.24E+0	-9.39E-1	-6.63E-1
ゴム製品製造	ゴム製品製造	3.62E+0	1.76E+1	1.88E+1	7.51E+0	-5.22E+0	-4.03E+0	5.00E+1	-1.75E+0	2.06E-1	8.57E+0	-1.28E+0	-7.37E-1	-1.04E+0
化学工業	化学工業	2.58E-1	1.16E+1	1.21E+0	2.94E+0	-1.77E+0	-1.57E+0	3.81E+0	2.05E+0	-3.10E-1	-2.00E-1	-5.87E-1	-1.74E-1	2.19E+0
皮革・同製品製造	皮革・同製品製造	-6.25E-1	5.31E-1	-1.23E+0	1.45E-1	-1.02E+0	-8.02E-1	2.56E+0	3.92E+0	-1.03E-1	5.99E+1	-1.96E-1	-5.79E-2	-1.29E-1
石油石炭製品製造	石油石炭製品製造	-1.58E+0	1.99E+1	-1.58E+0	5.10E-1	-3.09E+0	-2.27E+0	-2.55E-1	1.73E+0	-3.41E-1	-2.20E-1	3.19E+1	-1.91E-1	-4.27E-1
精密・医療機械製造	精密・医療機械製造	-5.95E-1	1.70E-1	4.21E+0	1.01E+0	2.81E+0	-1.72E+0	-7.38E-1	-5.16E-1	-2.22E-1	-1.43E-1	-4.20E-1	4.01E+1	-2.77E-1
その他	その他	-2.36E-1	1.22E+1	1.69E+1	2.22E+1	2.19E+0	-6.65E+0	6.80E+0	-1.08E+0	-1.39E+0	-9.00E-1	1.15E+0	-7.81E-1	8.58E+0

表 5 Z スコア行列 (企業間取引データ 2)

第 1 階層	第 2 階層	輸送機械製造	卸売 (1)	一般機械器具製造	金属製品製造	電気機械器具製造 (その他)	電気機械器具製造 (電球・照明器具製造)	電気機械器具製造 (蓄電池製造)	鉄・非鉄金属製造	ゴム製品製造	化学工業
輸送機械製造	輸送機械製造	7.04E+1	-3.40E+0	1.60E+1	2.15E+1	-2.03E+0	6.02E+0	-3.98E+0	1.30E+0	1.64E+0	-2.06E+0
卸売 (1)	卸売 (1)	-3.85E+0	1.18E+1	1.58E+1	2.64E+0	3.11E+1	1.72E+0	1.39E+1	2.37E+1	1.56E+1	7.43E+0
一般機械器具製造	一般機械器具製造	4.54E+0	-6.81E+0	6.30E+1	1.89E+1	6.85E+0	-8.51E-2	-5.55E+0	-2.60E+0	-1.68E+0	-1.26E+0
金属製品製造	金属製品製造	1.75E+1	-2.00E+0	1.04E+1	2.88E+1	-1.10E+0	3.33E+0	-1.61E+0	-1.17E+0	-7.53E-1	-1.30E+0
電気機械器具製造 (その他)	電気機械器具製造 (その他)	9.37E-1	3.28E+0	2.01E+1	1.24E+1	2.35E+1	3.51E-1	-2.87E+0	-1.67E+0	-1.57E-1	7.24E+0
電気機械器具製造 (電球・照明器具製造)	電気機械器具製造 (電球・照明器具製造)	3.62E+0	1.76E+1	1.88E+1	7.51E+0	-4.03E+0	5.00E+1	-1.75E+0	2.06E-1	8.57E+0	-1.28E+0
電気機械器具製造 (蓄電池製造)	電気機械器具製造 (蓄電池製造)	2.58E-1	1.16E+1	1.21E+0	2.94E+0	-1.57E+0	3.81E+0	2.05E+0	-3.10E-1	-2.00E-1	-5.87E-1
鉄・非鉄金属製造	鉄・非鉄金属製造	-6.25E-1	5.31E-1	-1.23E+0	1.45E-1	-8.02E-1	2.56E+0	3.92E+0	-1.03E-1	5.99E+1	-1.96E-1
ゴム製品製造	ゴム製品製造	-1.58E+0	1.99E+1	-1.58E+0	5.10E-1	-2.27E+0	-2.55E-1	1.73E+0	-3.41E-1	-2.20E-1	3.19E+1
化学工業	化学工業	-5.95E-1	1.70E-1	4.21E+0	1.01E+0	-1.72E+0	-7.38E-1	-5.16E-1	-2.22E-1	-1.43E-1	-4.20E-1
繊維工業	繊維工業	-2.36E-1	1.22E+1	1.69E+1	2.22E+1	-6.65E+0	6.80E+0	-1.08E+0	-1.39E+0	-9.00E-1	1.15E+0
皮革・同製品製造	皮革・同製品製造	-3.34E+0	-3.20E+0	1.52E+1	1.20E+1	-3.45E+0	-6.81E-1	-3.15E+0	-1.88E+0	3.19E+0	-2.66E-1
石油石炭製品製造	石油石炭製品製造	9.31E+0	2.81E+0	3.19E+0	5.80E+0	-2.23E+0	4.29E+0	5.39E-1	-7.52E-1	-4.86E-1	-1.42E+0
精密・医療機械製造	精密・医療機械製造	-7.96E-1	3.01E+0	6.56E-1	3.82E+0	3.93E-1	1.64E+0	-9.37E-1	-4.03E-1	-2.60E-1	-7.63E-1

