

# 異質なコミュニティをつなぐノードの評価指標

山野 泰子<sup>1,a)</sup> 坂田 一郎<sup>1,2,b)</sup>

**概要:** 従来ネットワーク分析におけるノードの重要性は、主にそのノードによって得られる情報量や、隣接関係に基づく媒介性からの議論が多くなされてきた。隣接関係に基づくノード評価には、主としてローカルなノード周辺情報が利用されてきたのに対し、ネットワークの重要な特徴である、コミュニティ構造はほとんど考慮されてきていない。一方で、現実のネットワークでは、閉じたコミュニティ内部での情報交換を繰り返すことで、同種の意見が増幅、強化されるエコチェンバー現象など、コミュニティ構造に起因する情報の偏りを踏まえた解析の必要性が高まっている。そこで本研究では、コミュニティの質的な差異に基づいて情報の希少性を定量化し、ノード評価に活用することを提案する。具体的には、共通隣接ノードからコミュニティ関連度を求め、ノードの持つ網羅性と掛け合わせることで、異質なコミュニティをつなぐノードを識別する指標である PW 値（重み付き P 値）を考案した。

## 1. 研究背景及び目的

広告や商品推薦をはじめ、通常我々が目にするデジタル情報は過去の検索や閲覧履歴等に基づいて推定され自動的に決定されている。ネット空間にはこの他にも、本人が気に入りに入れたリンク先や発信者リストからの情報をはじめ、選別された特定種類の情報が続けて届く仕掛けが随所にある。キュレーションの時代とも言える一方で、複眼的思考を妨げるフィルターバブルが起りやすいことにも注意が必要である。好みの情報にアクセスし、好みに合いそうな情報を推薦する仕組みの中では、自分やその所属コミュニティとは異なる視点を取り入れることが従来以上に難しくなっている課題が顕在化してきている。

ネットワークの構造的な空隙を測る指標である Burt の制約指標 [1] は、対象となるノードの次数と共通隣接ノード数が反映された値であり、コミュニティ差異は反映されていない。コミュニティ間リンク度を表すノード指標である P 値 [2] はノードのコミュニティ網羅性を評価する指標であり、コミュニティ間の差異を区別しているわけではない。

本研究では、こうした情報環境の課題および既存指標の限界を乗り越えるべく、ネットワーク内のノード評価にコミュニティ間リンクの視点を取り入れることを提案する。

その際、コミュニティをつなぐ網羅性に加え、どのようなコミュニティ間をつないでいるのかに着目し、関連度の低いコミュニティ間をつなぐノードであれば、代替の効かない希少なノードとして高く評価する。こうしたノードは異なる視点を合わせ持つ情報源とも捉えることができ、フィルターバブル回避に有用であることが想定される。

従来、複雑ネットワークにおけるノードの重要性は、主にそのノードによって得られる情報量や経路数に基づく中心性の観点から評価されてきた。また、ネットワークの構造予測には、ノードペアに加え共通隣接ノードなどの周辺リンク指標が用いられてきた。しかしこれらは、ノード周辺情報を参照してはいるものの、ネットワークを構成する複数コミュニティ間の関係性を捉えているとは言い難い。

そこで本研究ではコミュニティ差異を定量化し、異質なコミュニティ間をつなぐノードを評価する指標を作成することを目的とした。

## 2. 関連研究

### 2.1 構造的な空隙を捉える尺度

「弱い紐帯」は、頻繁に接する人々よりも、接する機会の少ない人々によって、転職に有用な情報を得られる確率が高いという調査結果から提唱された概念である [3]。ここでの「弱さ」の尺度は接触頻度であり、枝の重さにあたるが、これをネットワークの構造的な位置づけによって捉え直したのが Burt の制約指標である [1]。ここでは枝の相関事象としての紐帯の弱さではなく、コミュニティ間をつなぐ構造が、希少な情報をもたらす「構造的な空隙」の橋渡しとして重視されている。

<sup>1</sup> 東京大学未来ビジョン研究センター  
Institute for Future Initiatives, The University of Tokyo  
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033

