

## Graph of Graphs に対する二重畳み込みニューラルネットワーク

## Dual Convolutional Neural Network for Graph of Graphs Link Prediction

原田 将之介<sup>\*1</sup>  
Shonosuke Harada秋田 大空<sup>\*1</sup>  
Hirotaka Akita椿 真史<sup>\*2</sup>  
Masashi Tsubaki馬場 雪乃<sup>\*3</sup>  
Yukino Baba瀧川 一学<sup>\*4</sup>  
Ichigaku Takigawa山西 芳裕<sup>\*5</sup>  
Yoshihiro Yamanishi鹿島 久嗣<sup>\*1\*6</sup>  
Hisashi Kasima<sup>\*1</sup>京都大学  
Kyoto University<sup>\*2</sup>産業技術総合研究所  
National Institute of Advanced Industrial Science and Technology<sup>\*3</sup>筑波大学  
University of Tsukuba<sup>\*4</sup>北海道大学  
Hokkaido University<sup>\*5</sup>九州大学  
Kyushu University<sup>\*6</sup>理化学研究所 革新知能統合研究センター  
RIKEN Center for Advanced Intelligence Project

A graph is a general and powerful data representation for complex real-world phenomena ranging from chemical compounds to social networks; however, effective feature extraction from graphs is not a trivial task, and much work has been done in the field of machine learning and data mining. In this paper, we propose a dual convolution approach that extracts node representations of a *graph of graphs* (GoG) consisting of an external graph and internal graphs, by combining the external and internal graph structures in an end-to-end manner. Experiments on various types of link prediction tasks demonstrate the effectiveness of the proposed method.

## 1. はじめに

グラフは、化合物やソーシャルネットワークのような現実世界の複雑な現象をモデル化する一般的かつ強力なデータ表現である。しかしながら、データ解析においてそのようなグラフ構造の扱いは自明ではない。というのも、ほとんどの機械学習手法はデータは予め固定長の特徴ベクトルで表現されるという前提の上に成り立っているからである。従って、グラフ構造データの分析は多くの研究者の関心を集めている課題のひとつである。

グラフ構造は大きく 2 種類に分けることができる。データ内部の構造を表現する内部グラフ及びデータ間の関係性を表現する外部グラフである。内部グラフの例として化合物や Web ページ、外部グラフの例としてソーシャルネットワークや Web ページのリンク関係を表すネットワーク等がそれぞれ挙げられる。本研究では、両者の特徴を持つ図 1 のような graph of graphs (GoG) と呼ばれるグラフに対する適切な特徴表現の獲得を目的とする。GoG の例として、化合物相互作用ネットワークや Web ページリンクネットワークが挙げられる。

単純な一つのグラフの集合より複雑で一般的なグラフ構造を持つ GoG に対して、これまで異なる文脈で別々に研究が行われてきた内部、外部グラフの両者の特徴を共に考慮する手法を提案し、3 つの実 GoG データセットを使ったリンク予測の実験において、二重畳み込み法がベースライン手法と比較して低次元特徴ベクトルを用いた場合でも、ベースライン手法より優れていることを示す。

## 2. 問題設定

本研究において、GoG とは、グラフの各ノード自体が、ある別のグラフ構造を持つグラフ構造であると定義する。本論文では元になるグラフを外部グラフ、外部グラフのノードを構成するグラフを内部グラフと呼び、外部グラフにおけるノード間

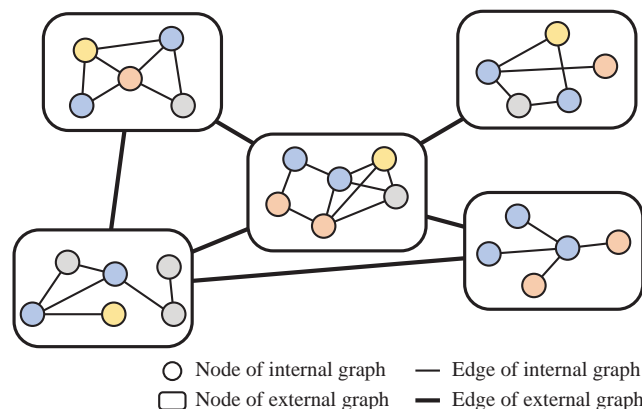


図 1: GoG の例

のリンク予測を行い、得られた特徴ベクトルの適切さを評価する。一般的に、GoG は内部グラフと外部グラフの 2 層だけでなく、何層にでもなりうるが、本研究では簡単のため 2 層の GoG のみを扱う。なお、本研究のアプローチは何層の GoG に対しても容易に応用可能である。

