

# 実ネットワークにおける役割情報のマルチラベル分類

## Multilabel Classification of Role Information in Real Network

劉庶<sup>\*1</sup> 鳥海不二夫<sup>\*1</sup>  
Shu Liu Fujio Toriumi

<sup>\*1</sup> 東京大学大学院  
The University of Tokyo

複雑ネットワークにおいて、ノードの構造的特徴や機能を集約したものを役割という。実ネットワークでは、1つのノードが複数の役割（多重役割）を有することが観測される。本研究では、ドメイン敵対的学習を用いた単一役割発見フレームワークを拡張し、ノードが持つ多重役割を発見できる手法を提案する。更に、実ネットワークの多重役割発見の大きな課題である事前ラベリング対象の選択に対して、コミュニティ抽出で分割したサブネットワークをソースネットワークおよびバリデーションネットワークに割り当てる手法を提案する。実ネットワークを用いた精度評価実験で、提案手法が上回った精度や安定した結果を達成したことを示す。

キーワード：複雑ネットワーク、役割発見、マルチラベル分類、コミュニティ分割、敵対的学習

In complex networks, role is the aggregation of structural features and functions of nodes. In real networks, it is observed that one node has a plurality of roles (multiplex roles). Therefore, in this study, we propose a method that extends a single role discovery framework using domain adversarial learning and discovers multiple roles for nodes. Furthermore, we propose a method to assign sub-networks divided by community extraction methods to a source network and a validation network, for the selection of datasets for pre-labeling, which is a major task of multiple role discovery in real networks. In an accuracy evaluation experiment using a real network, it indicates that the accuracy and stable result exceeded by the proposed method have achieved.

Keywords: Complex Network, Role Discovery, Multi-label Classification, Community Detection, Adversarial learning

## 1. はじめに

ネットワークは要素（ノード、節点）集合と要素間の相互関係（エッジ、枝）集合からなるデータ構造であり、要素間の関係性を直観的に表現できる利点がある。1998年に実世界に観測されたネットワークに近い性質を持つネットワークを簡単に生成できるモデル（ワッツ・ストログッツモデル）が提案されたことをきっかけに、実世界に存在する様々な巨大且つ複雑なネットワークを研究する複雑ネットワーク科学が注目されてきた。複雑ネットワークは、生物学や社会科学等に幅広く応用され、大規模で複雑なシステムのメカニズムや隠れた高次元知見を解明することができるといわれる。

複雑ネットワークのノードが持つ、ネットワークにおける構造的特徴や機能を役割と定義する。ネットワークトポロジーの文脈では、ノードの役割として、スター中心（star-center）、スター周辺（star-periphery）、クリークメンバー（clique-member）等が挙げられる一方、3ノードモチーフの文脈では、ソース（source）やシンク（sink）が考えられる。人間がわかりづらい複雑なネットワーク構造をより直観的に理解するために、ノードやエッジに役割（Role）を割り当てる、役割発見（Role Discovery）はネットワーク科学における研究分野の一つである [1]。役割発見は、グラフマイニングと探索的分析にとって重要であるだけでなく、多くの実用的なアプリケーションでも有用である。たとえば、IP トレース [2, 3] などの技術ネットワークの異常検出において、正常の役割（通常の構造パターン）に適合しないノードを異常という役割を割り当てて抽出することが考えられる [4]。また、オンラインソーシャルネットワークにおいて、「スター中心」の役割を有するユーザはインフルエンサーと見なすことができ、広告配信の効果が大きいと考えられる。

実世界のネットワークでは、構築方法によっては1つのノードが多重の役割を同時に持つことも考えられる。また、時間軸で変化するテンポラルネットワークでは、ノードの役割は時間に沿って変化することがあり、分割の粒度によっては、1つのタイムスライス内で多重の役割が観測される [1]。例えば、人をノード、人と人の繋がりをリンクで構築した人間関係ネットワークにおいて、平日の勤務時間では会社員で、それ以外の時間では父親の役割を果たすノードは、長いタイムスパンから見ると、会社員と父親の役割を担うノードと言える。また、語彙の共起ネットワークでは、語彙の品詞を役割ラベルとした場合、複数の品詞を持つ語彙が多重の役割を有することになる。テンポラルネットワークのみならず、ノードの多重役割現象は一般に存在する。本質的に、ノードの多重役割現象はネットワークのローカル構造により生じたものと考えれば、異なるローカル構造に異なる役割を担うため、グローバルスケールでは多重役割現象が観測される。

既存研究では多重役割現象について認識しているものの、多重役割の推定に具体的なアプローチは示されていない [1]。本研究では複雑ネットワークにおける役割発見にマルチラベル分類のアプローチを適用し、高精度に多重役割を予測することを目指し、以下2つを提案する。1つ目は、単一役割発見フレームワークである転移学習を用いたノードの役割発見 [5] を多重役割発見ができるように改良し提案することである。2つ目は、提案手法を実世界ネットワークに適用する際、適切なソースネットワーク、バリデーションネットワークの選定方法を提案することである。本論では、実世界ネットワーク（Wikipedia ネットワーク [6]）を用いて精度検証を行う。