<sup>2</sup> 東京大学イノベーション政策研究センター  
Innovation Policy Research Center, The University of Tokyo  
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033

a) yamano@ifi.u-tokyo.ac.jp

b) isakata@ipr-ctr.u-tokyo.ac.jp

構造的空隙を捉える尺度として提案されたのが、構造的空隙の欠如を表す「拘束度」である [4]. ローカルな拘束度は2つのノード間の相互情報量と、両ノードとその隣接ノードとの相互情報量の積和を足した値の二乗として定義される. ノード  $i$  に関するネットワーク内の全てのペアのローカルな拘束度を足し合わせた値は Burt の制約指標と呼ばれ、下記で定義される:

$$B_i = \sum_{j \in V_i, j \neq i} (p_{ij} + \sum_{q \in V_i, q \neq i, j} p_{iq} p_{qj})^2, \quad (1)$$

$$p_{ij} = \frac{a_{ij} + a_{ji}}{\sum_{k \in V_i, k \neq i} (a_{ik} + a_{ki})}. \quad (2)$$

$V_i$  はノード  $i$  の隣接ノード集合であり,  $a_{ij}$  はノード  $i$  からノード  $j$  への関係の強さを,  $p_{ij}$  は  $i, j$  間の関係量を  $i$  の持つ全関係量の総和で除した相互情報量を表す. 重みなし無向グラフの場合,  $p_{ij}$  は  $ij$  間にリンクがあれば  $1/(i$  の次数), なければ  $0$  となる. Burt の制約指標は対象となるノードの次数と共通隣接ノード数を反映し, 1パスおよび2パス先までのノード周囲情報に基づいて計算される.

## 2.2 コミュニティ内外のリンクとノードの役割

コミュニティ内外のリンクを区分し, その分布パターンをノードの役割分類に活用した指標がある [2]. 下記の式 (3)(4) で定義される, コミュニティ内次数の高さ (ハブ度) を表す  $Z$  値と, コミュニティ間リンク度 (コネクター度) を表す  $P$  値 (Participation coefficient) である,

$$Z_i = \frac{K_{S_i} - \bar{K}_{S_i}}{\sigma_{K_{S_i}}}, \quad (3)$$

$$P_i = 1 - \sum_{s=1}^{N_m} \left( \frac{K_{S_i}}{K_i} \right)^2. \quad (4)$$

$K_{S_i}$  はノード  $i$  のコミュニティ  $S_i$  内次数,  $\bar{K}_{S_i}$  はコミュニティ  $S_i$  の平均次数,  $\sigma$  はコミュニティ  $S_i$  内次数の標準偏差である.  $N_m$  はネットワーク内のコミュニティ数であり,  $K_i$  はノード  $i$  の総次数である.  $Z$  値は標準化されたノード  $i$  のコミュニティ内リンク度となる.  $P$  値はコミュニティ外リンクを持つほど大きくなり, コミュニティ内にしかリンクがなければ  $0$  となる.  $P$  値はノードのコミュニティ網羅性を評価する指標であり, コミュニティ間の差異を区別しているわけではない.

## 2.3 コミュニティの構造情報を用いたリンク予測

従来, リンク予測においてはノード類似度に基づく研究が盛んであったが, 近年ではコミュニティの構造情報がミッシング・リンクの予測に大きく寄与することが明らかになっている [5][6][7][8]. 最初の研究は, 同一コミュニティ内にあるノードペアは, 違うコミュニティにあるノードペアよりも類似度が高く, リンク確率が高いという仮説

に基づいて行われ, 予測精度の向上に成功している [5].

コミュニティ内の直接リンクのみを考慮した上記手法に対し, コミュニティ間の潜在的関連性を活用した予測手法が Com\_St である [8]. ここで予測に用いられた4つのコミュニティ関連度はいずれもローカルなノード類似度から求めたリンク予測結果よりも高い精度を達成したが, そのなかでも安定的に高い AUC を獲得した尺度は, 以下の式 (5) で定義される, コミュニティ間の共通隣接ノードから算出したジャガード係数 (Community Relevance Jaccard Coefficient: CRJC) である,

$$\text{CRJC}(c_i, c_j) = \frac{|\Gamma(c_i) \cup V(c_i) \cap \Gamma(c_j) \cup V(c_j)|}{|\Gamma(c_i) \cup V(c_i) \cup \Gamma(c_j) \cup V(c_j)|}. \quad (5)$$

$c_i, c_j$  は  $C = \{c_1, c_2, \dots, c_m\}$  で定義されるネットワーク内のコミュニティのいずれかの組み合わせを示しており,  $V(c_i) = \{v | v \in V, c_i \in C(v)\}$  はコミュニティ  $c_i$  内の全てのノードである.  $V$  はネットワークの全てのノード,  $C(v)$  は  $v$  を含む全てのコミュニティである.  $\Gamma(c_i)$  は  $c_i$  の隣接ノードであり,  $c_i$  内の全ノードの隣接ノードのユニーク集合となっている. CRJC の分子はコミュニティ単位での共通隣接ノード数であり, 分母は2つのコミュニティを構成するノードおよびその隣接ノードの和集合である.

