

# 構造に基づくリンクの役割発見

## Role discovery of the links based on the network structure

菊田 俊平<sup>\*1</sup>    鳥海 不二夫<sup>\*1</sup>    西口 真央<sup>\*1</sup>  
Shumpei Kikuta    Fujio Toriumi    Mao Nishiguchi

<sup>\*1</sup>東京大学工学系研究科

The University of Tokyo School of Engineering

This paper aims at the discovery of link roles in order to understand links on the network. This work presents a flexible, general framework including graph transformation, representation learning, role assignment, and sense making. We use Edge-dual graph to regard links as nodes and struc2vec to embed links based on roles. We show our model successfully embed the similar links into the low 2-dimensional space on visualization task. Furthermore, we assign roles to links and conclude the structure is critical for a better understanding of links. Future work includes the automatic algorithm to decide the number of clusters and apply our method to real-world datasets.

### 1. はじめに

ソーシャルネットワークにおいて、リンクは人と人との関係性を表し、そのリンクの性質は、「友人関係」、「上司と部下の関係」、「家族の関係」など多岐にわたる。これらのリンクの性質を理解することができれば、適切な推薦や行動を取ることが可能になるため、リンクの性質の理解は意義がある。

リンクの性質の理解には、リンクの構造が役立つと考えられる。なぜなら、二者の関係性は、二者の他の人との関係性の構造によって構築されると考えられるからである。例えば、所属企業が異なる二者 A, B が存在し、それぞれの所属する企業が頻繁に取引を行う場合、A, B はコミュニティをつなぐ役割を果たす。このとき、二者の間のリンクは、コミュニティをつなぐブリッジとしての役割を持ち、二者間の関係性が「得意先の関係」などと理解できる。

ネットワークにおけるノードやリンクの構造は、役割と言い換えられ、役割発見というテーマは、ノードを主な対象として、グラフ理論や社会科学の研究コミュニティにおいて発展し、様々な手法が提案されている [Rossi 15]。また、近年のネットワークの大規模化や自然言語処理 [Mikolov 13] での利用に伴い、低次元のベクトル空間に正確に射影することを目的とする、表現学習と呼ばれる研究領域が注目を集めている [Goyal 18][Cui 18]。表現学習は、分類やリンク予測のタスクにおいて利用されており、役割発見においても利用可能と考えられる。

しかし、従来の研究対象はノードであることが多く、リンクの理解に焦点を当てた研究は少ない。また、役割発見や表現学習の領域においても、リンクに対する研究は活発に行われていない。

本研究の目的は、リンクを類似した構造を持つリンク群に分割し、リンク群に役割をラベリングすることで、リンクの性質を理解することである。手法として、リンクの役割発見を行うための汎用的なフレームワークを導入する。提案フレームワークは、リンクをノードとして扱う辺双対グラフ及び、構造に注目した表現学習である struc2vec[Ribeiro 17] を利用している。実験では、ベンチマークネットワークにおける、可視化によって、類似した構造を持つリンクは近い表現を持つことを

確認し、クラスタリングによって、類似リンクにラベリングを行うことで、リンクの性質を理解できることが示す。

本論文は以下のように構成される。第2章では、関連研究について述べ、第3章において、提案手法の詳細を述べる。第4章において、実験とその結果・考察を述べる。最後に第5章で、結論と今後の展望について述べる。

### 2. 関連研究

本研究における関連研究として、ネットワークにおける役割発見とネットワーク表現学習について述べる。

#### 2.1 ネットワークにおける役割発見

ネットワークからの役割発見に関する研究は、これまでも数多く行われてきた [Rossi 15]。役割発見のアルゴリズムはグラフから直接役割を割り当てる手法と、グラフから特徴量を抽出し、その特徴量に応じて役割を割り当てる手法に大別される [Rossi 15]。前者の例として Blockmodel[White 76] は、役割グラフを作成することで、構造的に類似するノード群に分割する手法であり、多数の類似手法が提案されている。後者の例として rolx[Henderson 12] は、グラフにおける構造的な特徴量を抽出し、その特徴量に対して行列分解を行い、役割を割り当てる手法である。しかし、既存の研究において、対象はノードであり、リンクに対する手法は少ない。本研究においては、[Rossi 15] で提案されたノードの役割を割り当てるフレームワークをリンクに対して用いることができるように、拡張したフレームワークを提示する。

