5J-05

# ノード視点によるコミュニティ抽出手法のマルチレイヤネットワークへの拡張

盛 燎平 †　須貝 康雄 †
千葉大学大学院工学研究科 †

## 1 はじめに

ネットワークからコミュニティと呼ばれるリンクの繋がりが密な部分を抽出する研究が行われている。コミュニティ抽出手法の1つにCDB法がある[1)]。CDB法は任意のノードから局所的なネットワーク構造のみを用いてコミュニティを抽出できる。また，マルチレイヤネットワークと呼ばれる複数のネットワークを組み合わせてノード間の関係性を表すネットワークからのコミュニティ抽出手法も研究されているが，局所的なネットワーク構造のみからコミュニティを抽出する手法は少ない。そこで本稿では，CDB法をマルチレイヤネットワークへ拡張し，計算機実験により抽出したコミュニティを評価する。

## 2 関連研究

### 2.1 CDB法

CDB法は，ネットワークの任意のノードからコミュニティ濃度と呼ばれる評価値により隣接ノードをコミュニティに取り込む局所的コミュニティ抽出手法である。ネットワーク $G$ の部分グラフを $G_s$，$G_s$ 内のノード集合を $V_s$，$G_s$ 内のリンク集合を $E_s$，$V_s$ の平均次数を $k_{G_s}$，グラフ $G' = (V_s, E \backslash E_s)$ におけるノード $v_s \in V_s$ の平均次数を $k'_{G_s}$，$V_s$ と $V_s$ 外のノードを繋ぐリンク集合を $E_b$ とする時コミュニティ濃度 $Cd$ は式 (1) と定義される。

$$Cd = \frac{k_{G_s}}{k'_{G_s}} = \frac{2 \times |E_s|}{E_b} \tag{1}$$

$Cd > 1$ を満たす時，部分グラフは周囲に比べてリンクが密であるため，この時の $G_s$ をコミュニティと定義している。なお $Cd$ はリンク数の代わりにリンクの重みの総和を利用することにより，重み付きグラフに対応できる。

コミュニティ内のコミュニティ外のノードと隣接するノード（コミュニティ内境界ノード集合 $V_b^{\mathrm{in}}$）に注目したコミュニティ濃度 $Cd^{\mathrm{in}}$ が提案されている。$Cd^{\mathrm{in}}$ は式 (2) と定義されている。

$$Cd^{\mathrm{in}} = \frac{2 \times |E_s|/|V_s|}{|E_b|/|V_b^{\mathrm{in}}|} \tag{2}$$

CDB法のアルゴリズムを説明する。初めにコミュニティ抽出の視点となるシードノードを選択し，$G_s$ に追加する。次に $V_s$ と連結している $G_s$ 外のノードをコミュニティに追加したと仮定した時の $Cd$ を計算し，$Cd$ が最大となるノードを $G_s$ に追加する。これを $G_s$ に隣接するノードが無くなるまで繰り返す。

### 2.2 特許引用ネットワーク

文献[2)]では米国特許の内1963年1月から1999年1月までに取得された特許の取得年，出願年，譲受人，技術分野のクラスなどの情報と1975年から1999年に構築された特許の引用関係をまとめている。本稿ではこのデータセットの内，特許が分類されているクラス名がElectrical computers and digital processing systemsにより始まるクラスに属する特許の内，1975年から1995年に付与された特許を抽出して，ノードを特許，リンクを引用関係とする無向グラフ（以下ECDPS）として実験に用いる。ECDPSのノード数は9910，リンク数は54293である。

### 2.3 マルチレイヤネットワーク

マルチレイヤネットワークを対 $M = (G, C)$ と定義する。ここで，$G = \{G^{[\alpha]}; \alpha \in \{1, \ldots, k\}\}$ はグラフ $G^{[\alpha]} = (V^{[\alpha]}, E^{[\alpha]})$ の族，$C = \{E^{[\alpha\beta]} \subseteq V^{[\alpha]} \times V^{[\beta]}; \alpha, \beta \subseteq \{1, \ldots, k\}, \alpha \neq \beta\}$ は異なるレイヤ $G^{[\alpha]}$ と $G^{[\beta]}$ 間を繋ぐリンク集合，$V^{[\alpha]} = \{v_1^{\alpha}, \ldots, v_{N_\alpha}^{\alpha}\}$ はレイヤ $G^{[\alpha]}$ のノード集合，$E^{[\alpha]} = \{e_1^{\alpha}, \ldots, e_{L_\alpha}^{\alpha}\}$ はレイヤ $G^{[\alpha]}$ のリンク集合とする。

