# 転移可能なネットワーク埋め込み

黒川 茂莉† 米川 慧† 村松 茂樹† 小林 亜令†

†株式会社 KDDI 総合研究所 〒102-8460 東京都千代田区飯田橋 3-10-10

E-mail: †{ mo-kurokawa, ke-yonekawa, mura, kobayasi }@ kddi-research.jp

**あらまし** ネットワーク構造で表現されるデータの活用が広がっているが，企業や個人に散在するデータの統合活用は十分ではない. ネットワーク構造で表現されるデータには表や関係といった構造で表現されるデータと比べノード間の関連や関連の連鎖を表現できる特長があり，散在しているデータを統合することができればさらに価値を高めることができる．そこで，本論文では, ノードの対応関係に関する教師情報なしで 2 つのネットワークから得られる特徴を転移学習のフレームワークに組み込み統合する手法を提案する．複数のデータで実験的に評価する．

**キーワード** ネットワーク埋め込み，特徴ベクトル，転移学習，struc2vec

## 1. はじめに

人間関係ネットワークや論文の引用関係ネットワークに代表されるようにネットワーク構造で表現されるデータが溢れている．Web 上や実空間上の行動の遷移もネットワーク構造で表現することができる．ネットワーク構造で表現されるデータはグラフデータと呼ばれることも多いが，本論文ではネットワーク構造のデータと呼ぶ.

データ分析においては，一般的に表または関係といった構造のデータが扱われることが多い．近年，ネットワーク上のノードをネットワーク内の構造的な特徴を保存するように多次元空間に埋め込むネットワーク埋め込みの手法が考案され，ネットワーク構造のデータも表構造のデータに帰着させ，一般的なデータ分析のフレームワークに持ち込むことができるようになっている.

ネットワーク構造のデータはオープンなデータばかりではなく企業や個人の中にも散在しているデータも存在する．これらのデータを統合して分析することができればさらに価値を高めることができる．しかしながら，下記 a，b いずれかの理由から異なるネットワーク構造間でノードの対応付けをすることが現実的には難しく，統合分析は難しいのが現状である.

a) ネットワーク構造間でノードの集合が異なる

b) ネットワーク構造間でノードの集合が同じだが, 対応付けが不可である

a)のケースは，例えば個人毎の移動に関するネットワークで位置がノードとなるような場合に，個人間では移動範囲が異なり，従ってネットワーク構造間でノードの集合が異なる場合が相当する．b)のケースは，例えば異なる情報源の人間関係ネットワークだが，個人情報保護のためネットワーク構造間で直接的なノード同士の対応付けが不可である場合が相当する.

そこで，散在するネットワーク構造データをノードの対応付けをせず構造的な特徴の類似性のみで関連付ける技術へのニーズがあるが, まだ確立されていない. 本論文では，このニーズを踏まえ，ノードの識別子をネットワーク間でマッチングせず，ネットワーク内の構造的な特徴を元に 2 つのネットワークを関連付け，両ネットワークに跨ったネットワーク埋め込みを獲得する課題を解く.

この課題は教師なし転移学習の課題の一種と捉えられる．転移学習とは，元ドメインからの知識を目標ドメインでのタスクに適用し精度を向上させる学習手法である．本論文では 2 つのネットワークを関連付ける課題を考えるため，ドメインとはすなわちネットワークを指す．また，教師なしとは，本論文では，ネットワーク間を関連付ける教師情報がないことを言う.

このような転移学習を実現するためには，ネットワーク間で共通の多次元空間を張り，2 つのネットワーク間で共通の特徴ベクトルを得ることが必要になる．本論文ではネットワーク埋め込みの最新手法である struc2vec[1]で獲得される構造的な特徴の普遍性に着目する．struc2vec では，ネットワーク上離れた位置にあるノードでも近傍構造が似ていれば距離が近くなるように評価する．この点において，struc2vec ではネットワークに非依存な普遍性の高い特徴が得られていると考えられる．従って，教師なし転移学習において有効に機能すると期待する.

本論文では，struc2vec を拡張し，期待される転移学習を実現できるかを実験的に評価する．以降，第 2 章では問題定義について, 第 3 章では関連手法について, 第 4 章では提案手法について，第 5 章では実験について述べ，第 6 章でまとめを述べる.