## 2. 関連研究

### 2.1 役割発見

役割発見はノードをある種の同等性を持つノード群、つまり役割クラス、に分割するタスクである。同等性には、主にネットワーク構造から定義された構造的同等性（構造的同一、自己同形等価、レギュラー等価、確率的等価）と、ノードの特微量から定めた特徴的同等性がある。役割はそれぞれの同等性の定義に拠り、グラフベースの役割と特微量ベースの役割に分けられる。役割発見の従来手法では主に教師なし学習と教師あり学習のアプローチがある。教師なし学習アプローチはネットワーク構造からノードのベクトル表現を取得しクラスタリングしてから、各クラスに役割ラベルを割り当てる流れである。最も有名な手法はブロックモデルである [7, 8, 9, 10]。ブロックモデルは、同じ役割を持つノードを集約しブロック（ノード）で表し、ブロック間のエッジが役割間の相互作用を示す役割相互作用グラフを構築し、整形された目的関数を使用した最適化問題として定式化されることが一般的である。各ノードをいくつかの役割（またはブロック）の1つに割り当てるモデルもあれば、単一の役割のみに属するノードの仮定を緩和する混合メンバーシップ確率ブロックモデル [11] も提案された。Airoldiら [11] が提案したモデルは変分推論アルゴリズムで拡張した確率的ブロックモデルであり、明確にノードの多重役割を答えるのではなく、1ノードに複数の役割に所属する確率ベクトル（混合メンバーシップと言い、ベクトル要素の合計が1になる）を割り当てるソフトアサインメント方式が特徴である。また、隣接行列の行の間に対して何らかの形の類似性を用いて役割を計算する手法もある [12, 13]。これらの手法では主に、最初に隣接行列の行の各ペア間の類似性（または距離）を計算してから、出来上がった類似性マトリックスを使用してノードをクラスタリングする流れになる。また、隣接行列（または類似性行列）の固有ベクトルのサブセットを使用して役割を導出するスペクトル法もある [13]。

教師なし学習の手法は、個々のクラスに具体的な役割を割り当てる際に、主観的に推測する必要があり、役割の解釈性が低いことや割り当てた役割の妥当性を確認することが困難であるという欠点がある。

それに対して、教師あり学習アプローチは事前に定めた役割、および役割の割り当て基準や正例を通じて学習モデルを訓練し、テストデータの役割を予測するやり方になる。役割が必要に応じて定められるため、解釈性が高まることや、正例から学習することで予測する結果の精度を一定水準に保てるのが期待できる。本研究の先行研究である転移学習を用いたノードの役割発見 [5] は教師あり学習アプローチの一例であり、事前にラベリング済みのネットワーク（ソースネットワーク）の役割知識を役割が未知であるネットワーク（ターゲットネットワーク）に転移することを実現した。また、先行研究では、モデルの精度を検証するネットワーク（バリデーションネットワーク）を用いて、グリッドサーチによる最適なハイパーパラメータを選出する。

いずれのアプローチでも、ノードの単一役割発見を目的とし、多重役割発見はできない。多重役割の発見に近いアプローチは、役割を割り当てる際に、1ノードに1役割ラベルを割り当てる代わりに、1ノードに複数の役割に所属する確率ベクトル（ベクトル要素の合計が1になる）を割り当てる方式、つまり、ソフト割り当て方式を採用するアプローチになる。これは、ノードの多重役割の大まかなトレンドしか得られないため、多重役割発見に不十分である。

### 2.2 マルチラベル分類

マルチラベル分類タスクは、教師あり学習の一種であり、入力データに対応する出力データが二値（0と1）を複数並べたベクトルになる。ベクトルの要素数が対象となるラベルの数に対応し、0の場合が該当ラベルがネガティブで、1の場合が該当ラベルがポジティブであることを示す。例えば、画像物体検出の時、1枚の画像の中に複数の物体が存在し、同時に検出する時はマルチラベル分類を適用できる [14]。出力データが0または1の二値分類タスクや、出力データが1実数である多値分類タスクに比べると、マルチラベル分類は高次元性やラベル不均衡等の問題の影響を受けやすく、難易度が高いタスクである [15]。

入力データ空間を  $\mathcal{X}$ 、出力データ空間を  $\mathcal{Y}$  とした場合、マルチラベル分類器  $\mathcal{F}$  を  $\mathcal{F}: \mathcal{X} \rightarrow \mathcal{Y}$  に定義できる。入力は任意の  $X \in \mathcal{X}$  で、出力  $Z \in \mathcal{Y}$  が  $Z = \mathcal{F}(X)$  になる。