- 輸送機械製造
- 卸売 (1)
- 一般機械器具製造
- 金属製品製造
- 電気機械器具製造
- 鉄・非鉄金属製造
- ゴム製品製造
- 化学工業
- 繊維工業
- 皮革・同製品製造
- 石油石炭製品製造
- 精密・医療機械製造
- その他

図 3 業種中分類の色対応表

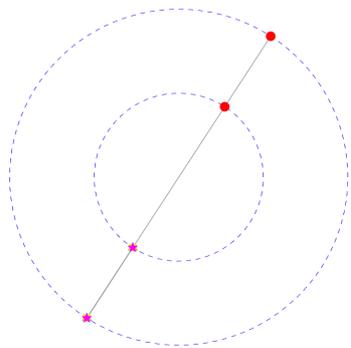
影響により、多くのノードが直線的に配置されている。企業間の取引数はべき乗分布に従うため、取引数の多い企業を含む業種ノード間のリンク重みは非常に大きくなる。逆に取引数の少ない企業を含む業種ノードが関与するリンク重みは非常に小さくなる。そのため、単純法では適切な可視化結果が得られないと考えられる。図 6(b) より Z スコア法では、単純法と比較して幾分かバラつきがあり直線的な配置は幾分か回避されている。さらに、各階層の同一業種ノードが原点から見て同一方向に配置されていることから、同業種間の実取引数が業種ノード比率より計算される期待値より大きいことがわかる。すなわち、統計的有意に同業種間取引が多いことがわかる。表 4 に全体の Mixing Matrix に対する Z スコア値行列をみても同様のことがわかる。図 6(c) より提案法では、Z スコア法よりさらにバラつきがあり、直線的な配置が完全に回避されている。また、図 6(c) の左半分側に輸送機械製造や一般機械器具製造、金属製品製造、電気機械器具製造、鉄・非鉄金属製造など、自動車

産業における主要業種が配置されており、右半分側にゴム製品製造や化学工業、繊維工業といった比較的一般的な業種が配置されており、同階層内における業種間の類似性も反映された可視化結果になっている。