## 2. 問題定義

図 1 に示すように, 元ドメイン, 目標ドメインの各々のネットワークを考える．元ドメインのネットワークを$G_S = \{V_S, E_S\}$と表し，$V_S = \{v_i^S; i = 1, \dots, n_S\}$は元ドメインのノード集合，$E_S$は元ドメインのエッジ集合を表す.

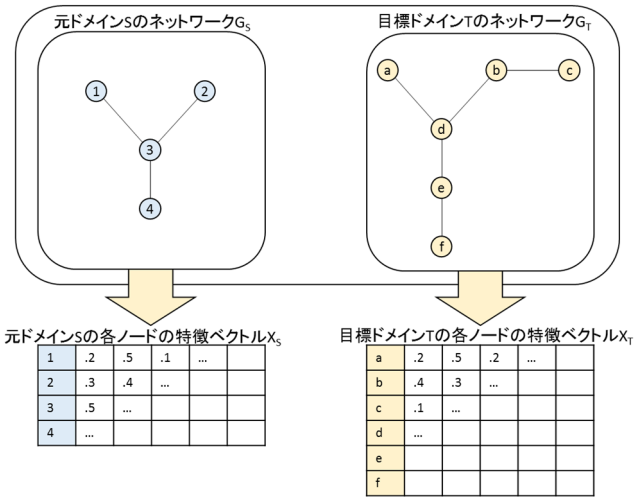



図 1 問題設定

$v_i^S$は元ドメインにおける$i$番目のノードを指す．一方，目標ドメインのネットワークを$G_T = \{V_T, E_T\}$と表し，$V_T = \{v_j^T; j = 1, \ldots, n_T\}$は目標ドメインのノード集合，$E_T$は目標ドメインのエッジ集合を表す．$v_j^T$は元ドメインにおける$j$番目のノードを指す．ここで，ネットワークは無向とする．元ドメインのノード数を$n_S$，目標ドメインのノード数を$n_T$と表す．また，元ドメインのネットワークの直径を$k_S^*$，目標ドメインのネットワークの直径を$k_T^*$と表す．

抽出する特徴ベクトルは目標ドメインの各ノードの特徴ベクトルであり，$X_T = \{\boldsymbol{x}_j^T; j = 1, \ldots, n_T\}$と表す．$\boldsymbol{x}_j^T$は，目標ドメインにおける$j$番目のノードの特徴ベクトルである．特徴抽出の過程で元ドメインの各ノードの特徴ベクトルも抽出し，これを$X_S = \{\boldsymbol{x}_i^S; i = 1, \ldots, n_S\}$と表す．$\boldsymbol{x}_i^S$は，元ドメインにおける$i$番目のノードの特徴ベクトルである．ここで，元ドメインの特徴の次元数を$m_S$，目標ドメインの特徴の次元数を$m_T$と表す．本論文では $m_S = m_T$ とし，特徴空間に対して以下の要件を求める．

**[要件 1]** 両ネットワーク間で構造的な特徴が類似しているノードには，距離が近い特徴ベクトルが割り当てられる．

**[要件 2]** 両ネットワークのスケールの違いに対応でき，スケールの小さいほうのネットワークの構造的な特徴が無視されることがない．

また，以上の転移学習が有効となる前提条件として，最終的な目的（例えば目標ドメインのノードの分類）を考えた場合，目標ドメインのノードに対し元ドメインで構造的な特徴が似ているノードの情報ほど寄与が大きいという条件がある．

## 3. 関連手法

ネットワーク埋め込みと総称される手法の多くは，ネットワークの構造からノードの構造的な特徴を文脈として考慮したシーケンス群を生成し，シーケンス群を元に言語モデル化を行いノードの特徴ベクトルを獲得する．言語モデル化では，ノードのシーケンス群をテキストデータでの文の集合とみなし，それらから文脈を表す潜在変数を獲得する．つまり，ステップとしては，ノードのシーケンス群を生成し，シーケンス群から各ノードの文脈を獲得するという 2 段階のステップになっており，結果として文脈を表す潜在変数の値として各ノードの特徴を得る．本論文では，ノードのシーケンス群を生成するステップを**「シーケンス化」**と呼び，シーケンス群から文脈を獲得するステップを**「言語モデル化」**と呼ぶ．