マルチラベル分類は、分類器のアンサンブル、データ変換、メソッド適応等、さまざまなアプローチを通じて実現できる。

BR (Binary Relevance) は分類器のアンサンブルに該当する手法であり、二値分類器をラベル毎に訓練した後、テストデータに対するそれらの二値分類器の出力を組み合わせることで予測ラベル集合を生成する。二値分類器は、任意の二値分類アルゴリズムを使用できる。個々のラベルが独立して並列的に訓練できる利点がある一方、ラベル間の相関関係を考慮できない欠点がある。そして、予測結果として、全てのラベルが全部ネガティブになってしまうこともあり得る。

LP (Label Powerset) はデータ変換のアプローチに該当する手法であり、マルチラベル分類を多値分類に向けてデータ変換を行う。具体的に、学習データに現れた各ラベルの有無を示す1と0を要素としたラベルベクトルを一意に定めた値（例：ラベルの組み合わせパターン番号）に置き換えて、多値分類器を該当値を予測できるよう訓練した後、テストデータに対応する予測結果（該当値）を元のラベルの組み合わせパターンに復元する。LPはBRと違って、ラベル間の相関関係を保存したままで学習することがメリットである。そして、多値分類としてタスクを解決するため、多値分類のノウハウをそのまま適用できることもメリットの1つになる。但し、元のラベルの組み合わせは論理上、 $2^k$  個もあり、ラベルが多い場合には  $2^k \gg n$  も考えられるため、極端の場合は多値分類問題に成り立たないことも考えられる。そして、学習データで現れたことがないラベルの組み合わせを検出できないため、学習データセットとテストデータセットが大きく偏った場合には、精度が低くなる懸念もある。

MLkNN は、k近傍法をマルチラベル分類に適応したメソッド適応の手法である [16]。k個の近傍インスタンスのラベルベクトルにより、該当インスタンスのラベルベクトルを予測する手法である。k近傍法と違って、各ラベルの事前確率と条件付き確率を含んだモデルを生成することが特徴である。

### 2.3 実ネットワークの役割

ノードの役割は、ネットワークトポロジーの文脈では、スター中心 (star-center)、スター周辺 (star-periphery)、クリークメンバー (clique-member) 等が挙げられる [1] 一方、有向ネットワークのモチーフの文脈では、ソース (source) やシンク (sink) が考える。更に実ネットワークにおいては、情報マイニングの観点から、より実用化できる役割種類を定義する必要があると考える。例えば、単語をノードに、単語の共起現象をエッジにした語彙ネットワークでは、単語の構造的特徴を文法的な機能の集約と見なせる。単語の文法的な基準で定義された概念の1つは品詞である [17] ため、単語の品詞を語彙ネッ

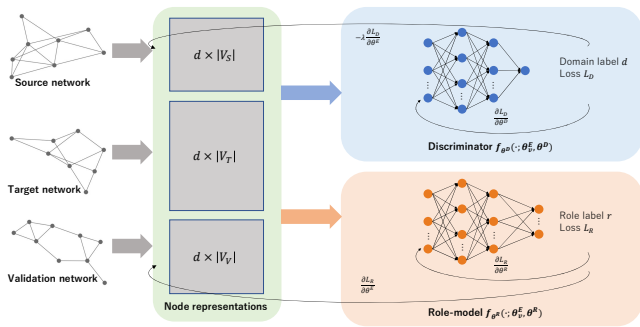


図 1: 転移学習を用いた単一役割発見手法 [5] の概略図

トワークにおける単語の役割にすることができる。ちなみに、1 単語が 2 つ以上の品詞ラベルを持つことがあるので、単語の品詞ラベルの予測は多重役割発見のタスクである。

あるターゲットネットワークが与えられた時、役割をどう発見できるかを考える。率直なアプローチとして、テストデータの特徴量からラベルを予測する問題として考える場合、テストデータの中からランダムサンプリングしたデータインスタンスにラベリングを行い、ラベル付きデータを教師データで識別器を訓練し、他のテストデータのラベルを予測する形になる。しかし、複雑ネットワークの役割の場合、以下の問題点が考えられる。

1. 複雑ネットワークのスケールフリー性で、少ないノードが高い次数を持つ現象があり、ランダムサンプリングでこういった特徴を持つノードが取れない可能性が高い。これは学習データの役割の欠損に繋がる。
2. 役割のラベリングは周辺ノードの役割によるものかしばしばあるため、ランダムサンプリングしたノードのみではなく、複数ホップ先までの周辺ノードも確認する必要がある。複雑ネットワークのsmall-world性から、ネットワーク全域に渡る考察が必要になる可能性があり、結局全ネットワークの内、大半のノードの役割は学習データの構築のためにラベリングすることになり、ランダムサンプリングの意味がなくなる。

このように、実ネットワークにおける役割発見は、役割自体が多様であり、ネットワーク種類に応じて定める必要があることや、複雑ネットワークのスケールフリー性やsmall-world性により、学習データセットの構築が課題である。