2 層の GoG は、 $\mathcal{G} = (\mathcal{V}, \mathcal{A}, \mathcal{E})$  として表されるグラフである。 $\mathcal{V}$  は外部グラフのノードの集合、 $\mathcal{A}$  は外部グラフの隣接リスト、 $\mathcal{E}$  は外部グラフのノード間のリンクの集合を表す。外部グラフ  $\mathcal{G}$  の  $i$  番目のノードは内部グラフ  $G_i \in \mathcal{V}$  をもつ。簡単のために、以降  $G_i = G = (\mathcal{V}, \mathcal{A}, \mathcal{E})$  と表記する。ここで、 $\mathcal{V}$  は内部グラフのノードの集合、 $\mathcal{A}$  は内部グラフの隣接リスト、 $\mathcal{E}$  は内部グラフのノード間のエッジの集合である。簡単のために、本研究では  $\mathcal{E}$  の種類については考慮しない。

本問題の入出力をまとめると以下ようになる：

入力： $\mathcal{G} = (\mathcal{V}, \mathcal{A}, \mathcal{E})$

出力： $G_i, G_j \in \mathcal{V}$  間にリンクが存在する確率

### 3. 関連研究

本研究は、これまで異なる背景で別々に研究が行われてきた内部グラフ及び外部グラフの統合を試みているため、これまでのグラフ学習の研究をそれら 2 種類のグラフの観点から述べる。

内部的なグラフ構造を持ったデータに対する分析は、データマイニング分野に端を発する。その根底にある考え方は、グラフの特徴は局所的な部分構造にあるとするものであり、そのような局所部分構造を見つけるような拡張が成された頻出パターンマイニング手法 [Inokuchi 00, Yan 02] が提案された。明示的にグラフの部分構造の特徴を見つけるグラフマイニングとは対照的に、カーネル法は暗黙的に特徴を抽出する。初期のグラフカーネル法 [Kashima 03, Gärtner 03] は、パスを特徴ベクトルとして用いた。昨今のグラフカーネルの最先端の手法 [Shervashidze 11] はより複雑な部分構造の形式を用いている。

ニューラルネットワークもまたグラフに対する柔軟な特徴抽出に成功している。Scarse らによるもの [Scarselli 09] に続いて、様々なグラフ畳み込みモデルが提案されてきた。例えば、Duvinaud らによるもの [Duvinaud 15] や Niepert らによるもの [Niepert 16] が挙げられる。

外部グラフ構造を持ったデータに対する分析は、しばしばリンクマイニングと呼ばれる [Getoor 05]。リンクマイニングの典型的なタスクには、リンク予測の他にランキング、クラスタリング、ノード分類等がある。近年、グラフ内でのノードの近接性を維持するノード埋め込みによる研究 [Socher 13, Yang 15] がますます行われている。特に Yang ら [Yang 15] の考え方は、評価実験におけるベースラインとして用いている。

グラフ構造を持つデータに対する研究は非常に幅広いため、ここで全ての範囲を網羅することは困難である。重要な点は、これまで内部グラフと外部グラフは別々に研究が行われてきていたという点である。

### 4. 二重畳み込みニューラルネットワーク

GoG の特徴表現を学習するために、本研究ではニューラルネットワークを用いて内部グラフ及び外部グラフを考慮した一貫学習を行う。

#### 4.1 内部グラフに対する畳み込み

$G$  の  $k$  番目のノード  $v_k$  の特徴ベクトルを  $\mathbf{v}_k \in \mathbb{R}^d$  とする。 $d$  は特徴ベクトルの次元を表し、モデルのハイパーパラメータの一つである。各  $\mathbf{v}_k$  の初期値  $\mathbf{v}_k^{(0)}$  はノードの種類 (例えば、水素や酸素) に応じてランダムに初期化され、その後、外部グラフに対するリンク予測の際の誤差逆伝搬で更新される。