## 2.4 本研究の位置づけ

本節では, ネットワークの構造的空隙を埋めるノードを検出する新たな指標を提案する. この際, 既存研究では活用されてこなかったコミュニティの構造情報を利用する. 近年試みられているコミュニティ関連度を用いたリンク予測は, 関連性の高いコミュニティであればリンクする確率が高いと考える点で, ノード類似度に基づいて同類を推薦する従来からの予測手法と設計思想が同じである. しかしこうした手法では, 弱い紐帯や構造的空隙といった概念で重視されてきた希少な情報をもたらすノードは検出されない. そこで本研究では異質性をもたらすコミュニティ間構造に着目し, ノード評価に活用することを提案する.

## 3. コミュニティ異質性を反映したノード指標

本研究では, 共通隣接ノードからコミュニティ関連度 [8] を求め, その逆数の対数をコミュニティ網羅性を表す  $P$  値 [2] と掛け合わせることで, コミュニティ差異を反映させた重み付き  $P$  値を提案する. 定義式は下記である:

$$PW_i = -P_i \log \sum_{j \in \Gamma(i)^{IC}, j \neq i} \frac{\text{CRJC}(c_i, c_j)}{L + \delta}. \quad (6)$$

$P_i$  は式 (4) で定義された  $P$  値であり,  $P_i$  の重みは, 式 (5) で示したコミュニティ関連度の平均の対数を負にした値である.  $\Gamma(i)^{IC}$  はノード  $i$  の隣接ノードのうち, ノード  $i$

の所属クラス以外に所属している隣接ノードを、 $L$  は  $\Gamma(i)^{LC}$  の数を表す。コミュニティ網羅性が高く、関連度の低いコミュニティをつないでいるノードであるほど、 $PW_i$  は大きくなる。以下に計算手順を示す。なお、計算過程で得られた行列は全てグラフのノード順に対応させている。

### 3.1 隣接コミュニティ行列

モジュラリティベースのコミュニティ分割手法 [9] により、全ノードの所属コミュニティ ID を取得する。隣接行列中、所属コミュニティの外にリンクが存在する要素位置に、隣接ノードのコミュニティ ID が入った行列を作成し、コミュニティ ID をコミュニティ関連度に置き換える。

### 3.2 非関連度と重み付き P 値

ノード毎にコミュニティ関連度の平均値を取得し、その対数を負にした値を非関連度とする。平均コミュニティ関連度が 0 の場合は、0 を 1 に置き換えた値を非関連度とする。計算式 (4) に基づき、隣接コミュニティ行列を用いてノード毎に P 値を取得する。P 値に非関連度を掛け、ノード毎の重み付き P 値を取得する。

## 4. 実験

### 4.1 データセット

コミュニティ構造を反映したノード指標である Z, P, PW 値の特性を比較検証するため、コミュニティ数とリンク率の異なるネットワークを生成し定量的評価を行った。なお、実験では次数による平均最短経路長への影響を抑えるため、次数分布がいずれも一様なネットワークを使用した。

RC(Relaxed Caveman) ネットワーク [10] では、コミュニティ内ノード数を 8, rewiring 率を 0.2 に設定し、コミュニティ数の異なる 11 個のネットワークを生成した。LFR(Lancichinetti-Fortunato-Radicchi) ネットワーク [11] では、ノード数、各ノードの次数、コミュニティ内ノード数をそれぞれ 100, 8, 10 に設定し、コミュニティ間リンク率の異なる 6 つのネットワークを生成した。

### 4.2 実験内容と結果

#### ノード削除後の平均最短経路長の順位相関

RC ネットワークを用いて、ノードを順に 1 つだけ削除した後の平均最短経路長を計算し、平均最短経路長の長い順にノード ID を並び替えて標準順位とした。さらに、Z 値、P 値、PW 値の 3 つの指標での評価順にノードを 1 つずつ削除した際の平均最短経路長を計算し、標準順位のノード順との違いを Spearman の順位相関係数で比較した。