### 3. 先行研究 (転移学習を用いたノードの役割発見 [5])

本研究の先行研究である転移学習を用いたノードの役割発見 [5] (概略図は図 1 に示す) のポイントを述べる。

#### 3.1 分散表現の取得

先行研究では、struc2vec[18] を用いて、ノードの構造的特徴を示す分散表現 (ベクトル形式) を取得する。図 1 の左側にある灰色の矢印は struc2vec になる。struc2vec は構造的に類似するノードを潜在空間にも近く配置するように埋め込み表現を取得する手法であり、ノードがネットワークにおける位置情報に依存しないため、ノードの役割発見に相応しいと考える。詳細については元の論文 [18] を参照。

### 3.2 ドメイン敵対的学習

struc2vec で取得したノードの分散表現はソースネットワークとターゲットネットワークのスケール等の違いが原因で、両ネットワークのドメインが異なり、ターゲットネットワークの役割推定のパフォーマンス低下に繋がる可能性がある。そのため、先行研究では、両ネットワークのドメインの違いを緩和するために、ドメイン敵対的学習フレームワークを使用する。ドメイン敵対的学習フレームワークには、図 1 の右上に示す Discriminator と、右下にある Role-model がある。Discriminator の目的は、ノードがソースネットワークに属するかターゲットネットワークに属するかを判断することである。Discriminator の入力 is ノードの分散表現であり、0 から 1 の間の信頼スコアを出力とする。ソースネットワークに属すると判断した場合は 0 に、ターゲットネットワークに属すると判断した場合は 1 に近づく。Role-model の目標は、ノードの単一役割ラベルを予測するとともに、両ネットワークのノード分散表現を類似させることで、gradient reversal layer[19] より、Discriminator を騙すことである。ドメイン敵対的学習フレームワークに従い、mini-max ゲームを行い、学習が終了した後、ターゲットネットワークのノード分散表現を Role-model の入力とし、対応する単一役割ラベルを予測する。

### 3.3 バリデーションネットワークを用いたモデル選択

先行研究では、通常の教師あり学習と同様に、検証用データを用いたモデル選択を可能にするバリデーションネットワークを用いたモデル選択手法を提案する。バリデーションネットワークのノード役割を予測し、真の役割ラベルと比較し、精度が最大なモデルを選択する。

## 4. 提案手法

### 4.1 分散表現取得手法の取替

先行研究で用いたノードの分散表現を取得する struc2vec[18] は構造的に類似するノードを類似するベクトル表現に変換できる一方、重み付きネットワークや有向ネットワークに対応しないことや、ヒューリスティックな手法であるため計算時間やスケラビリティに課題が残されている。role2vec[20] はノードの高次な構造的特徴の集約に基づきノードを役割にマッピングした後に、ランダムウォークで取得した役割シーケンスから Skip-Gram モデルで役割の分散表現を取得する。ノードの分散表現はマッピングされた役割の分散表現と同じにする。つまり、同一役割に割り当てられたノードはすべて同じベクトル表現になる。role2vec は 1 ノードが 1 役割に割り当てられるため多重役割を発見できないことや、写像した役割が実ネットワークの多種多様な役割発見のニーズに応え難いことが考えられるが、数が多いノードを少数の役割に写像することで構造的特徴に特化した潜在空間への埋め込みや計算量の削減が図れる。

本研究では struc2vec と role2vec を分散表現の取得に使用して後述する実験を行うことで、提案手法がコンポーネントの取替ができるフレキシブルなフレームワークであることを示す。

### 4.2 ニューラルネットワークにおけるマルチラベル分類

本研究で、先行研究の Role-model を多重役割検出向けにマルチラベル分類できるよう、以下の改良を行う。

#### 4.2.1 出力層の構造変更

Role-model は 3 層のニューラルネットワークからなり、出力層が全結合層で役割の数分のニューロンを持つ。単一役割発見の Role-model は多値分類であるため、出力層の活性化関数

が LogSoftmax 関数で、以下の式になる。

$$\text{LogSoftmax}(y_i) = \log \left( \frac{\exp(x_i)}{\sum_{j=1}^r \exp(x_j)} \right) \quad (1)$$

各インスタンスの  $r$  個のニューロンの出力の内、最も高いニューロンが対応する役割ラベルが該当インスタンスの予測役割となる。

Role-model をマルチラベル分類できるように、出力層の活性化関数を以下に示す sigmoid 関数に変更する。

$$\text{Sigmoid}(y_i) = \frac{1}{1 + e^{-x_i}} \quad (2)$$

sigmoid 関数が実数を  $(0, 1)$  の区間に写像でき、二値分類の出力層の活性化関数によく使われる。マルチラベル分類において、各インスタンスの  $r$  個のニューロンの出力がそれぞれ  $r$  個のラベルに対応し、各ニューロンの出力を  $[0, 1]$  の区間に写像し、閾値より大きい場合は該当ラベルをポジティブに、閾値より小さい場合は該当ラベルをネガティブにする。