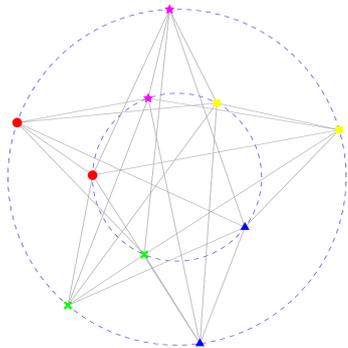
企業間取引データ 2 に対する結果を図 7 に示す。企業間取引データ 2 は、企業間取引データ 1 における電気機械器具製造業のみ小分類にしたデータである。図 7(a) より単純法では、企業間取引データ 1 と同様に企業間の取引数はべき乗分布に従うため、取引数の多い企業を含む業種ノード間のリンク重みは非常に大きくなり直線的な配置となった。単純法では業種間取引の全体構造を鮮明化する適切な可視化結果が得られなかった。図 7(b) より Z スコア法では、単純法と比較して幾分かバラつきがあり直線的な配置は幾分か回避されている。電球・照明器具製造や蓄電池製造の業種は小分類であるため、相対的に業種ノードに含まれる企業数が少なく取引数も少なくなる。表 5 の Z スコアの値を見ると、蓄電池製造などの取引数の少ない業種ノードにおいても特徴的な取引関係にある蓄電池製造同士のリンク重みは大きくなっている。Z スコアを用いることで、統計的有意に多い関係には正の大きな重みが付き、統計的有意に少ない関係には負の大きな重みが付くため、規模の格差に隠れてしまう業種間の取引に関する本質構造が見える可視化結果となった。図 7(c) より提案法では、Z スコア法よりさらにバラつきがあり、直線的な配置が完全に回避されている。また同階層内における業種間の類似性も反映された可視化結果になっている。

