

Functional Region設定問題への コミュニティ抽出法の適用

福本 潤也¹・岡本 佳洋²・安 康³

¹正会員 東北大学大学院准教授 情報科学研究科 (〒980-8579 仙台市青葉区荒巻字青葉6-6-06)

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

²非会員 茨城県土木部検査指導課 (〒310-8555 茨城県水戸市笠原町978-6)

E-mail: yo.okamoto@pref.ibaraki.lg.jp

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 仙台市青葉区荒巻字青葉6-6-06)

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

Functional Region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。本研究では、モジュラリティに基づくコミュニティ抽出手法をFR設定問題に適用する。空間相互作用はデータは地理的近接関係によって強く支配されており、FR設定問題でも地理的近接関係を明示的に考慮することが望ましい。本研究では、モジュラリティに基づくコミュニティ抽出手法の帰無モデルに代表的な空間相互作用モデルであるエントロピーモデルを用いることを提案する。市町村間通勤データに提案手法を適用し、FR設定手法としての有効性を明らかにする。

Key Words : *functional region, modularity, community finding, entropy model, commuting OD*

1. はじめに

Functional Region (以下, 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 (e.g. 都道府県, 市区町村) から構成される地域である¹⁾。身近な例として生活圈や通勤圏が挙げられる。FRを設定して各種統計を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。日本には政府公式のFRは存在しない。しかし、大半のOECD諸国には政府公式のFRが存在する²⁾。

FRを設定することで実務的・学術的な用途に役立てることもできる。実務的には、道州や定住自立圏等の各種圏域の設定、自治体の合併範囲の決定、複数の行政単位をまたぐ雇用政策や交通政策の策定や実施、後発地域の把握、等々の政策的な議論に役立つ基礎資料を提供できる。学術的には、都市構造の都市間比較、都市の成長速度の計測、都市集積の産業間比較、都市の階層性の把握、等々の分析に役立つ基礎資料を提供できる。公式にFRを設定しているOECD諸国の中には、実際にFR単位でデータを整備して社会経済分析や雇用構造分析を行い、地域政策の資料として活用している国がある²⁾。

これまで様々なFRの設定手法が提案されている。ただし、決定的な設定手法が確立されている訳ではない。FRを設定している国々は独自の基準を用い

て設定している。例えば、2つの市区町村間に“就業者の何%以上が相互に通勤する場合に同じFRに所属することにする”といった基準を用いている。これまで提案されてきたFR設定手法には、1) 最終的に決まるFRの数が恣意的に置かれる仮定によって左右される、2) 結果の妥当性を統計的に評価できない、といった問題がある。

本研究では、上記の問題に対処するため、ネットワーク科学の分野で発展してきたコミュニティ抽出法に着目する。コミュニティとは、あるネットワークにおいて高密度に張り巡らされた部分ネットワークのことを指す。コミュニティの抽出と、交通や経済取引などのフローで結びついた地理的単位の部分集合を抽出するFRの設定は、問題の構造が共通する部分が多い。数多くのコミュニティ抽出法が提案されているが、近年、最も多くの研究者が注目する方法の1つにモジュラリティに基づくコミュニティ抽出法がある。モジュラリティとは、巨大なネットワークからコミュニティを発見するために提案された指標である³⁾。モジュラリティを用いる利点に以下の3点がある：1) 分析結果の妥当性を評価する手法が用意されている、2) コミュニティ数が内生的に決まる、すなわち、分析者が恣意的に判断する必要がない、3) コミュニティを抽出するための効率的なアルゴリズムが用意されている。

モジュラリティに基づくコミュニティ抽出手法は、FR の設定手法としても利用可能であると考えられる。ただし、FR 設定問題に単純に援用することはできない。モジュラリティが WWW 等の地理的近接性の影響を考慮しないネットワークを対象としているのに対し、FR 設定問題は地理的近接性によって支配される空間相互関係を表すネットワークを対象にしているからである。モジュラリティを FR 設定問題に援用する場合、地理的影響を考慮可能にする拡張が求められる。