グラフ  $G = (V, A)$  におけるノードの特徴ベクトルを  $\mathbf{v}_k$  とし、畳み込みステップ  $t = 1, 2, \dots, T$  における  $\mathbf{v}_k$  を  $\mathbf{v}_k^{(t)}$  とし、各畳み込みステップにおいて以下のニューラルネットワークを用いて特徴を更新する。

$$\mathbf{v}_k^{(t+1)} = f_G \left( \mathbf{W} \mathbf{v}_k^{(t)} + \sum_{m \in A_k} \mathbf{M} \mathbf{v}_m^{(t)} \right) \quad (1)$$

ここで  $f_G$  は例えば ReLU のような非線形関数、 $A_k$  は  $k$  番目の隣接リストの索引、 $\mathbf{W} \in \mathbb{R}^{d \times d}$ 、 $\mathbf{M} \in \mathbb{R}^{d \times d}$  は学習される重み行列を示す。内部グラフに対する畳み込みについての関数の詳細な設定は、次節で説明する外部グラフに対する畳み込みの設定と共に、5.2 節で改めて述べる。

最終的に  $G = (V, A)$  における全てのノードの特徴を以下のように全ての畳み込みステップを通して足し合わせることで、内部グラフの特徴表現  $\mathbf{g}^{(T)} \in \mathbb{R}^d$  を得る。

$$\mathbf{g}^{(T)} = \sum_{k \in V} g \left( \sum_{t=0}^T \mathbf{v}_k^{(t)} \right) \quad (2)$$

ここで、 $T$  は内部グラフに対する畳み込みを行う回数、 $g$  は非線形関数である。続いて、 $i$  番目の内部グラフの特徴表現  $\mathbf{g}^{(T)}$  を以降では改めて  $\mathbf{g}_i^{(0)}$  と表記する。右肩の 0 は次節で説明する外部グラフに対する畳み込みの初期特徴ベクトルであることを表している。

#### 4.2 外部グラフに対する畳み込み

上述した内部グラフの特徴表現  $\mathbf{g}_i^{(0)}$  を用いて以下に示すような外部グラフに対する畳み込みを行う。

$$\mathbf{g}_i^{(\ell+1)} = f_G \left( \mathbf{U} \mathbf{g}_i^{(\ell)} + \sum_{n \in \mathcal{A}_i} \mathbf{V} \mathbf{g}_n^{(\ell)} \right) \quad (3)$$

ここで、 $f_G$  は非線形関数、 $\mathcal{A}_i$  は  $i$  番目の隣接リストの索引、 $\mathbf{U} \in \mathbb{R}^{d \times d}$  及び  $\mathbf{V} \in \mathbb{R}^{d \times d}$  は学習される行列パラメータを表す。そして、以下のように全ての畳み込みステップを考慮して新たな内部グラフの特徴表現を得る。

$$\mathbf{g}_i^{(L)} = g \left( \sum_{\ell=0}^L \mathbf{g}_i^{(\ell)} \right) \quad (4)$$

ここで  $L$  は外部グラフに対する畳み込みを行う回数、 $g$  は非線形関数を表す。また  $\mathbf{g}_i^{(L)}$  は内部グラフ  $G_i$  の外部グラフに対する畳み込みの畳み込みステップ  $L$  における特徴表現を表す。本手法の目的は外部グラフ  $G$  の特徴表現を獲得することではなく、内部グラフ  $G_i$  の適切な特徴表現を獲得することであることに注意されたい。本研究では  $\mathbf{g}_i^{(L)}$  を内部グラフ  $G_i$  の二重畳み込み特徴表現と呼ぶ。これは、次節で説明する外部グラフにおけるリンク予測の際に入力される特徴ベクトルとなる。

#### 4.3 リンク予測

2 つの内部グラフ  $G_i$  及び  $G_j$  間のリンク予測は、それらの二重畳み込み特徴表現  $\mathbf{g}_i^{(L)}$  及び  $\mathbf{g}_j^{(L)}$  を用いて行われる。以下のように入力として、2 つの特徴ベクトル  $\mathbf{g}_i^{(L)}$  及び  $\mathbf{g}_j^{(L)}$  を受け取り、ニューラルネットワークを用いてリンクの有無の度合いを出力する関数  $h$  によって 2 次元ベクトル  $\mathbf{y} \in \mathbb{R}^2$  を出力する。

$$\mathbf{y} = h \left( \mathbf{g}_i^{(L)}, \mathbf{g}_j^{(L)} \right) \quad (5)$$