既存手法の各手法はシーケンス化における文脈の考慮の仕方が異なる．DeepWalk[2], node2vec[3]は各ノードの近傍構造に着目している．struc2vec は近傍構造をさらに抽象化し，多層な中間構造の$k$層において$k$次の近傍構造を表現している．$k$次の近傍構造とは，各ノードから$k$ホップで到達できる範囲の近傍構造を意味する．従って，高階層ほど広域のノードの関係性を表現でき，高階層ほど多くのノードと関係がある影響力の強いノードが強い重みを持つ.シーケンス化の際は，影響力の強いノードほど高階層に遷移しやすく，高階層においては影響力の強いノード同士が遷移しやすいように設計されている．

言語モデル化では 2 層のニューラルネットワークを用いるのが一般的であり，特に Skip-Gram[4]がよく用いられる．Skip-Gram はシーケンス内の単語と文脈を”予測的に関連付け”する．”予測的に関連付け”とは，単語から文脈の予測確率が高くなるように潜在変数を学習するという意味である．

異なるドメインの特徴量の転移を行う技術として転移学習がある．シーケンス化の過程で転移を行う方法は筆者らが調べた限り存在しない．一方，言語モデル化の過程で転移を行う方法として BilBOWA[5]がある．BilBOWA は，word2vec の拡張であり，両ドメインで対応関係のある対コーパスを使ってドメイン間の関連付けを行う．BilBOWA を適用するためには，どのように対コーパスを作成するかが課題となる．

## 4. 提案手法

ネットワーク構造データからデータ内の各ノードの特徴ベクトルを学習する手法である struc2vec を拡張し，2 つのネットワーク間で共通の特徴ベクトルを獲得する転移学習手法を提案する．

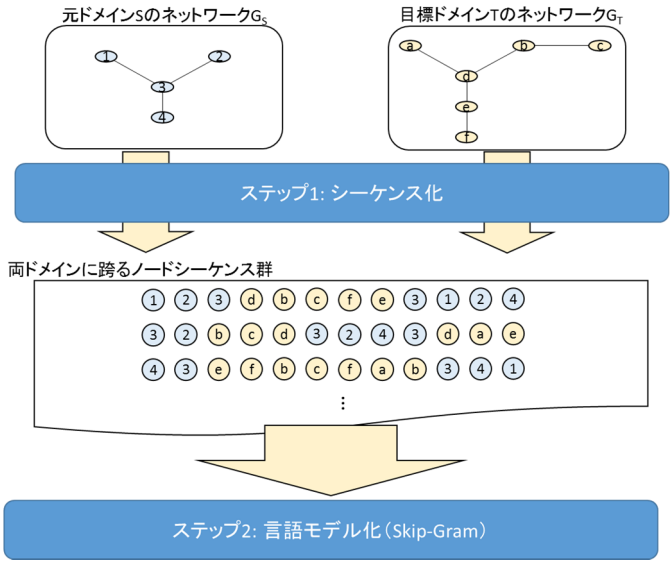



図 2 アプローチ 1 のフロー図

struc2vec のシーケンス化，言語モデル化の各ステップにおいてネットワーク間の関連付けを行うアプローチを提案する．それぞれを 4.1 節，4.2 節に示す．

### 4.1. アプローチ 1: シーケンス化における関連付け

本アプローチでは，図 2 に示す通り，ノードのシーケンス群を生成するシーケンス化のステップにおいてネットワーク間を関連付ける．struc2vec で構築される 2 つのネットワークの中間構造間に仮想的にエッジを張ることで両ネットワークを跨ったシーケンス化を可能とする．ネットワーク間の遷移は確率的に行われるため，ネットワーク間の関連付けはソフトな関連付けとなる．ここで，struc2vec で構成される中間構造の各階層を元ドメインのネットワークについては 1, …, $k_S^*$，目標ドメインのネットワークについては 1, …, $k_T^*$と索引付けする．両ドメイン共，最低階層の索引を 1 とする．