#### 4.2.2 損失関数の変更

マルチラベル分類タスクの損失関数についての研究が多く行われてきた [21, 22]。本研究では、BP-MLL という損失関数 [21] を採用する。BP-MLL はインスタンスに関連しないラベルを関連ラベルと予測した場合、ペナルティを課す。以下の式に示すよう、インスタンスに関連するラベルの出力と関連しないラベルの出力の差が大きければ大きいほど良い。

$$E = \sum_{i=1}^m E_i = \sum_{i=1}^m \frac{1}{|Y_i| |\bar{Y}_i|} \sum_{(k,l) \in Y_i \times \bar{Y}_i} \exp(-(c_k^i - c_l^i)) \quad (3)$$

$m$  は学習インスタンス数で、 $Y_i$  は  $i$  番目のインスタンスに対応するラベル集合である。 $\bar{Y}_i$  は  $Y_i$  の補集合で、インスタンス  $i$  に関連しないラベルの集合となる。 $c_k^i$  はインスタンス  $i$  のラベル  $k$  に対応するニューロンの出力である。

インスタンスに関連するラベルの出力と関連しないラベルの出力の全ペアの誤差を計上することで、関連するラベル集合の出力値がより高く、関連しないラベル集合の出力値がより低くなることになり、高精度にマルチラベル分類することが期待できる。

#### 4.3 コミュニティ抽出を用いた学習データの構築

実世界ネットワークにおける役割が多種多様でケースバイケースな状況がほとんどであり、更に、多重役割の場合、役割自体だけではなく、役割の共起状況もさまざまである。そのため、提案手法を適用するために、ソースネットワークとバリデーションネットワークの選定が非常に重要になる。本研究では、コミュニティ抽出を用いた学習データの構築手法を提案する。コミュニティの抽出は、ノードを内部が密に繋がるグループに分割するタスクで、Leiden 法 [23] と呼ばれるコミュニティ抽出アルゴリズムは、Louvain 法 [24] に基づき、コミュニティ内が必ず連結であることを保証する手法となる。

本研究では、連結したネットワークの分散表現がより良質に取得できるため、Leiden 法を用いて、実世界ネットワークの多重役割発見のための学習データを構築することを提案する。具体的に、実世界ネットワークにコミュニティ抽出手法を用いて分割したサブネットワークからランダム 2 つ選択し多重役割のラベリングを行い、それぞれソースネットワークとバリデーションネットワークにアサインすることで、他のサブネッ

トワーク、乃至全体のネットワークをターゲットネットワークに見なし、提案手法を適用し、多重役割予測を行う。コミュニティ抽出で取得した適切なサイズの 2 つのコミュニティに対して役割をラベリングし、それぞれソースネットワークとバリデーションネットワークに当てると考える。そして、ソースネットワークとバリデーションネットワークを除いた役割ラベルが付いていない部分をターゲットネットワークを見なし、提案手法を適用する。従って、実世界ネットワークにおける役割が多種多様でケースバイケースな状況においても、多重役割発見が期待できる。

## 5. 精度評価

実世界ネットワークを用いて多重役割発見の精度を確認する。ドメイン敵対的学習の精度評価とコミュニティ抽出で構築した学習データセットの精度評価を含む。

### 5.1 比較手法

本研究で使用した比較手法は、関連研究で紹介した BR (Binary Relevance), LP (Label Powerset), MLkNN になる。ドメイン敵対的学習の精度評価として、struc2vec 或いは role2vec より取得したソースネットワークの分散表現を入力とし、バリデーションネットワークの分散表現を検証データとし、比較手法を用いてターゲットネットワークの分散表現から役割ラベルを予測する。具体的に、BR では、各ラベルに対応するサポートベクトルマシンの二値分類器を学習させて予測させ、グリッドサーチによるハイパーパラメータチューニングを行う。LP では、各ラベルの組み合わせパターンに対応するランダムフォレストの多値分類器を学習させて予測させ、同様にグリッドサーチによるパラメータチューニングを行う。MLkNN はマルチラベル分類の手法のため、直接に各データセットを用いて精度検証を行う。

コミュニティ抽出で構築した学習データセットの精度評価として、ネットワーク全体の分散表現を struc2vec 或いは role2vec より取得した後、提案手法のソースネットワーク（バリデーションネットワーク）と同じサイズの訓練データ（検証データ）をランダムサンプリングして構築し、提案手法のターゲットネットワークに対して予測を行う。10 回のランダムサンプリングを行い、それぞれに BR, LP, MLkNN モデルで予測する。

### 5.2 提案手法のバリエーション

提案手法の Discriminator の有無、または Discriminator の種類によって、以下 4 つのバリエーションを提案手法とし、比較手法との精度比較を行うとともに、各バリエーションの性能を考察する。