本研究では、モジュラリティに基づくコミュニティ抽出法に地理的近接性の影響を組み入れた新たな FR の設定手法を提案する。ケーススタディを通して提案手法の有用性を明らかにする。

2. モジュラリティの定義と問題

(1) モジュラリティの定義

モジュラリティとは、ネットワークを複数のコミュニティに分割する場合に、ネットワーク全体がリンク高密度なコミュニティに分割されているかどうか評価するための指標である。式(1)で定義される。

$$Q = \frac{1}{M} \sum_{c \in C} \sum_{i, j \in c} [A_{ij} - P_{ij}] \quad (1)$$

ただし、 i はノードのインデックス、 c はコミュニティのインデックス、 C はコミュニティの集合、 A_{ij} は観測ネットワーク (e.g. 隣接行列, OD 行列), M は観測ネットワークの総リンク数、 P_{ij} は観測ネットワークとの比較に用いる帰無モデルネットワークを表す。

モジュラリティを提案した Girvan & Newman³⁾ は帰無モデルネットワークとしてランダムネットワークを使用している。ランダムネットワークに基づく P_{ij} は式(2)で与えられる。

$$P_{ij} = \frac{k_i^{\text{out}} k_j^{\text{in}}}{M} \quad (2)$$

ただし、 k_i^{out} はノード i の出次数、 k_j^{in} はノード j の入次数である。

モジュラリティに基づくコミュニティ抽出法では、2つの代表的なアルゴリズムが用意されている。1つは、モジュラリティが増加するようにノードやコミュニティを逐次統合する Greedy Algorithm⁴⁾ (GA) である。もう1つは、モジュラリティ行列の固有ベクトルを用いてコミュニティを逐次分割するモジュラリティ・スペクトラル・クラスタリング⁵⁾ (MS) である。

(2) FR設定問題への適用

Girvan & Newmanの帰無モデルを単純にFR設定問題に適用する場合に生じる問題を具体例を通して確認する。

a) 対象データ

平成17年国勢調査による通勤者データをOD行列化した通勤ODデータを用いる。データの集計単位は市区町村単位である。就業者および通学者の合計を通勤者数として扱う。なお、通勤者数が10人未満の場合は0人とみなしている。

以下では、宮城県、東北6県、全国の3つの通勤ODにGAを適用し、地理的近接性を考慮しなかった場合の影響について検証する。GAのアルゴリズムについては3.(3)で後述する。

b) 適用結果と問題点

図-1は、それぞれ宮城県、東北6県、全国の通勤ODにGAを適用して得られた宮城県内のFRである。図-1より、分析対象範囲が広がるに従い、抽出されるFRの規模も大きくなるのが分かる。

この原因は、ランダムネットワークの帰無モデルがデータ数に依存して決まるためである。通勤ODのようなデータでは自治体数が増えれば総通勤数 M が増加するが、各自治体の出通勤数 k_i^{out} や入通勤数 k_j^{in} はほとんど変化しない。そのため、 P_{ij} は自治体数が増加するにつれて減少し、近接している地域が同じFRに所属しやすくなる。通勤ODのような地理的近接性によって支配されるネットワークにモジュラリティに基づくコミュニティ抽出法を単純に適用する場合に生じる問題である。