シーケンス化の際，struc2vec では，同階層内でのノード間の遷移，同ノードの階層間での遷移を行い，シーケンスを生成している．これに対し，本アプローチでは，さらにネットワーク間での同階層（同索引の階層）のノード間の遷移を加える．遷移確率の定められたノード間の仮想的なエッジは図 3 の通りである．線が混線するため，ネットワーク間の遷移については，元ドメインのネットワークの 1 ノードからの遷移のみ図示し，それ以外のエッジについては線を省略している．図 3 のような関係に基づいて，ネットワークを跨ったノードのシーケンス群を生成できるようにする．

ここで，階層$k$における元ドメイン，目標ドメインのノード$v_k^S, v_k^T$ 間の遷移確率$p_k(v_k^S, v_k^T)$は，各ノードの当該階層$k$における重要度$w_k(v_k^S), w_k(v_k^T)$を加味し，重要度

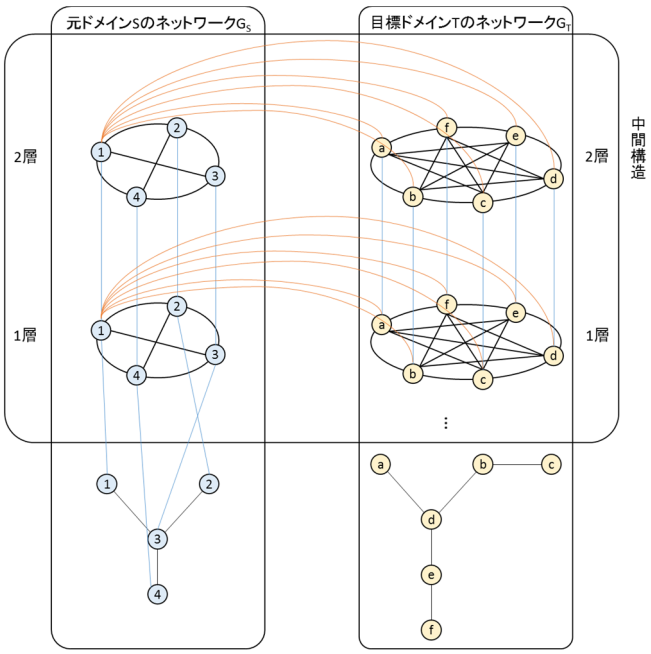



図 3 アプローチ 1 のシーケンス化におけるネットワーク間の関連付け

が近いノードほど遷移しやすいよう以下のように定める．

$$p_k(v_k^S, v_k^T) \propto \mathrm{e}^{-\left|w_k(v_k^T)-w_k(v_k^S)\right|} \quad (1)$$

各ノード$v_k^S, v_k^T$の重要度$w_k(v_k^S), w_k(v_k^T)$は，struc2vec で計算される各ネットワークのノード間のエッジの重みを利用して以下のように計算する．

$$w_k(v_k^S) = \frac{\Gamma_k(v_k^S)}{|E_S|} \quad (2)$$

$$w_k(v_k^T) = \frac{\Gamma_k(v_k^T)}{|E_T|} \quad (3)$$

ここで，$\Gamma_k(v_k^S), \Gamma_k(v_k^T)$ は階層$k$においてエッジに付与された重みの平均より大きな重みを持つ$v_k^S, v_k^T$ からのエッジの本数を表す．この本数を階層$k$におけるエッジの本数$|E_S|, |E_T|$ で割り正規化したものが重要度であり，これは$k$ホップで到達できる範囲における当該ノードの影響力の強さを表している．

struc2vec で計算される各ネットワークのノード間のエッジの重みは各ネットワークのノードの近傍構造を考慮して計算されており，これをネットワーク間でのノード間の遷移確率に反映することで，間接的に各ネットワークの構造的な特徴が反映されることになる．