図 1 は、左から順に 2 個、3 個、12 個のコミュニティを持つ RC ネットワークでのノード削除実験の結果である。図の横軸は Z 値、P 値、PW 値でのノード順位、縦軸は標

準順位である。Z 値、P 値、PW 値の順位はそれぞれ緑、青、赤色のドットで示している。Z 値は 3 つのネットワークのいずれにおいてもランダムな分布を示し、標準順位との相関がなかった。コミュニティ数が 2 の場合は、P 値と PW 値いずれにおいても青と赤のドットが重なっていた。コミュニティ数が 3 の場合は、中間の順位においては P 値と PW 値で差異があったが、下位と上位では P 値と PW 値の順位が同じであった。コミュニティ数が 12 の場合は、下位中位上位のいずれにおいても P 値と PW 値の順位に差異があり、PW 値が P 値より高い順位相関を示した。

表 1 表 2 は、コミュニティ数 cn の異なる RC ネットワークとコミュニティ間リンク率  $\mu$  の異なる LFR ネットワークを生成し、同様に順位相関を比較した結果である。Z 値はいずれのネットワークにおいても順位相関が低かった。全てのネットワークにおいて PW 値の順位相関が同じか高かった。また、LFR ネットワークではコミュニティ間リンク率が上がるほど、順位相関が低くなった。

表 1 コミュニティ数の異なる RC ネットワークでの順位相関

cn	Z	P	PW
2	-0.141	<b>0.850</b>	<b>0.850</b>
3	-0.148	0.827	<b>0.838</b>
4	0.048	0.507	<b>0.733</b>
5	-0.174	0.676	<b>0.825</b>
6	-0.194	0.663	<b>0.823</b>
7	-0.208	0.792	<b>0.863</b>
8	0.006	0.727	<b>0.792</b>
9	-0.119	0.737	<b>0.864</b>
10	-0.068	0.785	<b>0.880</b>
11	-0.022	0.711	<b>0.846</b>
12	-0.089	0.763	<b>0.864</b>

表 2 リンク率の異なる LFR ネットワークでの順位相関

$\mu$	Z	P	PW
0.1	-0.424	0.617	<b>0.895</b>
0.2	-0.569	0.596	<b>0.838</b>
0.3	-0.435	0.549	<b>0.766</b>
0.4	-0.357	0.404	<b>0.504</b>
0.5	0.019	0.258	<b>0.280</b>
0.6	-0.021	0.082	<b>0.163</b>

## 5. 考察

弱い紐帯や構造的空隙で提示されたコミュニティ間リンクの重要性はそのリンクがもたらす情報の希少性にある。本研究では、情報の希少性をコミュニティ間の共通隣接ノードの少なさとして捉え直し、ノード評価に活用した。共通隣接ノードが少なく、関連度の低いコミュニティ同士を結びリンクを持つノードを削除した場合、平均最短経路長は長くなることが予想される。なお、媒介中心性が高い

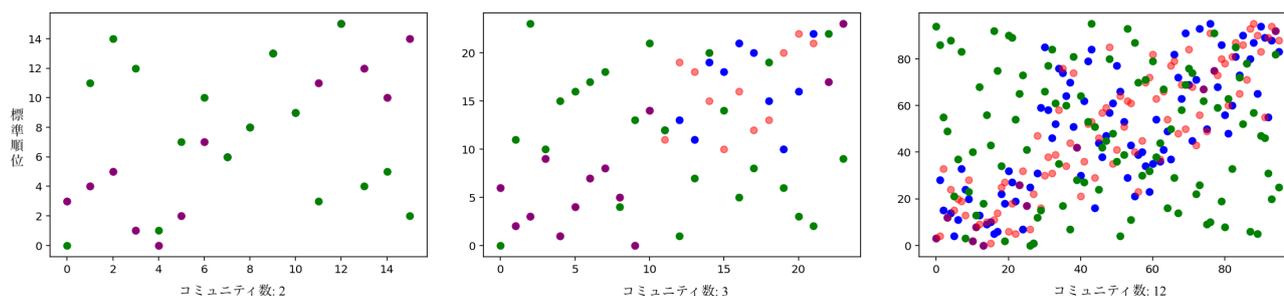


図 1 ノード削除時の平均最短経路長の指標別順位 (Z 値：緑, P 値：青, PW 値：赤色)