#### 2.2 ネットワーク表現学習

自然言語処理 [Mikolov 13] を端緒としたネットワーク表現学習は、現在多数のアルゴリズムが提案されている [Perozzi 14][Grover 16]。ただし、多くのアルゴリズムが隣接しているノードを近くに埋め込む、homophily の性質に基づいており、ネットワークの大域的構造を保存することを目的とした手法は少ない。しかし、役割を探索する場合、ネットワーク上で距離が離れていても果たす役割が似ていれば、近くに埋め込まれる必要がある。グラフにおける役割に注目した手法として、struc2vec [Ribeiro 17] や、グラフ畳み込み手法 [Kipf 16] が提案されているが、本研究においては、属性情報やラベルがないことから、構造情報のみを利用する struc2vec を用いる。

連絡先: 菊田 俊平, 東京大学工学系研究科, kikuta@torilab.net

さらに、本研究においてはリンクの表現を得る必要があるが、リンクの表現学習の手法は少なく、既存手法 [Abu-El-Haija 17] はリンク予測に特化している。また、リンクの分散表現は、二つのノードの分散表現間の二項演算によって取得し、リンク予測などのタスクに用いられることがあるが、明示的にリンクの表現を得ておらず、有向グラフである場合、非対称な関係性を適切に埋め込むことはできない。そこで、本研究においては、ノードの手法をリンクに用いるためにグラフを変形する手法を導入する。

### 3. 提案手法

本章では、リンクの役割を発見するため、以下の汎用的なフレームワークを導入する。

- グラフ変形: リンクとノードの役割を反転させる
- 表現学習: ノードの分散表現を自動的に得る
- クラスタリング: 分散表現に応じてリンク群に分割する
- 役割割当: リンク群に対して適切なラベルを貼る

本フレームワークは目的や制約に応じて、柔軟に手法を入れ替えることができる性質を持つ。例えば、本研究ではクラスタリングの際にハードクラスタリングを用いるが、ソフトクラスタリングの手法も同様に用いることができる。

#### 3.1 グラフ変形

本研究では、ノード用のネットワーク表現学習をリンクに対して用いるために、辺双対グラフを用いる。古典的なグラフ理論に起因する手法であり、脳科学 [Bogdanov 14] や、現在のグラフ理論 [Yuan 17] においても利用されている。辺双対グラフの作成方法は以下の通りである。

$$G' = (\{e' | e \in E\}, \{e'_1 e'_2 | e_1, e_2 \in E, e_1 \text{ と } e_2 \text{ は隣接する}\}) \quad (1)$$

ここで、 $E$  は元のグラフにおけるリンクのリスト、 $e_1$  と  $e_2$  が隣接するとは、ノードを共有することを示す。図 1 に、辺双対グラフの例を示す。元のグラフである [1] のリンクとノードを入れ替えた辺双対グラフが [2] である。

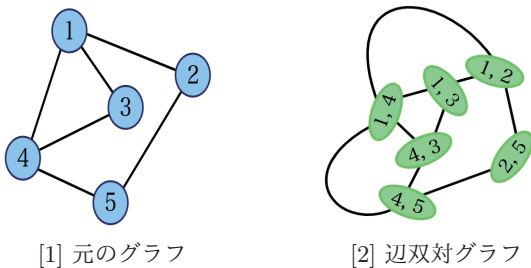


図 1: 辺双対グラフ作成過程

#### 3.2 表現学習

本研究において、分散表現を得る手法として struc2vec [Ribeiro 17] を用いる。struc2vec は、ネットワーク上で位置に関係なく、構造的に似ている二つのノードを潜在空間上で近くに、似ていない二つのノードを遠くに埋め込む表現学習の手法である。struc2vec では、二つのノ