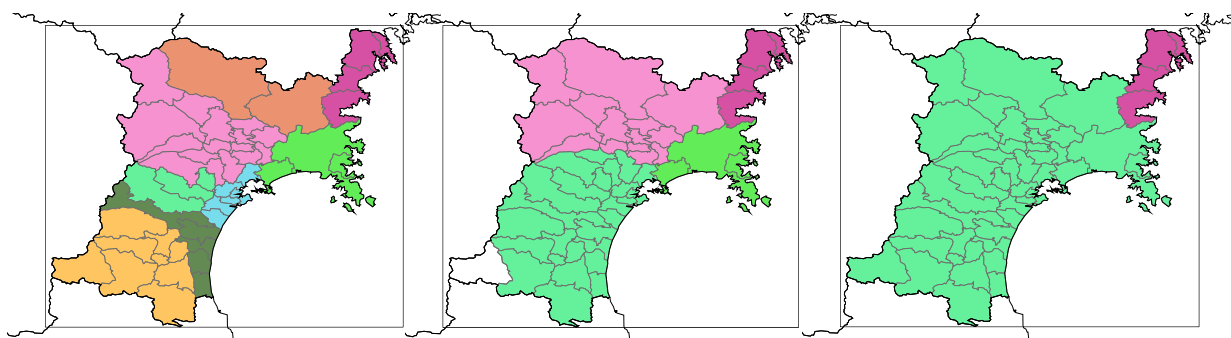


図-1：既存手法による宮城県内の FR (左から宮城県、東北 6 県、全国の通勤 OD)

分析対象範囲の変化によって抽出されるFRが異なることは好ましくない。これに対処するには、通勤ODに適した帰無モデルを採用する必要がある。具体的には、地理的近接性を考慮可能な帰無モデルを採用する必要がある。

3. 提案手法

(1) エントロピーモデルの適用

本研究では地理的近接性を考慮したモジュラリティを定義するにあたり、帰無モデルにエントロピーモデルを用いることを提案する。エントロピーモデルに基づくと、自治体間の通勤数は式(3)で与えられる。

$$\begin{aligned} q_{ij} &= a_i b_j O_i D_j \exp(-\gamma C_{ij}) \\ a_i &= \left[\sum_j b_j D_j \exp(-\gamma C_{ij}) \right]^{-1} \\ b_j &= \left[\sum_i a_i O_i \exp(-\gamma C_{ij}) \right]^{-1} \end{aligned} \quad (3)$$

ただし、 $O_i = k_i^{out}$ 、 $D_j = k_j^{in}$ 、 C_{ij} は自治体*ij*間のインピーダンス、 γ は地理的近接性の減衰パラメータである。

式(1)の P_{ij} を q_{ij} に置き換えた式(4)が本研究で用いる地理的近接性を考慮したモジュラリティである。

$$Q = \frac{1}{M} \sum_c \sum_{i,j \in c} [A_{ij} - q_{ij}] \quad (4)$$

(2) 自治体の隣接関係の考慮

国勢調査の通勤ODデータを吟味すると、日常的な通勤が不可能である遠距離の自治体間でも外れ値的に正の通勤トリップが存在することがある。この時、式のモジュラリティに基づくコミュニティ抽出法を適用すると、遠く離れたFRが一つのFRに統合されてしまう。外れ値的なデータによって遠隔地の自治体が飛び地FRを形成することは望ましくないと考えられる。そこで、本研究では自治体の隣接関係を考慮することで、飛び地FRが形成されないようにする。具体的には、自治体間の隣接関係を X_{ij} で表し、式(5)で定義される自治体間の隣接関係も考慮したモジュラリティを用いる。

$$Q = \frac{1}{M} \sum_c \sum_{i,j \in c} X_{ij} [A_{ij} - q_{ij}] \quad (5)$$

ただし、 X_{ij} は自治体*i*と*j*が隣接している場合に1、それ以外の場合に0になる対称行列である (X_{ii} は1)。

(3) FR設定手法

本研究では、式(5)を用いたGAとMSによるFRの設定手法を提案する。本稿では紙面の都合上、MSの説明は割愛する。

GAでは、仮コミュニティを結合した時にモジュラリティの増加量が最大となる仮コミュニティ同士を逐次結合することでコミュニティを抽出する。任意のコミュニティのペア*i, j*を統合した場合のモジュラリティの増加量は式(6)で与えられる。