**BPMLL**: Discriminator を使わない。つまり、ドメイン敵対的学習フレームワークを適用しない。

**BPMLL\_SD2**: 先行研究の提案手法と同じ、二値分類の Discriminator を 1 つ組み込む。ソースネットワークからの分散表現か、ターゲットネットワークとバリデーションネットワークの分散表現かを識別する。

**BPMLL\_SD3**: 多値分類 Discriminator を 1 つ組み込む。Discriminator の出力ニューロンが 3 つあり、それぞれソースネットワーク、ターゲットネットワーク、バリデーションネットワークに対応する。

**BPMLL\_DD**: ダブル Discriminator 方式で、ソースドメインとターゲットドメインを識別する二値分類の Discriminator 1 つと、ソースドメインとバリデーションドメインを識別する二値分類の Discriminator 1 つからなる。

表 1: Statistics of Wikipedia network and communities

Name	# Node	# Edge	# Role
Wiki_all	4,777	184,812	40
Wiki_c1	958	7,946	18
Wiki_c2	1,633	8,550	18
Wiki_c3	442	1,982	18

### 5.3 評価指標

本研究では、F-measure を用いて、精度評価を行う。

$$F\text{-measure} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

$$Precision = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i \cap Z_i|}{|Z_i|}, Recall = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i \cap Z_i|}{|Y_i|} \quad (5)$$

真のラベルマトリックスを  $Y \in \{0, 1\}^{n \times k}$  で、分類器の出力を  $Z \in \{0, 1\}^{n \times k}$  で表す。これはマルチラベル分類向けのインスタンススペースの評価指標である。

### 5.4 実験データ

表 2: Wikipedia datasets

No.	Source	Target	Validation
Wiki-1	Wiki_c1	Wiki_c2	Wiki_c3
Wiki-2	Wiki_c1	Wiki_c3	Wiki_c2
Wiki-3	Wiki_c2	Wiki_c1	Wiki_c3
Wiki-4	Wiki_c2	Wiki_c3	Wiki_c1
Wiki-5	Wiki_c3	Wiki_c1	Wiki_c2
Wiki-6	Wiki_c3	Wiki_c2	Wiki_c1

本論では Wikipedia ネットワーク [6] を実験データにする。Wikipedia ネットワークは、Wikipedia の一部分をコーパスとした単語の共起ネットワークであり、4,777 個のノードと 184,812 個のエッジ、および、40 種のノード役割ラベル (単語の品詞ラベル) を有する [6]。全体のネットワーク (Wiki\_all) に対して、コミュニティ抽出を行い、その結果から 3 つのサブネットワーク (Wiki\_c1, Wiki\_c2, Wiki\_c3) を無作為に選出し、それぞれソースネットワーク、ターゲットネットワークとバリデーションネットワークにした。各ネットワークの統計情報は Table1 に示す通りであり、6 通りの実験データを Table2 に示す。6 つのデータセットに対し、それぞれ 10 回の実験を行う。また、3 サブネットワークにおいて、共通する役割ラベル 18 個を予測対象とした。

### 5.5 実験結果

#### 5.5.1 struc2vec で取得した分散表現を用いた実験

struc2vec で取得した分散表現を用いた実験の F-measure の結果は図 2 に示す。横軸の BPMLL.S2V, SD2.S2V, SD3.S2V, DD.S2V は提案手法の各バリエーションで 10 回の実験結果を箱ひげ図で示したものに対し、横軸の MLkNN, BR(GS), LP(GS) の箱ひげ図はコミュニティ抽出で構築した学習データセットの精度評価の対比組である。また、ドメイン敵対的学習の精度評価の比較手法として、MLkNN, BR(GS), LP(GS) の結果はそれぞれ赤、ピンク、黄色の横線で表す。

ドメイン敵対的学習の精度評価として、提案手法の Discriminator を含むバリエーション (BPMLL\_SD2, BPMLL\_SD3,

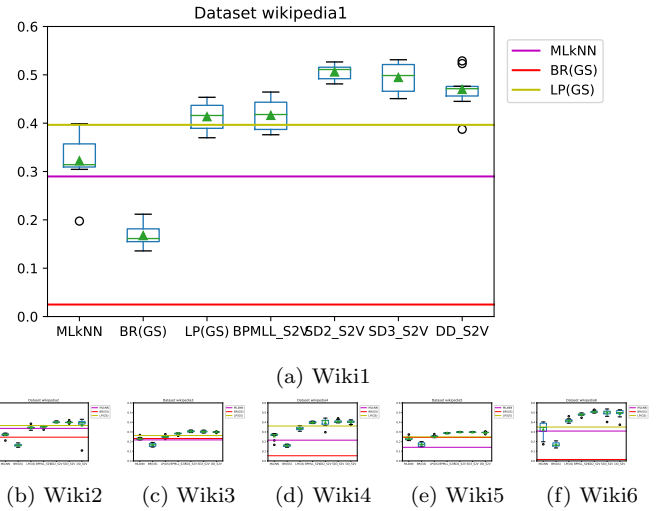


図 2: Wikipedia ネットワークの実験の F-measure 結果 (struc2vec)