ード  $u, v$  の構造的類似度  $f_k(u, v)$  を計算するために、以下の式を用いる。

$$f_k(u, v) = f_{k-1}(u, v) + g(s(R_k(u)), s(R_k(v))) \quad (2)$$

$$k \geq 0, |R_k(u)|, |R_k(v)| > 0$$

ここで、 $k$  はノードから  $k$  ホップ先を示し、 $R_k(u)$  は、ノード  $u$  の  $k$  ホップ先のノード群である。 $s(R_k(u))$  は、 $R_k(u)$  の次数集合を昇順に並び替えた配列であり、 $g$  は、配列同士の距離を測定する関数である。struc2vec では、関数  $g$  として、動的時間伸縮法 [Rakthanmanon 13] を用いており、以下の式で表される。

$$d(a, b) = \frac{\max(a, b)}{\min(a, b)} + 1 \quad (3)$$

ここで、 $a, b$  は配列  $A, B$  の要素であり、 $d$  の和が最小となるように組み合わせる。ノード  $u, v$  は構造的距離が近いと、配列同士の類似性が大きくなるため、 $d(a, b)$  は小さくなり、 $f$  も小さくなる。次に、 $k^*$  層から成る多層グラフを作成するが、ノード間の重みを以下のように定義する。

$$w_k(u, v) = e^{-f_k(u, v)} \quad (4)$$

$$k = 0, \dots, k^* - 1$$

ここで  $k$  は  $k$  層目であることを示し、 $w$  はリンクの重みを表す。また、同一のノードの上下の層とのリンクの重みを以下のように定義する。

$$w(u_k, u_{k+1}) = \log(\Gamma_k(u) + e)$$

$$w(u_k, u_{k-1}) = 1$$

$$\Gamma_k(u) = \sum_{v \in V} \mathbb{1}(w_k(u, v) > \bar{w}_k)$$

$$k = 0, \dots, k^* - 1 \quad (5)$$

得られた多層グラフ上をノード  $u$  を始点とするランダムウォークすることを考える。確率  $q$  で、現在の層上を遷移し、確率  $1 - q$  で上下の層における同一ノードに遷移する。第  $k$  層目における遷移確率は、式 (4)(5) で得た重みを用いて、以下のように定義する。

$$p_k(u, v) = \frac{e^{-f_k(u, v)}}{Z_k(u)}$$

$$Z_k(u) = \sum_{\substack{v \in V \\ v \neq u}} e^{-f_k(u, v)} \quad (6)$$

$Z_k(u)$  は確率とするための正規化項である。

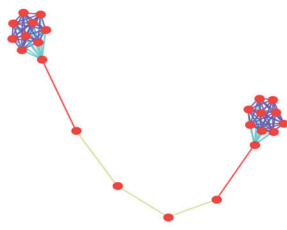
$$p(u_k, u_{k+1}) = \frac{w(u_k, u_{k+1})}{w(u_k, u_{k+1}) + w(u_k, u_{k-1})}$$

$$p(u_k, u_{k-1}) = 1 - p(u_k, u_{k+1}) \quad (7)$$

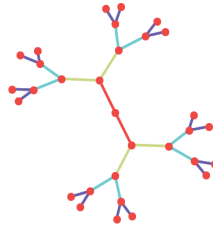
ランダムウォークによって類似する構造を持つノードに遷移しやすいため、ノード  $u$  の文脈はノード  $u$  に類似するノードによって構成される。文脈が与えられた上で、分散表現を得る際に Skip-gram [Mikolov 13] を用いる。最適化手法として階層型ソフトマックスを用いる。

#### 3.3 クラスタリング

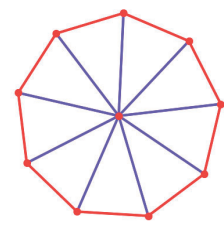
役割数  $r$  が与えられた前提で、分散表現から類似性を持つ  $r$  個に分ける手法として、非階層型クラスタリング手法である  $k$ -means 法 [Hartigan 79] を用いる。ここで、struc2vec の性質によって、構造的に類似しているリンクは近い表現を持つため、同じクラスターに帰属していると考えられる。



[1] Barbell graph



[2] 階層型グラフ



[3] リング・スター構造

図 2: 可視化の結果

### 3.4 役割割当

同様の役割を持つリンク群に分割した後、それらに役割のラベルを与える必要がある。本研究では、役割を割り当てる手法を導入する。クラスターにおける、ネットワーク指標の代表値からクラスターを特徴付ける手法であり、各クラスターに対して、双対グラフ上の次数, pagerank, 媒介中心性, 近接中心性, クラスター係数を計算し、それらに有意な差が得られれば、それをクラスターの特徴と考え、役割を割り当てる。