本アプローチでは 2 章で定めた要件に対応した工夫を行っている．

**[要件 1 への対応]** シーケンスを生成する際のネットワーク間のノード間の遷移確率に各ネットワークの構造的な特徴を反映している．従って，それに基づき生成されるシーケンスを介して構造的な特徴の類似性が最終的な特徴ベクトルに反映される．

**[要件 2 への対応]** スケールが大きいネットワークほど階層が深くなる性質がある．そこで，図 3 に示すように低階層から順にネットワーク間の関連付けを行うことで，ネットワークのスケールの影響を低減している．さらに前述の通り，階層$k$におけるネットワーク間でのノード間の遷移確率の計算においてもネットワークのスケールの影響を低減するように正規化を行っている．

### 4.2. アプローチ 2: 言語モデル化における関連付け

本アプローチでは，図 4 に示す通り，シーケンス化は各ネットワークで独立に行い，生成されたシーケンス群の間で比較を行い類似したシーケンス群を対コーパスとして記憶する．各ドメインで得られたノードシーケンス群を各ドメインのコーパスとし，対コーパスと合わせて言語モデル化における転移学習手法である BilBOWA に入力し，ネットワーク間の関連付けを行う．言語モデル化では対コーパスは固定されるため，ドメイン間の関連付けはハードな関連付けとなる．

対コーパスを作成する際は，シーケンス化の過程で得られる情報を用いてノードの重要度を計算し，利用する．ノードの重要度は式(2)(3)と同様に計算し，最低階層の重要度を用いる．ノードシーケンスをノードの重要度を要素とする重要度シーケンスに変換し，重要度シーケンス間の類似度をユークリッド距離で評価する．対コーパスには類似度が閾値以上のシーケンスを追加する．

本アプローチでは 2 章で定めた要件に対応した工夫を行っている．

**[要件 1 への対応]** 対コーパスをノード間の構造的な特徴の類似性に基づいて作成している．従って，対コーパスを介して構造的な特徴の類似性が最終的な特徴ベクトルに反映される．

**[要件 2 への対応]** 対コーパスを作成する際に考慮する構造的な特徴はネットワークのスケールに依存しない低次の構造的な特徴とすることで，ネットワークのスケールの影響を低減している．

## 5. 実験

実験では，提案手法を struc2vec の公開プログラム[7]を拡張して実装し，評価する．評価では，ベンチマークデータを用いた特徴量の良さの定性評価，ノード分類問題での有効性評価という 2 つの評価を実施する．

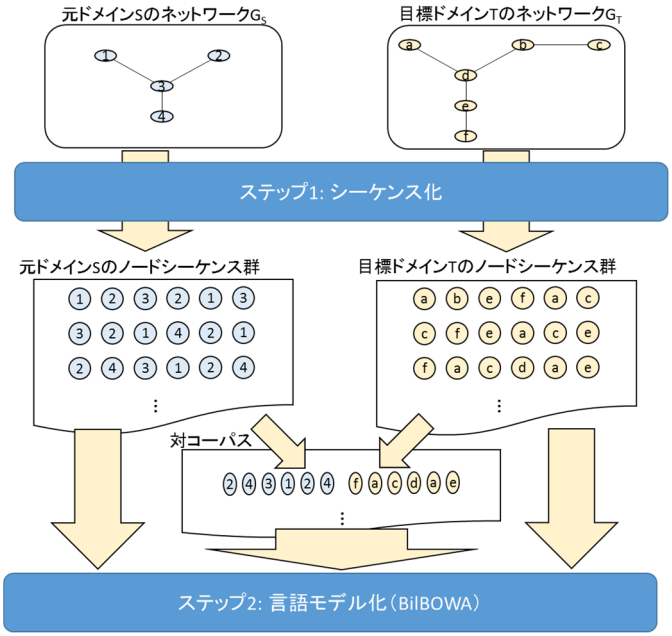



図 4 アプローチ 2 のフロー図

### 5.1. 提案手法の各種パラメータの設定

実験においてパラメータを以下のように設定する．

・ノード当たりのシーケンス数：**20**

・シーケンスの長さ：**80**

以上は，struc2vec[1] におけるミラー化した Karate network[6] への適用の際のパラメータ設定に合わせた．その他のパラメータは以下のように設定する．