5. おわりに

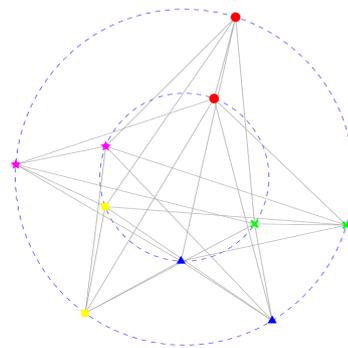
本稿では、企業間あるいは業種間の取引関係や産業間の投入



(a) 単純法

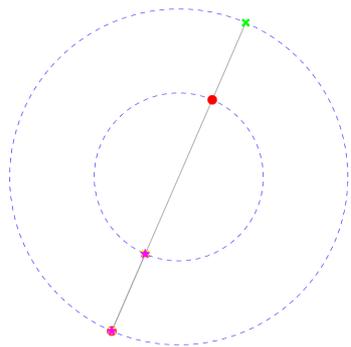


(b) Zスコア法

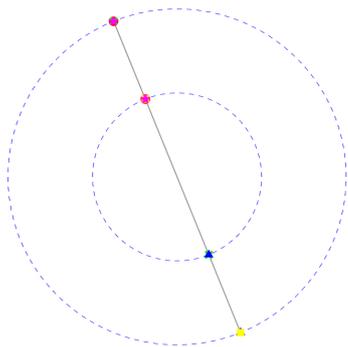


(c) 提案法

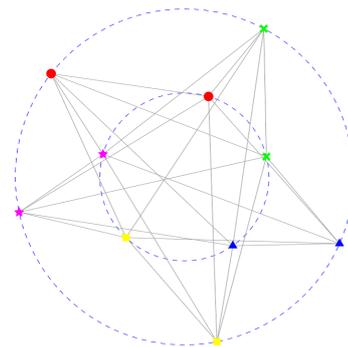
図4 人工データ1



(a) 単純法

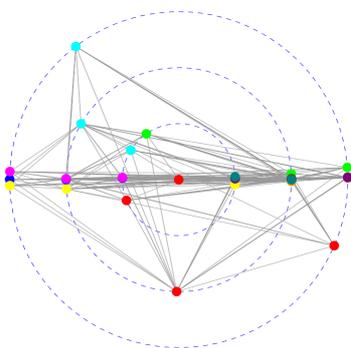


(b) Zスコア法

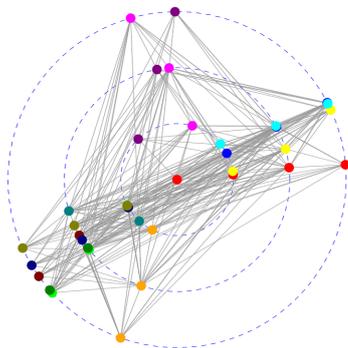


(c) 提案法

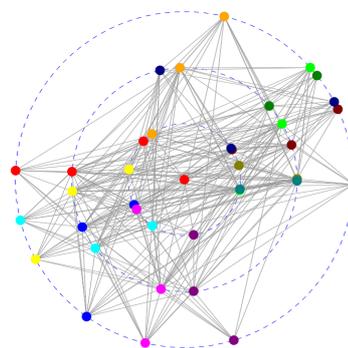
図5 人工データ2



(a) 単純法

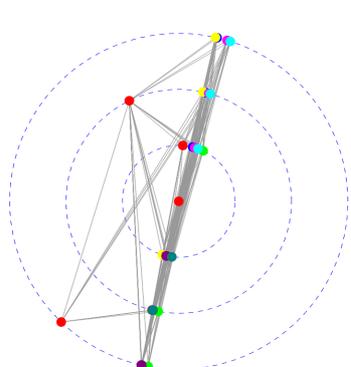


(b) Zスコア法

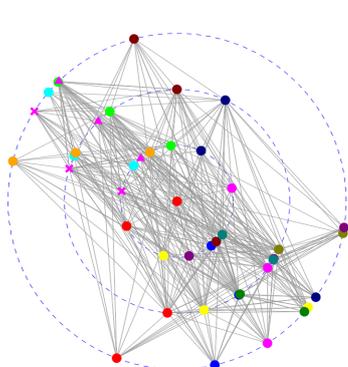


(c) 提案法

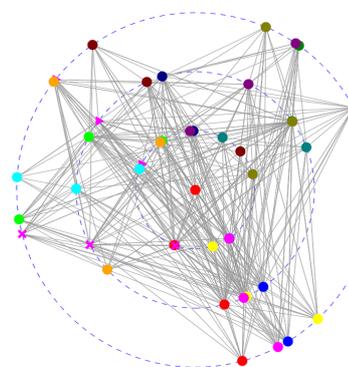
図6 企業間取引データ1(中分類)



(a) 単純法



(b) Zスコア法



(c) 提案法

図7 企業間取引データ2(電気機械器具製造業のみ小分類)

量など、オブジェクト間あるいはカテゴリ間の相互関係を効果的に可視化することにより全体構造や法則性を把握するのに有用な方法を提案した。現実のデータでは取引関係の多重度や関係の強さがべき乗分布に従うなど、規模の格差が大きい場合が多く、規模の大きな重みが支配的となり、解釈可能な可視化結果を得ることが困難な場合がある。企業間取引ネットワークおよび人工データを用いた実験より、重みの Z スコアを用いた提案法は規模の大きさにより隠れてしまう取引関係の本質構造をとらえられる可視化結果が得られることを示す。今後は、より多様なデータを用いて評価し、提案法の有効性・有用性を検証していきたい。また、ハイパボリックタンジェント関数のパラメータを変化させた評価も実行していきたい。

謝辞 本研究は、株式会社豊田中央研究所との共同研究および、科学研究費補助金基盤研究 (C)(No.23500128) の補助を受けた。企業間取引データの作成については、帝国データバンク産業調査部産業分析課の協力をいただいた。

文 献

- [1] 大西立顕, 高安秀樹, 高安美佐子: 企業間ネットワークの数理構造, 応用数理, Vol. 20, No. 3, pp. 223–235 (2010).
- [2] 久保田大和, 伏見卓恭, 斉藤和巳: 2 部グラフ可視化の高速化, 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM2011) (2011).
- [3] 久保田大和, 伏見卓恭, 斉藤和巳, 風間一洋: 重み付き 2 部グラフの可視化法, ネットワークが創発する知能研究会 (Jwein'11) (2011).
- [4] Newman, M. E. J.: Mixing patterns in networks, *Physical Review E*, Vol. 67, No. 2, pp. 026126+ (2003).
- [5] Torgerson, W.: Multidimensional scaling: I. Theory and method, *Psychometrika*, Vol. 17, pp. 401–419 (1952). 10.1007/BF02288916.
- [6] Kleinberg, J. M.: Authoritative sources in a hyperlinked environment, *J. ACM*, Vol. 46, pp. 604–632 (1999).
- [7] Kamada, T. and Kawai, S.: An algorithm for drawing general undirected graphs, *Inf. Process. Lett.*, Vol. 31, pp. 7–15 (1989).