さらに、softmax 層によってリンクの確率を以下のように与える。

$$p_t = \frac{\exp(y_t)}{\sum_k \exp(y_k)}$$

ここで  $t \in \{0, 1\}$  は 2 値 (すなわちリンクが存在しているか否か) のラベル、 $p_t$  は  $t$  の確率をそれぞれ表している。関数  $h$  について詳細は 5.2 節で述べるが、入力される  $i$  及び  $j$  について、 $h$  によって対称性は維持されていることに注意されたい。

### 5. 評価実験

#### 5.1 データセット

GoG に対する内部グラフ及び外部グラフを考慮する一貫学習の有用性を確認するため、リンク予測を行う。リンク予測実験で用いた 3 つの実データセットについて説明する。

### 5.1.1 標的共有ネットワーク

標的共有ネットワークはタンパク質と化合物の相互作用についてのデータセット [Takigawa 11] である。元々のグラフは2部グラフであるが、この実験では、このデータセットでは同じタンパク質と作用する薬同士を繋いだ薬剤のみのネットワークに変換して用いる。化合物に含まれる原子数が64以下の化合物3,918を用いた。化合物のペア  ${}_{3918}C_2 = 7,673,403$  のうち、リンクがあるものは35,562であり、残りは負例として扱う。上述のように、薬剤のみのネットワークに変換しているが、本研究では便宜上このデータを標的共有ネットワークと呼ぶ。

### 5.1.2 薬剤相互作用ネットワーク

薬剤化合物のネットワークで、2つの化合物間で、同時に服用すると相互作用や、人体に悪影響のある反応が見られる場合にその2つの化合物の間にエッジが存在する。化合物に含まれる原子数が64以下の化合物1,993を用いる。薬剤ペア  ${}_{1993}C_2 = 1,985,028$  のうち、リンクがあるものは186,555であり、残りは負例として扱う。

### 5.1.3 代謝化合物ネットワーク

代謝化合物のネットワークであり、2つの化合物が代謝経路上の酵素反応において、基質及び生成物の対である場合にそれらの間にリンクを張る [Kotera 14]。化合物に含まれる原子数が64以下の化合物5,920を用いる。化合物ペア  ${}_{5920}C_2 = 17,520,240$  のうち、リンクがあるものは5,041、リンクがないものは220,096であり、残りは正解ラベルが与えられておらず、今回の実験では扱わない。

## 5.2 提案手法の設定の詳細

2つの畳み込みネットワーク内で用いる非線形関数には選択の幅がある。本実験では以下のような選択を行う。

式(1)における関数  $f_G$  として ReLU を用いる。また行列パラメータ  $\mathbf{W}$  及び  $\mathbf{M}$  を次数 (すなわち  $|A_k|$  及び  $|A_m|$ ) とその畳み込みステップに応じて初期化する。重み行列パラメータをそれぞれの次数について初期化するという考え方は、Duvinaudら [Duvinaud 15] の考えに基づいている。式(2)及び式(4)における関数  $g$  として softmax 関数を用いる。式(1)及び式(3)における関数  $f_G$  として softmax 関数を用い、行列パラメータ  $\mathbf{U}$  及び  $\mathbf{V}$  を畳み込みステップに応じて初期化する。 $\mathbf{W}$  及び  $\mathbf{M}$  とは異なり、 $\mathbf{U}$  及び  $\mathbf{V}$  については本実験で扱ったデータセットでは、外部グラフが巨大で各ノードが非常に大きな次数を持っているために次数を考慮しない。また今回の実験では  $\mathbf{U}=\mathbf{V}$  としている。4.3節における関数  $h$  として以下のような関数を用いた。関数  $h$  に対して入力される  $\mathbf{g}_i^{(L)}$  及び  $\mathbf{g}_j^{(L)}$  は  $(\mathbf{g}_i^{(L)} + \mathbf{g}_j^{(L)}) \oplus (\mathbf{g}_i^{(L)} \odot \mathbf{g}_j^{(L)})$  と処理された後、ニューラルネットワークの入力として扱われる。ここで、 $\oplus$  は2つのベクトルの連結、 $\odot$  はアダマール積を表す。ニューラルネットワークの構造は (128, 64, 2) であり、全ての非線形活性化関数は ReLU を用いる。

