

ODフローの有向性と重複を考慮した 機能地域検出手法の開発

佐藤 加斐¹・福本 潤也²・李 想³

^{1,3} 非会員 東北大学大学院博士前期課程 情報科学研究科 (〒980-8579 宮城県仙台市青葉区荒巻字青葉6-6-06)

E-mail:lixiang@plan.civil.tohoku.ac.jp

² 正会員 東北大学大学院准教授 情報科学研究科 (〒980-8579 宮城県仙台市青葉区荒巻字青葉6-6-06)

E-mail:fukumoto@plan.civil.tohoku.ac.jp

交通や経済取引等の相互作用を通じて、一体的・補完的に機能する複数の地理的単位から構成される機能地域 (FR: Functional Region) の検出手法がこれまでに地理学等で開発されてきた。既存手法の多くが抱える課題として、ODフローの有効性を明示的に考慮していない点と、機能地域間の重複を明示的に考慮していない点が挙げられる。本研究では、フローデータの非対称性とFRの重複を考慮できるFR設定手法を提案する。両者を考慮しない既存の機能地域検出手法の適用結果と比較することで、FR設定結果を通じた地域構造の理解に与える影響を実証的に明らかにする。

Key Words : *Functional Region, Modularity, Bipartite Graph, Overlapping Community Detection*

1. はじめに

機能地域 (Functional Region: FR) とは、交通や経済取引等の相互作用を通じて、一体的・補完的に機能する複数の地理的単位から構成される地域である。FRを設定し、各種統計をFR単位で集計することで社会経済活動の実態をより正確に把握することが可能となる。

既存のFR設定手法は、通勤や人口移動などの地域間相互作用を表すフローデータを入力データとして用いる。FR設定のアプローチの違いにより、ルールベース手法とアルゴリズム手法に大きく分かれる。ルールベース手法では、FRが満たすべき要件を定め、要件に従ってFRの設定ルールを定める。FRの要件やルール設定に分析者の恣意性が入りやすいが、人々の直観と比較的合致するFRが設定されやすい。他方、アルゴリズム手法はグラフ理論やクラスタ分析で開発されたアルゴリズムを用いてFRを設定する手法である。ルールベース手法より分析者の恣意性が入りにくい反面、設定結果がアルゴリズムの仮定に強く依存しやすい。

アルゴリズム手法に属する多くの既存手法には、フローデータの非対称性を十分考慮していない点と、機能地域の重複の可能性を除外している点において、課題があると考えられる。フローデータは一般に非対称的なデータである。しかし、多くのアルゴリズム

手法ではアルゴリズムを適用可能とするため、非対称的なフローを平均化して対称的なフローへと変形してから入力データとして用いている。代表的な機能地域の一つである通勤圏を設定する場合、郊外から都心へと向かう通勤と都心から郊外へと向かう通勤のうち、直感的には前者がより重要であると考えられる。しかし、既存手法では前者と後者を等しく取り扱っており、後者の大きさにも通勤圏の設定結果が依存してしまう。

また、地域間相互作用データは近接する地理的単位間にフローが集中するデータである。隣接する地理的単位間には必ずと言ってよいほど正のフローが存在する。FRの境界は必ずしも自明でなく、境界部に位置する地理的単位がどちらのFRに所属すべきか判断しかねる場合も少なくない。例えば、ある都心に複数方面からの通勤が集中しており、方面別に通勤圏が設定されるような場合、都心は方面別通勤圏のいずれにも所属すると考えるのが自然である。しかし、既存のFR設定手法では、1つの地理的単位が2つ以上の機能地域に所属しないという仮定を置いており、重複した通勤圏を検出することはできない。

本研究の目的は、フローデータの非対称性とFRの重複を考慮できるFR設定手法を提案し、両者を考慮しないことがFR設定結果に与える影響と、FR設定結果に基づく地域構造の理解に与える影響を実