$$B_{ij} = A_{ij} - q_{ij} \quad (6)$$

また、自治体の隣接関係を考慮した場合、モジュラリティの増加量は式(7)で与えられる。

$$B_{ij}^X = X_{ij} [A_{ij} - q_{ij}] \quad (7)$$

以下、隣接関係を考慮したGAの手順を示す。

- i) 全自治体を独立した仮コミュニティとみなし、コミュニティ $V = \{V_1, \dots, V_M\}$ とする。ただし、 M は全ノード数、 $|V| = M$ である。
- ii) 式(7)を算出する。
- iii) $(l, m) = \arg \max_{l, m} (\mathbf{B}_X + \mathbf{B}_X^T)$ となるコミュニティのペア (V_l, V_m) を見つける。
- iv) $B_{lm}^X + B_{ml}^X \leq 0$ ならば、アルゴリズムを終了する。
- v) $B_{lm}^X + B_{ml}^X > 0$ ならば、 (V_l, V_m) を統合し、以下のように変数を更新した後(ii)に戻る。
 - $V = \{V_1, \dots, V_M, V_{(lm)}\}$, $V_l = V_m = \emptyset$
 - $|V| = |V| - 1$
 - $B_{(lm)(lm)} = B_{ll} + B_{lm} + B_{ml} + B_{mm}$
 - $B_{(lm)k} = B_{lk} + B_{mk}$
 - $B_{k(lm)} = B_{kl} + B_{km}$
 - $X_{(lm)(lm)} = 0$
 - $X_{(lm)k} = X_{k(lm)} = \max(X_{lk}, X_{mk})$

4. ケーススタディ

(1) 地理的近接性データ

本研究では、地理的近接性データとして、市町村間の所要時間を用いる。近接性データは、国土交通省の総合交通分析システムNITASで作成した。

各自治体間の所要時間には、自動車と鉄道をそれぞれ利用した所要時間のうち、時間が最小となる交通モードの所要時間を使用する。所要時間を計測するための起終点にはそれぞれの役所または役場の地点データを用いる。

(2) 適用結果

減衰パラメータ γ を0.04に固定してGAを適用した結果を図-2に示す。図-1と比較すると、分析対象範囲が拡大しても抽出されるFRは大きく変化しないことが分かる。地理的近接性を考慮することで分析対象範囲の変化に対する頑健性が向上することが確認できる。