BPMLL\_DD) は BPMLL および比較手法に比べると、F-measure 平均値が高い結果になっている。図 2b に示す Wiki2 の場合、BPMLL\_SD2 と BPMLL\_SD3 は精度が一番高い比較手法である LP の F-measure の 1.3 倍の精度を実現した。BPMLL は Wiki1 と Wiki3 以外の場合では、比較手法より高い精度を実現したが、Discriminator を含む提案手法に比べると、全ての場合に F-measure 平均値が低くなっている。これは、提案手法のドメイン敵対的学習フレームワークがソースネットワークのノードの分散表現と、ターゲットネットワークおよびバリデーションネットワークのものを類似させることで、よりはっきりとした分類境界を学習できたことを示すと考えられる。

コミュニティ抽出で構築した学習データセットの精度評価として、比較手法 (ランダムサンプリング) の内、精度が最も高かった LP(GS) は Discriminator 無しの BPMLL と同程度の精度を達している。同じ比較手法の場合、ランダムサンプリング (箱ひげ図) とコミュニティ抽出手法で構築したデータセット (横線) は、ランダムサンプリングのほうが高い精度の場合 (Wiki1, Wiki6) もあれば、コミュニティ抽出手法で構築したデータセットのほうが高い精度の場合 (Wiki2, Wiki3, Wiki4) もある。これはランダムサンプリングの揺らぎやターゲットネットワークの多重役割分布の特徴による結果であると考察する。また、ランダムサンプリングで構築した実験データセットでは、訓練データと検証データに存在する役割しか予測できないため、役割の種類をコントロールできなく、予測する役割数はトライ毎に異なり、16 から 20 までであった。

また、提案手法の Discriminator を含むバリエーション (BPMLL\_SD2, BPMLL\_SD3, BPMLL\_DD) 同士で比べると、各組み合わせパターンにおいては、平均値や分散、および、異常値が変化しているが、全体としては特に大きな傾向性が見当たらない。

#### 5.5.2 role2vec で取得した分散表現を用いた実験

role2vec で取得した分散表現を用いた実験の F-measure の結果は図 3 に示す。図 3 の見方は前述した図 2 の見方と同じである。

role2vec の結果は struc2vec の結果と似た傾向を示すが、以下で異なる部分を抜粋して説明する。ドメイン敵対的学習の精度評価として、提案手法の Discriminator を含むバリエーション

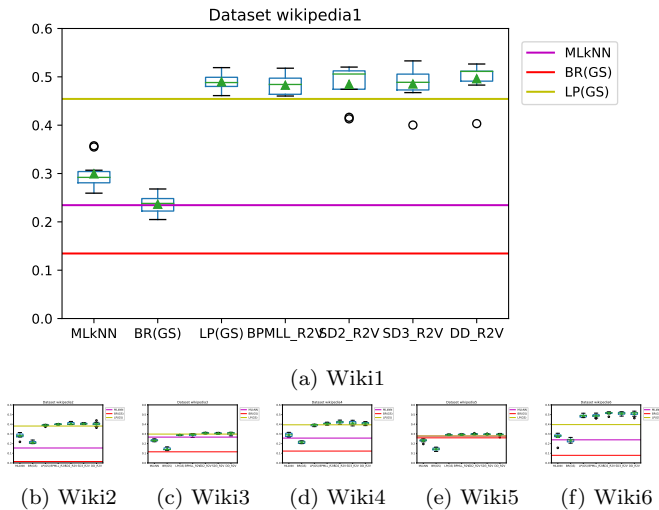


図 3: Wikipedia ネットワークの実験の F-measure 結果 (role2vec)

ン (BPMLL\_SD2, BPMLL\_SD3, BPMLL\_DD) は BPMLL および比較手法に比べると、F-measure 平均値が僅かに高い結果か、同等な精度になっている。その内、BPMLL\_SD2 が他に比べるとやや高い精度である。これは、role2vec で取得した分散表現は複数ノードに共有しているため、説明変数自体のばらつきが小さく、簡単な予測モデルでもある程度予測できることが考えられる。

一方、提案手法の Discriminator を含むバリエーション (BPMLL\_SD2, BPMLL\_SD3, BPMLL\_DD) のデータセット毎の予測結果として、struc2vec と role2vec が同レベルの精度に達している。これは各データセットの構造的情報が多重役割への寄与度が異なることが原因であると考えられる。

## 6. 終わりに

本研究では、複雑ネットワークにおける多重役割発見する手法を提案した。また、実世界ネットワークの役割発見タスク向け、コミュニティ抽出で学習データの構築手法を提案した。実世界ネットワークの Wikipedia ネットワークにおいて、提案手法が高い精度を達成し、様々な実世界ネットワークに適用できることが明らかになった。