・言語モデル化において文脈を考慮する際の窓のサイズ：**10**

窓のサイズは struc2vec[1]が **5** としているのに対し，提案手法では 2 ネットワークに跨る分，2 倍の **10** としている．この意味は，同じ比率で元ドメインのノードと目標ドメインのノードがシーケンスに含まれた場合，文脈を考慮する際に 5:5 で参照されることになり，struc2vec に近しい状況となる．

アプローチ 1 固有のパラメータは以下のように設定する．シーケンス化において次ノードを選択する際，以下の比率で遷移のパターンを選択する．struc2vec の公開プログラムの実装では①:③=0.3:0.7（②は本論文の追加部分のため 0）であり，①は同一ネットワークからの特徴抽出の主要部分であるためそのままとし，③の比率の一部を②ネットワーク間での遷移に回すように設定した．なお，③で同ノードに遷移した場合，struc2vc と同様，同ノードをシーケンスに追加することはしない．

①同階層内でのノード間の遷移：**0.3**

②ネットワーク間での同階層（同索引の階層）のノード間の遷移：**0.2**

③同ノードの階層間での遷移：**0.5**

アプローチ 2 固有のパラメータは以下のように設定する．まず，シーケンス化において次ノードを選択する際，以下の比率で遷移のパターンを選択する．この比率は，struc2vec の公開プログラムの実装に合わせている．

①同階層内でのノード間の遷移：**0.3**

②同ノードの階層間での遷移：**0.7**

さらに，対コーパスに追加するシーケンス数の元ドメインのシーケンスに占める割合を **0.2** とする．

比較手法である struc2vec のパラメータは，窓のサイズを **5** とした以外は提案手法の設定と合わせている．

### 5.2. ベンチマークデータでの評価

ネットワーク埋め込みの評価でベンチマークとして用いられるミラー化した Karate network[6]のエッジを一部切断して評価する．当該ネットワークは 68 個のノードから構成され，構造は図 5 の通りである．これに対してオリジナルの Karate network は，Karate club 内のメンバーをノードとし club 外での交友関係をエッジとして表したネットワークであり，34 個のノードから構成される．ミラー化した Karate network ではこれをミラー化し 2 倍の 68 ノードのネットワークを構築した上で，対応関係のあるノード 1 とノード 37 にエッジを追加している．

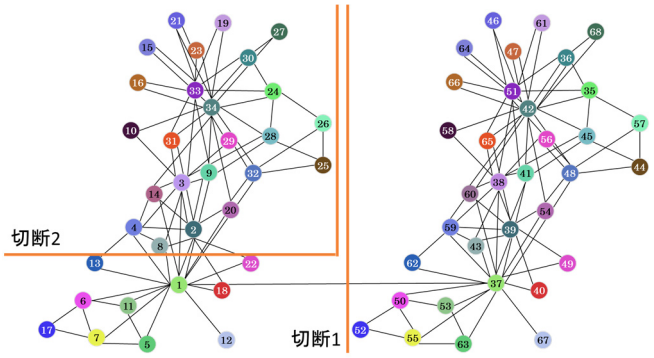
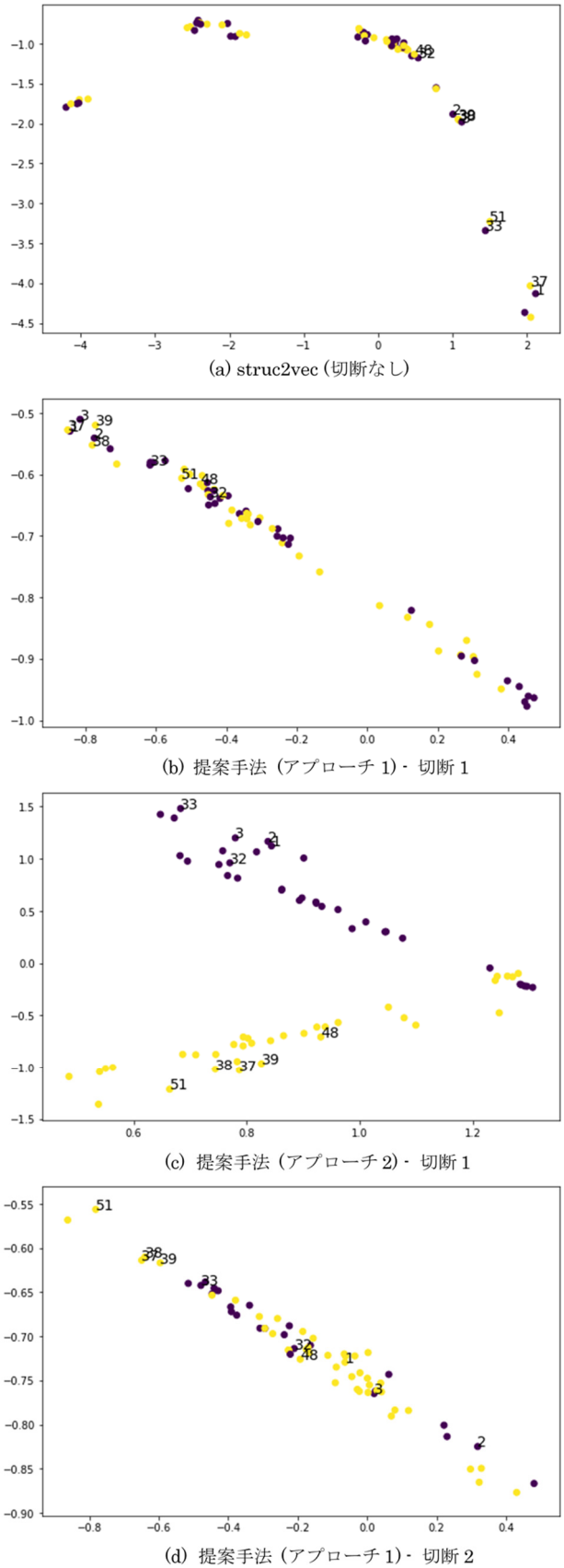