ノードを削除しても平均経路長は長くなるが、媒介中心性が反映するのはノードを経由する経路数であり、コミュニティ構造ではないため、今回の比較対象からは除外した。

図 1 でコミュニティ数が 2 の場合に PW 値と P 値の順位が等しくなったのは、コミュニティ関連度が 1 つしかなく、PW 値が P 値に同じ重みを掛けた値となったためである。これに対し、コミュニティ数が 3 の場合に中間順位で P 値と PW 値が一致しなかったのは、コミュニティ関連度が 3 つになり差異が生じたためと考えられる。ただし、下位では P 値と PW 値の順位が重なり、相関が見られなかった。これは全てのリンクがコミュニティ内にしか存在しないノードは P 値が 0 であり、その結果 PW 値も 0 になって順位に意味がなくなるためである。上位ノードで P 値と PW 値が同じか近くなったのは、これらのノードが 3 つのコミュニティ全てにリンクを持ち関連度の差が小さくなったためと考えられる。コミュニティ数が 12 の場合は、コミュニティ外リンクの割合が相対的に高くなり、P 値と PW 値の順位は一致しなかった。

表 2 で、コミュニティ間リンク率が上がるほど順位相関が低くなったのは、P 値、PW 値がともにコミュニティ間リンクを評価する指標だからだと考えられる。コミュニティ間リンク率が低いほどコミュニティ間差異は大きくなるため、PW 値は P 値より優位であった。コミュニティ外リンクが増え、コミュニティ構造が消失するリンク率 0.5, 0.6 では、P 値、PW 値ともに標準順位との相関がなかった。

コミュニティ数とリンク率の違う 17 種類のネットワーク全てにおいて、PW 値が P 値の順位相関以上の値を示したのは、PW 値が関連度の低いコミュニティをつなぐノードを高く評価しているためと考えられる。これにより、PW 値は他に代替の効かない希少なノード、すなわちそのノードを介さないと到達できない経路を、P 値よりも正確に識別し評価していることが伺えた。

## 6. 結論

本研究の目的は、コミュニティ差異を踏まえたノード指標を作成し、その有効性を検証することであった。まず、共通隣接ノードからコミュニティ関連度を求め、ノードの

持つ網羅性と掛け合わせることで、新たなノード指標 PW 値を作成した。次に、平均最短経路長の順位比較により、P 値と PW 値で順位に差異が出る範囲が確認された。さらに、順位相関係数の比較分析を行い、コミュニティ数およびコミュニティ間リンク率の異なる全てのネットワークにおいて、PW 値が P 値より相関が高く、PW 値が希少なリンクの識別向上に寄与したことが検証された。

謝辞 本研究は JSPS 科研費 19K01884 の助成を受けたものである。

## 参考文献

- [1] R.S. Burt. *Structural holes: The social structure of competition*. Harvard university press, 2009.
- [2] R Guimera and LAN Amaral. Functional cartography of complex metabolic networks. *nature*, Vol. 433, No. 7028, p. 895, 2005.
- [3] M.S. Granovetter. The strength of weak ties. In *Social networks*, pp. 347–367. Elsevier, 1977.
- [4] R.S. Burt. Structural holes and good ideas. *American journal of sociology*, Vol. 110, No. 2, pp. 349–399, 2004.
- [5] B. Yan and S. Gregory. Finding missing edges in networks based on their community structure. *Physical Review E*, Vol. 85, No. 5, p. 056112, 2012.
- [6] C.V. Cannistraci et al. From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks. *Scientific reports*, Vol. 3, p. 1613, 2013.
- [7] L. Lü et al. Toward link predictability of complex networks. *Proceedings of the National Academy of Sciences*, Vol. 112, No. 8, pp. 2325–2330, 2015.
- [8] J. Ding et al. Prediction of missing links based on community relevance and ruler inference. *Knowledge-Based Systems*, Vol. 98, pp. 200–215, 2016.
- [9] 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.
- [10] DJ Watts. *Small worlds: the dynamics of networks between order and randomness*, Vol. 9. Princeton university press, 2004.
- [11] A. Lancichinetti and S. Fortunato. Community detection algorithms: a comparative analysis. *Physical review E*, Vol. 80, No. 5, p. 056117, 2009.