## 3 提案手法

CDB法をマルチレイヤネットワークに対応させるために $Cd^{\mathrm{in}}$ の拡張を行う。各レイヤの隣接構造を考慮すると $Cd^{\mathrm{in}}$ は式 (3) と表す事ができる。

$$Cd^{\mathrm{in}} = \frac{2 \times \sum_{\alpha \in G} |E_s^{[\alpha]}|/|V_s|}{\sum_{\alpha \in G} |E_b^{[\alpha]}|/|V_b^{\mathrm{in}}|} \tag{3}$$

またマルチレイヤネットワークではレイヤ毎に重要度が異なる場合がある。そこでレイヤ $G_\alpha$ に対する重みを $w^{[\alpha]}$ とすると $Cd^{\mathrm{in}}$ を式 (4) と表すことができる。

$$Cd^{\mathrm{in}} = \frac{2 \times \sum_{\alpha \in G} w^{[\alpha]} |E_s^{[\alpha]}|/|V_s|}{\sum_{\alpha \in G} w^{[\alpha]} |E_b^{[\alpha]}|/|V_b^{\mathrm{in}}|} \tag{4}$$

式 (3),(4) はレイヤ間を繋ぐリンク集合 $C$ を考慮していない。そこで形式的に $C$ とその両端のノードを集めた部分ネットワークをレイヤ $G^{[\mathrm{inter}]}$ とし，レイヤの1つとして扱う事により $C$ を考慮できる。この評価値を用いてCDB法と同様にコミュニティ抽出を行う。

## 4 計算機実験

ECDPSを図1のように譲受人毎のレイヤに分割し，マルチレイヤネットワークとした。図1にて，円は特許，実線は同じ譲受人の特許同士の引用関係，破線は譲受人が異なる特許の引用関係（$C$ に相当）を表している。こ

An extension of the community detection method based on view point of a node to multilayer networks
†Ryohei Mori, †Yasuo Sugai, Graduate School of Engineering, Chiba University



Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.

の様に分割したときレイヤ数 $|G| = 924$，$|C| = 48586$ となる。

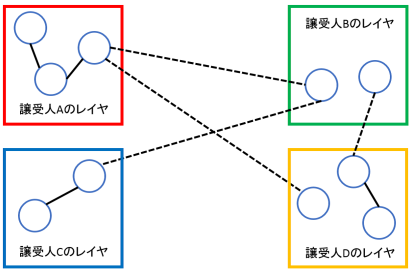


図 1: ECDPS を分割する時の模式図

譲受人に着目してコミュニティを抽出する場合を考える。譲受人が同じ特許は関係性が強く，異なる特許は関係性が弱いと考えられる。レイヤ毎の重み付けは，

**一様型**：$w^{[sa]} = 1$, $w^{[\mathrm{others}]} = 1$, $w^{[\mathrm{inter}]} = 1$

**増加型**：$w^{[sa]} = 1.1$, $w^{[\mathrm{others}]} = 1$, $w^{[\mathrm{inter}]} = 1$

**混在型**：$w^{[sa]} = 1.05$, $w^{[\mathrm{others}]} = 1$, $w^{[\mathrm{inter}]} = 0.95$

の 3 通りにより実験を行う。ここで，$G^{[sa]}$ はシードノードとなる特許の譲受人のレイヤ，$G^{[\mathrm{others}]}$ は $G^{[sa]}$ 以外のレイヤ，$G^{[\mathrm{inter}]}$ は $C$ とその両端のノードからなるレイヤである。なお，譲受人が存在しない特許及び引用関係は $G^{[\mathrm{inter}]}$ に追加した。

### 4.1 特許番号 5168569 をシードノードとした場合

ECDPS の次数の最頻値は 3 である。特許番号 5168569 は次数 3 のノードの内，離心中心性が最小である。すなわち一般的なノードの中でも最もネットワークの中心に近いノードである。特許 5168569 の譲受人は 211480 である。

$Cd^{\mathrm{in}}$ の推移を図 2 に，表 1 の 2〜4 行目に $Cd^{\mathrm{in}}$ が極大値になる時のコミュニティ内の譲受人 211480 の特許数 $N_{sa}$ と割合を示す。 シードノードの譲受人の割合が