図 5 ミラー化した Karate network

本節の評価では，図 5 で切断 1，2 と示した 2 つの切断パターンで元ドメインと目標ドメインに分割し，これに対し提案手法を適用する．結果として得られた 2 次元の特徴ベクトルについて，構造的な特徴が同一のノードに対して割り当てられる特徴ベクトルの距離が近いかどうかを評価した．切断 1 の場合，線の左側を元ドメイン，右側を目標ドメインとし，切断 2 の場合，線の左上側を元ドメイン，右下側を目標ドメインとする．元ドメインと目標ドメインのネットワークのスケールについて，切断 1 はバランスがとれているのに対し，切断 2 はバランスがとれておらず，より難易度の高い設定となっている．



(a) struc2vec (切断なし)

(b) 提案手法 (アプローチ 1) - 切断 1

(c) 提案手法 (アプローチ 2) - 切断 1

(d) 提案手法 (アプローチ 1) - 切断 2

図 6 ミラー化した Karate network への適用結果

結果は図 6 の通りであった．図 6 の紫のノードは元ドメインのノード，黄のノードは目標ドメインのノードを示す．また，図 6 で番号を併記しているノードは対応関係のあるノードであり，さらに当該ネットワークで中心的な構造的な特徴を持つノードである．それぞれ 1-37，2-39，3-38，33-51，32-48 という対応関係がある．また，2-39，3-38 は構造的な特徴が近いことが分かっており，これらは近い距離に布置されることが望ましい．

図 6 より，以下の結果が読み取れた．

・切断なしの理想的な環境における struc2vec の適用結果では，対応関係のあるノードはほぼ同じ位置に布置されている．また，2-39，3-38 も近い位置に布置されている．それに対し，切断ありの状況に適用した提案手法の場合は，理想環境下での struc2vec の適用結果と比べると，離れた位置に布置されている．

・切断ありでドメイン間のバランスがとれた環境（切断 1）においてアプローチ 1 とアプローチ 2 を比較すると，アプローチ 1 のほうが対応関係のあるノード同士の距離は近い．アプローチ 1 では 2-39，3-38 も近い位置に布置されている．これは，対コーパス作成においてノードシーケンスの関係を固定してしまうアプローチ 2 より，ノードシーケンスを生成する際に都度確率的に関連するノードを選択するアプローチ 1 のほうが表現力において優位である可能性を示している．