## 4. 提案手法の評価と応用

本研究における実験のスキームは以下の通りである。まず、類似した構造を持つリンクが近い表現を持つことを確認するために、可視化を行う。その結果を考察し、クラスターリングから得られる結果は有意義であることを確認した上で、役割発見のタスクを解く。

### 4.1 ベンチマークネットワークの可視化

本節では、(4, 10)-Barbell グラフ, 階層型グラフ, リング・スター構造グラフに対して、提案手法を適用し、類似した構造を持つリンクが近い表現を持っていることを確認する。

ここで、(n, k)-Barnell グラフとは、n 個のノードを用いて完全グラフを二つ作成し、それらが k 個のノードによって媒介されるグラフである。また、階層型グラフとして、深さ 4 の完全二分木, リング・スター構造としてスター構造の周りをリング状にリンクで囲ったグラフを用いる。階層型, リング・スター構造グラフは、4, 2 本の構造的に同等なリンクを持つため、クラスター数を 4, 2 と定めた。Barbell グラフにおいては、構造的に一致したリンクは 5 本であるが、構造的に類似しているが一致はしていないリンクが近い表現を持つことを確認するため、クラスター数を 4 に設定した。

元のグラフを構造に基づいてリンクを色分けした結果は図 2 の通りである。barbell グラフにおいて、構造的に同等なリンク群 (紫, 緑, 赤) は同じクラスターに所属し、類似した構造を持つリンク群 (黄) も同じクラスターに所属している。さらに、階層型グラフ, リング・スター型グラフにおいても、構造的に同等なリンク群は同じクラスターに所属していることが確認できる。したがって、可視化のタスクによって、類似した構造を持つリンクは近い表現を持つことが分かる。

### 4.2 リンクの役割発見

本節では、第三章で導入した手法を用いて、リンクの役割発見のタスクを行う。データセットは Zachary's Karate Club[Zachary 77] を用いる。ノードが 34 個, リンクが 78 本の重みなし無向グラフである。各ノードはクラブ内のメンバーを表し、リンクはメンバー同士の結びつきを表す。クラスターの個数は 4 に設定した。

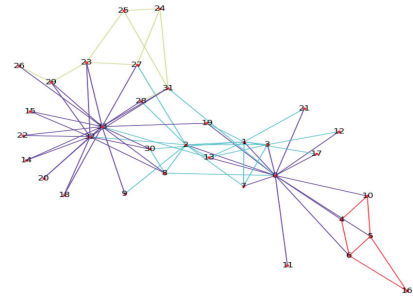


図 3: Karate Club 可視化の結果

クラスターリングを可視化した結果は図 3 の通りであり、また各クラスターのネットワーク指標の代表値は表 1 の通りである。クラスター 1 は、中心性の指標が大きく、またクラスター係数も大きいことから、コミュニティの中心からコミュニティ内の他のノードへ貼られるリンク (紫) と推測できる。また、クラスター 2 は、1 よりも媒介中心性が大きく、その他の指標が 1 よりも小さいことから、コミュニティ間のブリッジの役割を果たすリンク (水色) と推測できる。さらに、クラスター 3, 4 は、クラスター係数が小さい特徴を持ち、中心性指標は全て小さい。したがって、コミュニティ内で末端のリンクであると推測できる。さらに、リンクの構造からリンクの性質が理解できることを示す。Karate Club のデータセットは、二人の指導者によって分裂したクラブにおいて、メンバーが半分ずつ分裂したクラブに所属している。クラスター 1 は、コミュニティ内の中心的なリンクであることから、「リーダーからコミュニティ内のメンバーとの関係性」、クラスター 2 は、コミュニティ間のブリッジの役割を果たすことから、「双方のコミュニティへのつながり」、クラスター 3, 4 はリーダーとは直接かかわっておらず、媒介的でもないリンクであるといえる。すなわちグループ間をつなぐわけでも、リーダーとの関係を作るわけでもない周辺的な関係性であり、「コミュニティ内の味方同士の関係性」と解釈できる。以上より、本フレームワークによってリンクの役割をラベリングし、リンクの役割からリンクの性質が理解できる。

## 5. おわりに

本研究では、リンクの構造・役割に注目することで、リンクの性質を理解することを示した。また、リンクの役割発見に用いることができる柔軟なフレームワークを示し、その中で、辺双対グラフを用いることで有意義なリンクの分散表現を得られることを示した。今後取り組むこととして、データから役割の数を自動的に決定する手法がある。現時点では役割数  $r$  を既知としているが、それを自動的に求めるアルゴリズムが求めら