本研究の今後の課題として、ドメイン敵対的学習フレームワークについて提案した多値分類 Discriminator 方式やダブル Discriminator 方式の最適化、ハイパーパラメータチューニングの効率化、多重役割発見の精度向上、分散表現を効率的に取得できる手法の適用が考えられる。

## 参考文献

- [1] Ryan A Rossi and Nesreen K Ahmed. Role discovery in networks. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 27, No. 4, pp. 1112–1131, 2014.
- [2] Priya Mahadevan, Dmitri Krioukov, Marina Fomenkov, Xenofontas Dimitropoulos, KC Claffy, and Amin Vahdat. The internet as-level topology: three data sources and one definitive metric. *ACM SIGCOMM Computer Communication Review*, Vol. 36, No. 1, pp. 17–26, 2006.
- [3] Ryan Rossi, Sonia Fahmy, and Nilothpal Talukder. A multi-level approach for evaluating internet topology generators. In *2013 IFIP Networking Conference*, pp. 1–9. IEEE, 2013.
- [4] Ryan A Rossi, Brian Gallagher, Jennifer Neville, and Keith Henderson. Modeling dynamic behavior in large evolving

- graphs. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 667–676, 2013.
- [5] Shumpei Kikuta, Fujio Toriumi, Mao Nishiguchi, Shu Liu, Tomoki Fukuma, Takanori Nishida, and Shohei Usui. Framework for role discovery using transfer learning. *Applied Network Science*, Vol. 5, No. 1, pp. 1–19, 2020.
- [6] Jie Zhang, Yuxiao Dong, Yan Wang, Jie Tang, and Ming Ding. Prone: Fast and scalable network representation learning. In *IJCAI*, Vol. 19, pp. 4278–4284, 2019.
- [7] Phipps Arabie, Scott A Boorman, and Paul R Levitt. Constructing blockmodels: How and why. *Journal of mathematical psychology*, Vol. 17, No. 1, pp. 21–63, 1978.
- [8] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, Vol. 5, No. 2, pp. 109–137, 1983.
- [9] Krzysztof Nowicki and Tom A. B Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, Vol. 96, , 9 2001.
- [10] Vladimir Batagelj, Andrej Mrvar, Anuška Ferligoj, and Patrick Doreian. Generalized blockmodeling with pajek. *Metodoloski zvezki*, Vol. 1, No. 2, pp. 455–467, 2004.
- [11] Edoardo Maria Airoidi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed membership stochastic blockmodels. *Journal of machine learning research*, 2008.
- [12] Ronald S Burt. Positions in networks. *Social forces*, Vol. 55, No. 1, pp. 93–122, 1976.
- [13] Ulrik Brandes and Jürgen Lerner. Structural similarity: Spectral methods for relaxed blockmodeling. *Journal of classification*, Vol. 27, No. 3, pp. 279–306, 2010.
- [14] Qiaoyu Tan, Yezi Liu, Xia Chen, and Guoxian Yu. Multi-label classification based on low rank representation for image annotation. *Remote Sensing*, Vol. 9, No. 2, p. 109, 2017.
- [15] Francisco Herrera, Francisco Charte, Antonio J Rivera, and Maria J Del Jesus. Multilabel classification. In *Multilabel Classification*, pp. 17–31. Springer, 2016.
- [16] Min-Ling Zhang and Zhi-Hua Zhou. Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, Vol. 40, No. 7, pp. 2038–2048, 2007.
- [17] Mark Baker and William Croft. Lexical categories: Legacy, lacuna, and opportunity for functionalists and formalists. *Annual Review of Linguistics*, Vol. 3, pp. 179–197, 2017.
- [18] Leonardo FR Ribeiro, Pedro HP Saverese, and Daniel R Figueiredo. 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, 2017.
- [19] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, Vol. 17, No. 1, pp. 2096–2030, 2016.
- [20] Nesreen Ahmed, Ryan Anthony Rossi, John Lee, Theodore Willke, Rong Zhou, Xiangnan Kong, and Hoda Eldardiry. Role-based graph embeddings. *IEEE Transactions on Knowledge and Data Engineering*, 2020.
- [21] Min-Ling Zhang and Zhi-Hua Zhou. Multilabel neural networks with applications to functional genomics and text categorization. *IEEE transactions on Knowledge and Data Engineering*, Vol. 18, No. 10, pp. 1338–1351, 2006.
- [22] Rafał Grodzicki, Jacek Mańdziuk, and Lipo Wang. Improved multilabel classification with neural networks. In *International Conference on Parallel Problem Solving from Nature*, pp. 409–416. Springer, 2008.
- [23] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, Vol. 9, No. 1, pp. 1–12, 2019.
- [24] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, Vol. 2008, No. 10, p. P10008, 2008.