証的に明らかにすることである。既存の FR 設定手法として、ネットワーク科学分野で提案されたモジュラリティ最大化法に基づく手法を取り上げる^{1),2)}。最初に、フローデータの非対称性の影響を明らかにするため、OD データをネットワークの重み行列とみなして 1 部ネットワークと 2 部ネットワークを生成する。それぞれのネットワークに対してモジュラリティ最大化に基づく FR 設定手法を適用し、2 つの結果を比較することでフローデータの非対称性の考慮の有無の影響を明らかにする。次に、FR の重複の影響を明らかにするため、ネットワーク科学分野で研究の蓄積が進んでいる重複を考慮したコミュニティ抽出法³⁾を援用して FR を設定する。重複の表現方法等が異なる 3 種類の手法を適用することで、重複を考慮することの影響と手法ごとの仮定の影響を出来るだけ区別する。最後に、2 部ネットワークに対して重複を考慮したコミュニティ抽出法を適用し、フローの非対称性と機能地域の重複をいずれも考慮することの有用性を示す。

2. 使用データと分析手法

(1) 使用データ

a) OD通勤データ

本研究では平成12年国勢調査における宮城県内の市区町村間通勤者数データを用いる。平成の大合併で誕生した市区町村を集計単位とすると、結果の解釈が難しいため、最新のデータを用いていない。

b) 二部ネットワークの生成

OD表を行列 T で表すと、二部ネットワークは以下の行列 $T2$ で表される。

$$T2 = \begin{bmatrix} \mathbf{O} & \mathbf{T} \\ \mathbf{T}^r & \mathbf{O} \end{bmatrix} \quad (1)$$

行列 $T2$ は対称行列であり、フローの対称性を仮定する任意のFR設定手法が適用可能である。二部ネットワークでは、分析対象範囲に含まれる全ての市区町村のフローの起点としての役割と終点としての役割がそれぞれノードとして表現される。

(2) 分析手法

本研究では、既存のFR設定手法としてモジュラリティ最大化に基づく手法^{1),2)}を用いる。重複を考慮するFR設定手法として、以下で説明するWang他⁴⁾、Chen他⁵⁾、Nepusz他⁶⁾のコミュニティ抽出法を援用する。

a) モジュラリティ最大化法

モジュラリティ最大化法は、代表的なコミュニティ抽出法の1つである。観測ネットワークと帰無ネットワークの重み行列を A と P で表す場合、次式で定義されるモジュラリティが最大となるように、

各ノードが所属するコミュニティを決定する。

$$Q = \sum_{i,j} (A_{ij} - P_{ij}) \delta(c_i, c_j) \quad (2)$$

ただし、 c_i はノード i の所属コミュニティを表すインデックスである。

b) Wang他の方法

同手法では、最初に既存のFR設定手法（本研究の場合、モジュラリティ最大化法）を用いて重複を考慮しないFR設定結果を求める。次に、求められた各FRについて、境界部の地理的単位の所属FRを1つずつ変更しながら式(3)の変化量を観察する。最後に、式(3)の変化量に基づいて境界部の地理的単位のFRへの帰属度 u_{ik} ($0 \leq u_{ik} \leq 1$) を設定する。 u_{ik} は地理的単位のFR k への帰属度を表し、式(4)を満たす。

$$S(k) = \sum_k^{in} / \left(\sum_k^{total} \right)^r \quad (3)$$

$$\forall i: \sum_k u_{ik} = 1 \quad (4)$$

式(3)の r はパラメータである。 $r=1$ の場合、式(3)はあるFRに所属する地理的単位を起終点とするフローの内、FR内の地理的単位を起終点とするフローの割合を意味する。式(3)が大きな値をとるほど、明確なFRが存在するとの前提に基づいてFRを設定している。

c) Chen他の方法

同手法ではゲーム理論の均衡問題として重複するコミュニティ抽出結果を求める。FR設定問題の文脈で解説すると、地理的単位をゲームのプレイヤー、所属するFR集合を戦略とみなす。地理的単位が複数のFRに所属することを戦略として認めることでFRの重複を表現している。モジュラリティの定義式から導出される関数を地理的単位の利得関数として定義することで、モジュラリティとほぼ等価な関数をポテンシャルとするポテンシャルゲームを定義できる。地理的単位が所属コミュニティ集合を局所的に変更する局所的最適応答プロセスを繰り返すことで、最終的なFR設定結果が得られる。

d) Nepusz他の方法

同手法では、全ての地理的単位について FR への帰属度 u_{ik} を 0 から 1 の実数で表し、式(5)を目的関数、式(4)を制約条件とする最小化問題を解くことで帰属度ベクトル $\{u_{ik}\}$ を求める。