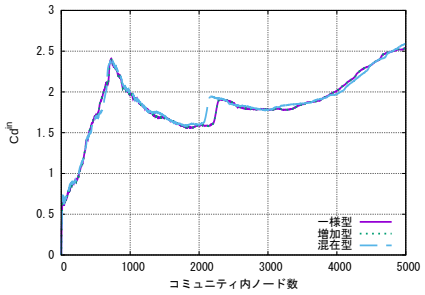


図 2: シードノードが特許 5168569 の場合の $Cd^{\mathrm{in}}$

向上することはなく，混在型では低下する結果となった。考えられる理由として，(i)$G^{[211480]}$ の連結成分が小さく，他のレイヤに移りやすい事 (ii) 特許 5168569 は譲受人 211480 の特許との引用関係が存在しない事が挙げられる。

### 4.2 譲受人 280070 からシードノードを選択する場合

(i) の検証を行う。譲受人 280070 は ECDPS 内で最大の連結成分を持つレイヤである。特許番号 5226156 をシードノードとして実験する。この特許は 5168569 と同様，次数 3 かつ同じ譲受人の特許との引用関係が存在しない。$Cd^{\mathrm{in}}$ が極大値になる時のコミュニティに含まれる譲受人 280070 の特許数 $N_{sa}$ と割合を表 1 の 5〜7 行目に示す。増加型と混在型いずれの場合にもシードノードの譲受人の特許の割合が増加しなかった。

### 4.3 シードノードがレイヤ $G^{[211480]}$ に存在する場合

(ii) の検証を行う。レイヤ $G^{[211480]}$ 内にリンクが存在する特許番号 4718006 をシードノードとして実験した。$Cd^{\mathrm{in}}$ の推移を図 3 に，譲受人 211480 の特許数 $N_{sa}$ と割合を表 1 の 8〜10 行目に示す。増加型の場合は 0.682%，混在型の場合は 0.489%の増加が見られた。このことから，重み付けを大きくしたレイヤ内にシードノードと隣接するノードがある場合，そのレイヤのノードの割合が大きくなる。また 4.1 節と 4.2 節の結果から，重み付けを大きくしたレイヤにシードノードと隣接するノードが無い，すなわちシードノードとそのレイヤの関係性が小さい場合，そのレイヤに含まれるノードがコミュニティに含まれる割合は増加しない。

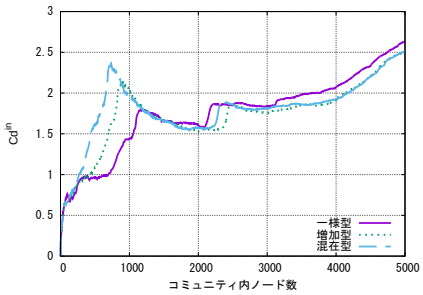


図 3: シードノードが特許 4718006 の場合の $Cd^{\mathrm{in}}$

表 1: シードノードが属する譲受人の割合

| 特許 | 重み付け | $\lvert V_s \rvert$ | $N_{sa}$ | 割合 [%] |
|---|---|---|---|---|
| 5168569 | 一様型 | 725 | 10 | 1.379 |
| | 増加型 | 725 | 10 | 1.379 |
| | 混在型 | 740 | 10 | 1.351 |
| 5226156 | 一様型 | 797 | 108 | 13.551 |
| | 増加型 | 806 | 109 | 13.524 |
| | 混在型 | 814 | 107 | 13.145 |
| 4718006 | 一様型 | 1217 | 17 | 1.367 |
| | 増加型 | 732 | 15 | 2.049 |
| | 混在型 | 916 | 17 | 1.856 |

## 5 おわりに

本稿ではノード視点によるコミュニティ抽出手法である CDB 法をマルチレイヤネットワークに拡張した。レイヤ毎に異なる重み付けを行い，抽出されたコミュニティ内のノードを比較した結果，シードノードが重み付けを大きい値にしたレイヤ内にリンクを持っている場合，そのレイヤのノード数がコミュニティに含まれる割合が増加した。これは，レイヤ視点によりネットワークからコミュニティを抽出できることを意味し，ある技術分野に注目した ECDPS からのコミュニティの抽出などに適応できる。今後の課題として，レイヤに付与する重みの考察が挙げられる。

### 参考文献

1) 高橋篤, 荒井幸代:“任意ノードから視点からのコミュニティ抽出手法”. 第 23 回人工知能学会全国大会論文集, 2009.

2) Bronwyn H Hall, Adam B Jaffe, Manuel Trajtenberg; “The nber patent citation data file: Lessons, insights and metholodogical tools”. Technical report, National Bureau of Economic Research, 2001.



Copyright ©2017 Information Processing Society of Japan.
All Rights Reserved.