## 5.3 比較手法

比較手法として、内部グラフに対する畳み込みのみを行う手法 (internal)、Yang らによるもの [Yang 15] の踏襲で外部グラフにおけるノードの特徴ベクトルを外部グラフにおけるリンク関係から学習する外部グラフへの埋め込み (external)、化合物の部分構造に対してインデックスを割り当て特徴表現を獲得する Morgan fingerprint [Morgan 65] (morgan) を用いる。今回の実験では Morgan fingerprint は 2048 次元、それ以外は 64 次元の特徴ベクトルを用いた。これは Morgan fingerprint を小さな次元で適用すると衝突により、性能が大きく低下するためである。

## 5.4 実験結果

訓練データの大きさを変えながら評価したものを図 2a、図 2b、図 2c に示す。横軸が訓練データ (リンク) のサイズ、縦軸は AUC である。なお、訓練データは正例、負例の割合をそれぞれ 50% とした。

### 5.4.1 標的共有ネットワーク

標的共有ネットワークでは、外部グラフの埋め込み手法の予測性能は低い。これは、ネットワークが疎であるために単純にノードの組み合わせのみを考えるだけでは予測が難しいためだと考えられる。一方で、Hashed Morgan fingerprint は訓練データが小さい場合は提案手法を上回る性能を発揮しており、fingerprint に基づく手法の頑健性を示している。しかし、訓練データのサイズが大きくなると、提案手法が Hashed Morgan fingerprint を上回る性能を示している。これは外部グラフに対する畳み込みによって提案手法の予測性能が大きく向上したためであると考えられる。また内部グラフに対する畳み込みのみの手法を提案手法が上回っており、二重畳み込みが有効に働いていることが分かる。

### 5.4.2 薬剤相互作用

薬剤相互作用では、ネットワークが密であることから外部グラフのみで特徴学習を行う外部グラフの埋め込み手法が優れた性能を示しているが、訓練データサイズが小さくなるにつれて、その性能を大きく低下させている。一方で、訓練データサイズが大きい時は提案手法と外部グラフの埋め込み手法はほぼ同様の性能を示しており、また訓練データサイズが小さい時であっても、内部グラフ及び外部グラフ両方の情報を取り入れているため大幅な性能の低下は免れており、頑健性を示している。

### 5.4.3 代謝化合物

代謝化合物ネットワークは本研究の中で最も疎なデータセットであり、提案手法が効果を発揮していないだけでなく逆に悪化させてしまう結果を示している。これは、外部グラフが非常に疎であるために、情報を取り入れることが困難になり有効な特徴ベクトルを抽出できていないためと考えられる。同様に外部グラフのスパース性によって外部グラフの埋め込み手法も大きく性能を低下させていることが分かる。一方でベースライン手法である Hashed Morgan fingerprint と内部グラフに対する畳み込みは外部グラフのスパース性の影響を受けない特徴ベクトルを用いているためか、優れた性能を示している。

## 6. おわりに

本研究では、これまであまり研究対象としては扱われてこなかった GoG に対して内部グラフ及び外部グラフを共に考慮し、一貫で特徴表現を学習する二重畳み込みニューラルネットワークを提案し、特徴表現学習の問題について取り組んだ。3つの実データに対するリンク予測実験で提案手法の有用性を示した。提案手法である二重畳み込みによるアプローチは、標的共有ネットワーク及び薬剤相互作用ネットワークという一定のリンク密度があるネットワークで高い予測性能を示し、特徴ベクトルの次元がベースライン手法の1つである Hashed Morgan fingerprint に比べて低次元であっても同等以上の性能を示した。また提案手法が、外部グラフが非常に疎であるとき、情報を十分に取り入れることが困難になるため性能が大きく低下するというを確認した。

将来の展望として、ノード分類やノードクラスタリング等のタスクについても応用する所存である。本実験では、外部グラフに対する畳み込みの際用いる訓練データから得られる完全