表 1: Zachary's Karate Club における各クラスターのネットワーク指標の代表値

label	betweenness		closeness		clustering_coefficient		degree		pagerank	
	mean	median	mean	median	mean	median	mean	median	mean	median
1	5.1	4.5	6.6	6.6	7.6	7.9	2.2	2.2	1.5	1.5
2	6.1	5.8	5.9	5.9	6.3	5.9	1.6	1.6	1.2	1.2
3	2.2	2.3	4.0	4.4	5.3	5.2	0.8	0.9	0.8	0.8
4	1.6	1.7	3.1	3.1	5.6	5.2	0.6	0.6	0.7	0.7
scale	$\times 10^{(-2)}$		$\times 10^{(-2)}$		$\times 10^{(-1)}$		$\times 10^{(-1)}$		$\times 10^{(-2)}$	

れる。また、得られたリンクの性質からノードへの理解を深めることができると考えられる。さらに、現実世界から得られたネットワークデータに本手法を適用し、役割が得られることを示す。

## 参考文献

- [Abu-El-Haija 17] Abu-El-Haija, S., Perozzi, B., and Al-Rfou, R.: Learning edge representations via low-rank asymmetric projections, in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1787–1796 ACM (2017)
- [Bogdanov 14] Bogdanov, P., Dereli, N., Bassett, D. S., Grafton, S. T., and Singh, A. K.: Learning about Learning: Human Brain Sub-Network Biomarkers in fMRI Data, *arXiv preprint arXiv:1407.5590* (2014)
- [Cui 18] Cui, P., Wang, X., Pei, J., and Zhu, W.: A survey on network embedding, *IEEE Transactions on Knowledge and Data Engineering* (2018)
- [Goyal 18] Goyal, P. and Ferrara, E.: Graph embedding techniques, applications, and performance: A survey, *Knowledge-Based Systems*, Vol. 151, pp. 78–94 (2018)
- [Grover 16] Grover, A. and Leskovec, J.: node2vec: Scalable feature learning for networks, in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864 ACM (2016)
- [Hartigan 79] Hartigan, J. A. and Wong, M. A.: Algorithm AS 136: A k-means clustering algorithm, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 28, No. 1, pp. 100–108 (1979)
- [Henderson 12] Henderson, K., Gallagher, B., Eliassirad, T., Tong, H., Basu, S., Akoglu, L., Koutra, D., Faloutsos, C., and Li, L.: Rolx: structural role extraction & mining in large graphs, in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1231–1239 ACM (2012)
- [Kipf 16] Kipf, T. N. and Welling, M.: Semi-supervised classification with graph convolutional networks, *arXiv preprint arXiv:1609.02907* (2016)
- [Mikolov 13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J.: Distributed representations of words and phrases and their compositionality, in *Advances in neural information processing systems*, pp. 3111–3119 (2013)
- [Perozzi 14] Perozzi, B., Al-Rfou, R., and Skiena, S.: Deepwalk: Online learning of social representations, in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 701–710 ACM (2014)
- [Rakthanmanon 13] Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., and Keogh, E.: Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping, *ACM Transactions on Knowledge Discovery from Data (TKDD)*, Vol. 7, No. 3, p. 10 (2013)
- [Ribeiro 17] Ribeiro, L. F., Saverese, P. H., and Figueiredo, D. R.: 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 ACM (2017)
- [Rossi 15] Rossi, R. A. and Ahmed, N. K.: Role discovery in networks, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 27, No. 4, pp. 1112–1131 (2015)
- [White 76] White, H. C., Boorman, S. A., and Breiger, R. L.: Social structure from multiple networks. I. Blockmodels of roles and positions, *American journal of sociology*, Vol. 81, No. 4, pp. 730–780 (1976)
- [Yuan 17] Yuan, W., He, K., Guan, D., and Han, G.: Edge-dual graph preserving sign prediction for signed social networks, *IEEE Access*, Vol. 5, pp. 19383–19392 (2017)
- [Zachary 77] Zachary, W. W.: An information flow model for conflict and fission in small groups, *Journal of anthropological research*, Vol. 33, No. 4, pp. 452–473 (1977)