$$f = \sum_{i,j} w_{ij} (\tilde{s}_{ij} - s_{ij})^2 \quad (5)$$

ただし、 w_{ij} はノード i, j の重要度を表すパラメータ、 \tilde{s}_{ij} は観測フローによって作成されるノード i, j の類似度である。 s_{ij} は式(6)で定義され、ノード i, j の類似度の推測値を意味する。

$$s_{ij} = \sum_k u_{ik} u_{jk} \quad (6)$$

本研究では、モジュラリティと類似した関数を用いて \tilde{s}_{ij} を設定することで、他の手法とFR設定結果の比較が行いやすいようにする。

3. 分析結果

(1) フローの非対称性を考慮することによる影響

モジュラリティ最大化を1部グラフと2部グラフに適用した結果を図1に示す。図1より、全体的な傾向は概ね同じだが、仙台市周辺の機能地域が異なることが分かる。1部グラフにモジュラリティ最大化法を適用した場合、太白区と泉区が別々の機能地域に所属するという直感から外れた結果が得られた。その理由は、泉区・太白区はいずれも青葉区への発生通勤量が大きいものの、両者間の通勤量が相対的に少ないためである。一方、2部グラフに適用した結果では、太白区と泉区のOノードがいずれも青葉区と同一の機能地域に所属しており、それぞれのDノードは異なる機能地域に所属している。上記結果は、太白区と泉区では、青葉区への発生通勤量が大きい一方で、郊外部の自治体からの集中通勤量も大きいという実態を特徴を捉えている。

(2) 重複を考慮することによる影響

重複を許容しないモジュラリティ最大化法による結果と重複を許容するChenの手法による結果を図2に示す。図2より、以下の2つが分かる。第一に、県北部において、重複を許容しない場合には個々の機能地域が相対的に大きいのに対し、許容する場合にはそれらが分割されて相対的に小さい機能地域が検出されていることである。分割された機能地域にはそれぞれ築館町・迫町、古川市・小牛田町という2つの中心があり、重複を許容することでそれぞれを中心とする機能地域が検出された。第二に、太白区・若林区・名取市の所属コミュニティが重複を許容する場合と許容しない場合で異なることが分かる。重複を許容しない場合、ある市区町村の所属コミュニティが決定すると、周辺市区町村の所属コミュニティがそれに引きずられて決まる。機能地域の境界部が不自然な区分結果となっている可能性を示唆していると解釈できる。

Nepuszらの手法による結果とChenらの手法による結果を図3に示す。図3より、両者の結果は比較的似ているが、他方、前者では複数の機能地域に所属する自治体が増加し、範囲の広い機能地域が検出されていることが分かる。この原因は、Nepuszらの手法では各市町村の機能地域への帰属度を0から1の実数で連続的に表しているのに対し、Chenらの手法では0か1のいずれかで表しているためである。

(3) 両者を同時に考慮することによる影響

1部グラフにChenの手法を適用した結果と2部グラフにChenの手法を適用した結果を図4に示す。2部グラフにおける結果に注目すると、流出・流入通勤量が多い仙台市近辺において機能地域が細分化され、

複数の機能地域に所属する自治体の数が増加していることが分かる。この原因の1つは、2部グラフを用いることで機能地域間の重複が一定程度考慮されているためである。ある市区町村のOノードとDノードが異なる機能地域に所属する場合、地理空間上では複数の機能地域が重複された分割結果が得られたと解釈することもできる。2部グラフを用いることで重複を一定程度反映されるにも関わらず、さらに、重複を考慮した機能地域区分手法を適用することで、過度の細分化が起きてしまった可能性がある。もう1つの原因は、上記とも関連するが、2部グラフではOノード間あるいはDノード間に結びつきの強さが定義されない点がある。通勤量の大きさで機能地域を区分する場合、OノードとDノードの間にはノードを結びつける力が働く一方、Oノード同士あるいはDノード同士を結びつける力や離反させる力が全く考えられていない。その結果、2部グラフに重複を考慮した機能地域区分結果が過度に細分化されたと考えられる。上記の課題を克服するには、2部グラフを用いた機能地域の定義式を再検討し、それに対応したアルゴリズムを開発する必要があると考えられる。