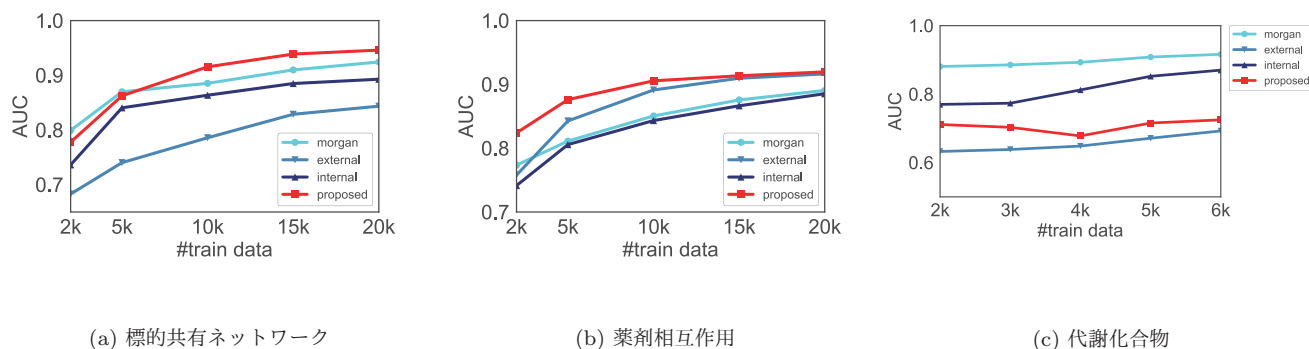


図 2: 各データセットにおける訓練データの変化に対する AUC の推移

な隣接行列について、経験上性能が向上したため 2 割の隣接関係を無視した隣接行列を用いている。隣接行列の運用法による影響については今後詳しく調べていく必要があると考える。上記も含めたハイパーパラメータの調査や、比較手法として新たな比較手法の採用を検討する。

## 参考文献

- [Duvenaud 15] Duvenaud, D. K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., and Adams, R. P.: Convolutional Networks on Graphs for Learning Molecular Fingerprints, in *Advances in Neural Information Processing Systems (NIPS)* (2015)
- [Gärtner 03] Gärtner, T., Flach, P., and Wrobel, S.: On graph kernels: hardness results and efficient alternatives, in *Learning Theory and Kernel Machines*, pp. 129–143 (2003)
- [Getoor 05] Getoor, L. and Diehl, C. P.: Link mining: a survey, *SIGKDD Explorations*, Vol. 7, No. 2, pp. 3–12 (2005)
- [Inokuchi 00] Inokuchi, A., Washio, T., and Motoda, H.: An apriori-based algorithm for mining frequent substructures from graph data, in *Proceedings of the Fourth European Conference on Principles of Data Mining and Knowledge Discovery (PKDD)* (2000)
- [Kashima 03] Kashima, H., Tsuda, K., and Inokuchi, A.: Marginalized kernels between labeled graphs, in *Proceedings of the 20th International Conference on Machine Learning (ICML)* (2003)
- [Kotera 14] Kotera, M., Tabei, Y., Yamanishi, Y., Muto, A., Moriya, Y., Tokimatsu, T., and Goto, S.: Metabolome-scale prediction of intermediate compounds in multistep metabolic pathways with a recursive supervised approach, *Bioinformatics*, Vol. 30, No. 12, pp. i165–i174 (2014)
- [Morgan 65] Morgan, H. L.: The generation of a unique machine description for chemical structures—a technique developed at chemical abstracts service, *Journal of Chemical Documentation*, Vol. 5, No. 2, pp. 107–113 (1965)
- [Niepert 16] Niepert, M., Ahmed, M., and Kutzkov, K.: Learning convolutional neural networks for graphs, in *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pp. 2014–2023 (2016)
- [Scarselli 09] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G.: The graph neural network model, *IEEE Transactions on Neural Networks*, Vol. 20, No. 1, pp. 61–80 (2009)
- [Shervashidze 11] Shervashidze, N., Schweitzer, P., van Leeuwen, E. J., Mehlhorn, K., and Borgwardt, K. M.: Weisfeiler-Lehman graph kernels, *Journal of Machine Learning Research*, Vol. 12, No. Sep, pp. 2539–2561 (2011)
- [Socher 13] Socher, R., Chen, D., Manning, C. D., and Ng, A.: Reasoning with neural tensor networks for knowledge base completion, in *Advances in Neural Information Processing Systems (NIPS)* (2013)
- [Takigawa 11] Takigawa, I., Tsuda, K., and Mamitsuka, H.: Mining significant substructure pairs for interpreting polypharmacology in drug-target network, *PLoS ONE*, Vol. 6, No. 2, p. e16999 (2011)
- [Yan 02] Yan, X. and Han, J.: gSpan: Graph-based substructure pattern mining, in *Proceedings of the Second IEEE International Conference on Data Mining (ICDM)* (2002)
- [Yang 15] Yang, B., Yih, W.-t., He, X., Gao, J., and Deng, L.: Embedding entities and relations for learning and inference in knowledge bases, in *Proceedings of the Third International Conference on Learning Representations (ICLR)* (2015)