東北6県の通勤ODを対象に減衰パラメータ γ を変化させながらFRを抽出した結果を図-3に示す。図-3

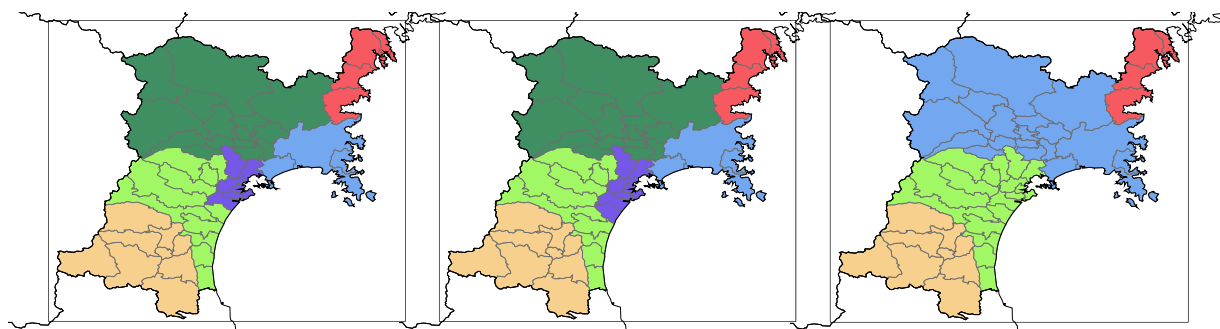


図-2：提案手法による宮城県内のFR（左から宮城県，東北6県，全国の通勤OD， $\gamma=0.04$ ）

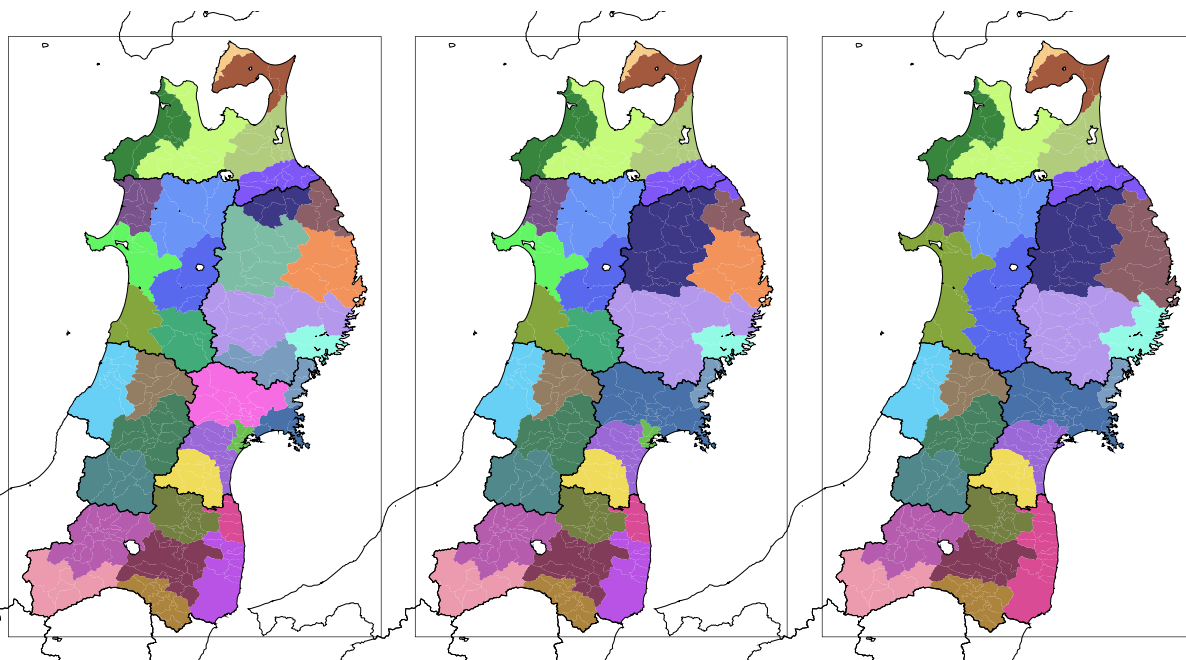


図-3：東北6県のFR（左から $\gamma=0.04$ ，0.035，0.03）

より， γ が減少するにつれ，より大きなFRが抽出されることが分かる． γ が大きいときは近接する自治体間の帰無モデル上の通勤が大きくなる．一方， γ が小さくなると帰無モデル上の通勤は広範囲に分散する．その結果， γ の変化に伴って抽出される抽出されるFRの範囲も変化することになる．

FRの範囲の変化を見ることで，FRの階層性や複数のFRに重複して含まれる自治体（オーバーラップ）の存在を確かめることもできる．例えば，宮城県に着目すると， γ が0.04から0.035に変化したとき，石巻市を含むFRと古川市を含むFRが統合したFRが形成されている．この統合したFRが階層性，オーバーラップのどちらかの構造を示していると考えられる．

5. 結論

本研究では，モジュラリティに基づくコミュニティ抽出法に地理的近接性の影響を組み入れた新たなFR設定手法を提案した．ケーススタディを通し

て，以下の2点を明らかにした：1) 地理的近接性を考慮したことで，データ対象の空間範囲の変化に依存しないFRを設定可能となった，2) 減衰パラメータ γ を変化させることで，FRの階層性やオーバーラップの存在を示した．

今後の課題として，FRの階層性を明示した設定手法，オーバーラップを許容する設定手法へ拡張していく必要がある．

参考文献

- 1) Brown, L.A. and Holmes, J.: The delimitation of functional regions, nodal regions, and hierarchies by functional distance Approach, *Journal of Regional Science*, Vol.11, pp.57-72, 1971.
- 2) OECD: *Redefining Territories, The Functional Regions*, OECD, 2002.
- 3) Girvan, M. and Newman, N.E.J.: Community structure in social and biological networks, *PNAS*, Vol.99, pp.7821-7826, 2002.
- 4) Newman, M.E.J.: Fast algorithm for detecting community structure in networks, *Physical Review E*, Vol.69, 066133, 2004.
- 5) Newman, M.E.J.: Modularity and community structure in networks, *PNAS*, Vol.103, pp.8577-8582, 2006.