4. 結論

本研究では、2部グラフを用いてフローデータの

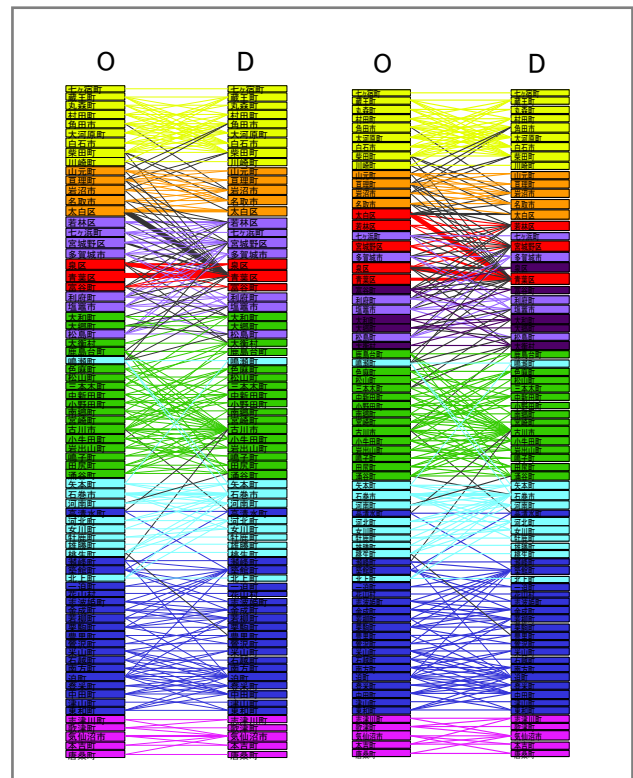


図1：1部グラフにおける結果(左)と2部グラフにおける結果(右)

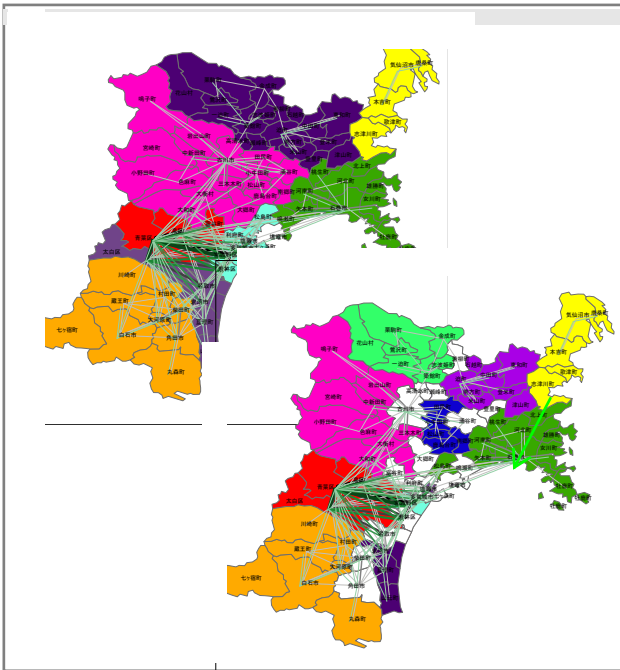


図 2 : モジュラリティ最大化法による結果(上)と Chen の手法による結果(下) (縞線部で重複)