・アプローチ 1 について切断ありでドメイン間のバランスがとれた環境（切断 1）とドメイン間のバランスがとれていない環境（切断 2）を比較すると，切断 2 のほうが対応関係のあるノード同士の距離が離れている．個別に見ると 32-48 については距離が近いが，それ以外は離れている．このことから，ドメイン間のバランスがとれていない環境では，ネットワーク間のスケールが異なり，中間構造の階層の深さも異なるため，ネットワーク間での遷移が上手く行えていない可能性を示している．この点において，アプローチ 1 も改善の余地がある．

### 5.3. ノード分類問題での有効性評価

ネットワーク上のノードにラベルが付与されており，ノードの特徴ベクトルからそのラベルを予測する問題を考える．この問題において，提案手法により教師なしで獲得した特徴ベクトルが，分類精度にどれだけ寄与するかを評価する．

評価用のデータとしては，struc2vec の評価で用いられている空港ネットワーク（ブラジル；**B**, ヨーロッパ；**E**）を用いる．空港ネットワークの各ノードに対し利用率に基づく 4 区分のラベルが付与されている．各ネットワークのノード数は **B** が 131，**E** は 399 であり，エッジ数は **B** が 1,074，**E** は 5,995 である．

提案手法は **B**，**E** 両方のネットワークを参照し，エリア間の転移（**B→E**）を意図した転移学習を行う．まず **B**，**E** 両方のネットワークに提案手法（アプローチ 1）を適用し，各ノードについて 2 次元の特徴ベクトルを作成する．さらに，各ノードのラベルを参照し，分類器を学習する．一方，比較対象である struc2vec は **E** のネットワークのみを参照し，2 次元の特徴ベクトル作成と分類器学習を行う．分類器は L2 正則化付きロジスティック回帰とし，正則化項の重みは学習データ内での交差検証により設定した．4-fold 交差検証により評価を行い，各 fold では，検証用のラベルとして目標ドメイン（**E**）からランダムに 25% のラベルを選択し，学習時には参照しない．

結果の confusion matrix は表 1 の通りであった．精度に関してはほぼ同等（201/399 と 200/399）であるが，ラベル=3 に関しては提案手法のほうが優位である．転移学習における分類問題に関しては，元ドメインから問題に十分に有用な知識を取り出せない場合は負の転移が起こる場合があることが知られており，この結果からは正の転移と負の転移が混在した状況が起こっていることが読み取れる．この例では，利用率の最高（ラベル=3）と最低（ラベル=0）に関する分類に関しては正の転移ができているが，利用率が中程度のラベル=1 または 2 の分類に関しては正の転移ができていないか，負の転移が発生している状況が想定される．

表 1 ノード分類の confusion matrix

| | | **E** に struc2vec 適用 | | | | **B**,**E** に提案手法適用 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 |
| 正解 | 0 | 96 | 3 | 0 | 0 | 94 | 3 | 0 | 2 |
| | 1 | 49 | 21 | 13 | 16 | 54 | 11 | 4 | 30 |
| | 2 | 11 | 12 | 31 | 45 | 11 | 6 | 8 | 74 |
| | 3 | 8 | 9 | 32 | 53 | 1 | 6 | 8 | 87 |

## 6. おわりに

本論文では，企業や個人に散在するネットワーク構造データの統合活用を目指し，ノードの対応関係に関する教師情報なしで 2 つのネットワークから得られる特徴を転移学習のフレームワークに組み込み統合する手法を提案し，初期的な評価を行った．今後はさらに詳細な実験を行い，更なる精度向上を目指す．

## 参 考 文 献

[1] Ribeiro, L. F., Saverese, P. H., & Figueiredo, D. R. (2017, August). struc2vec: Learning node representations from structural identity. In

Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 385-394).

[2] Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 701-710).

[3] Grover, A., & Leskovec, J. (2016, August). node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 855-864).

[4] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).

[5] Gouws, S., Bengio, Y., & Corrado, G. (2015). Bilbowa: Fast bilingual distributed representations without word alignments. In Proceedings of the 32nd International Conference on Machine Learning (ICML-15) (pp. 748-756).

[6] Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. Journal of anthropological research, 33(4), 452-473.

[7] https://github.com/leoribeiro/struc2vec