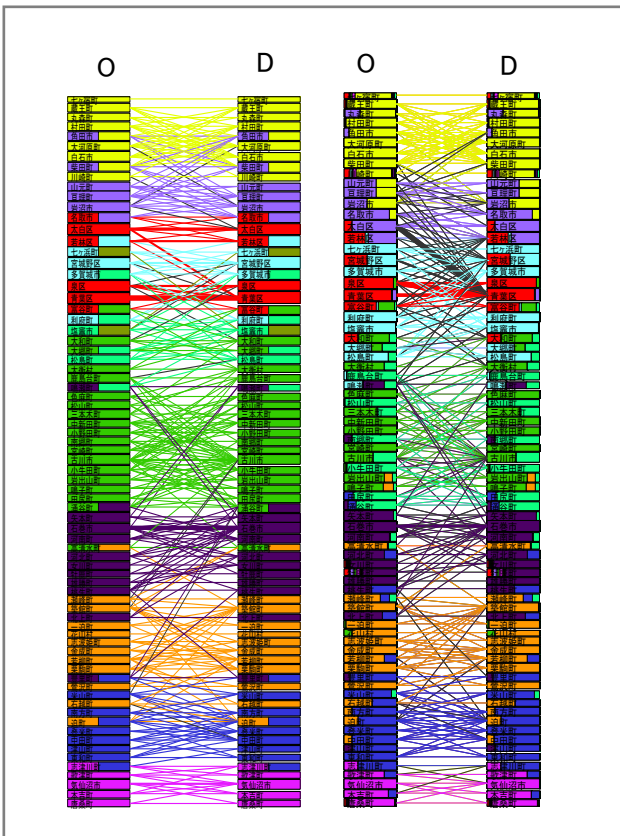


図 3 : 1 部グラフに Chen の手法を適用した結果(左) と 1 部グラフに Nepusz の手法を適用した結果(右)

有向性を考慮することが機能地域検出結果に与える影響を、コミュニティの重複を許容したコミュニティ抽出法を適用することで機能地域同士の重複を許容することが機能地域の検出結果に与える影響を、

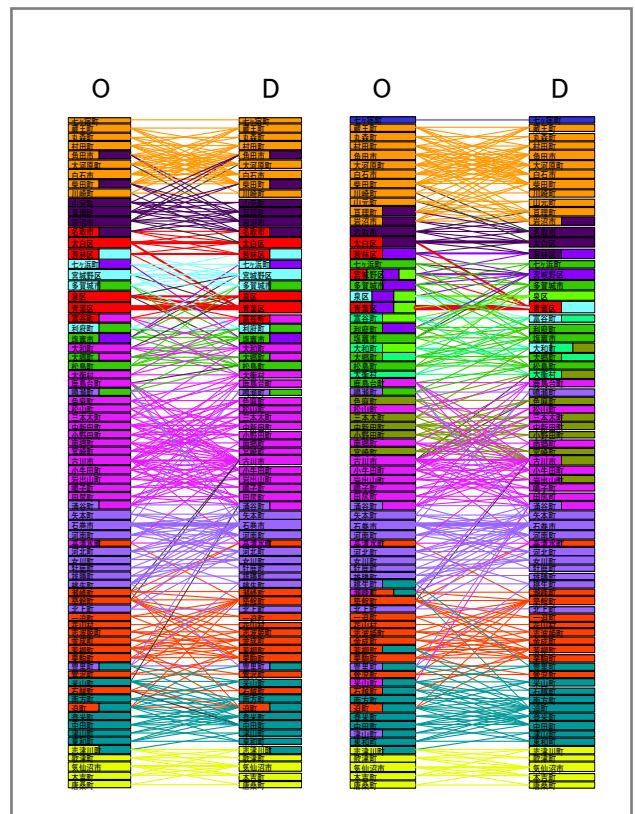


図 4 : 1 部グラフに Chen の手法を適用した結果(左) と 2 部グラフに Chen の手法を適用した結果(右)

それぞれ検証した。

分析結果として、フローの有向性と重複を別々に考慮することと既存手法では検出できなかった機能地域の構造を検出できることを明らかにした。一方、両者を同時に考慮すると、意味解釈の難しい結果が得られることが分かった。

参考文献

- 1) Farmer, C.J.Q. and Fotheringham, A.S.: Network-based functional regions, *Environment and Planning A*, Vol.43, pp.2723-2741, 2011.
- 2) 福本潤也, 岡本佳洋: コミュニティ抽出法と空間相互作用モデルを組み合わせた機能地域区分手法の提案, *土木学会論文集D3*, Vol.68, pp.427-436, 2012.
- 3) Xie, J., Kelley, S., and Szymanski, B. K.: Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Comput. Surv.* 45, 4, Article 43, 35 pages, 2013.
- 4) Wang, X., Jiao, L., and Wu, J.: Adjusting from disjoint to overlapping community detection of complex networks. *Physica A*388, 5045- 5056, 2009.
- 5) Chen, W., Liu, Z., Sun, X., and Wang, Y.: A game-theoretic framework to identify overlapping communities in social networks. *Data Mining Knowl. Discov.* 21, 2, 224-240, 2010
- 6) Nepusz, T., Petroczi, A., Negyessy, L., and D Bazso, F.: Fuzzy communities and the concept of bridgeness in complex networks. *Phys. Rev. E* 77, 1, 2008.