

## グラフ研磨を用いた乗降パターンによる駅のクラスタリング

### Railway Station Clustering based on Origin-Destination Patterns using Graph Polishing

細江 美欧\*・桑野 将司\*\*・森山 卓\*\*  
Mio Hosoe\*, Masashi Kuwano\*\*, Taku Moriyama\*\*

With the development of ICT, interest in traffic policy planning by utilizing large varieties of accumulated big data has been increasing. In recent years, graph polishing has been proposed as a new methodology for graph clustering. Graph polishing is one of the graph clustering methods. This method can be used to extract groups that are similar or related to each other by clarifying the cluster structures in the data. This study classifies railway stations by applying the graph polishing to smart card data that has been introduced in Kagawa Prefecture, Japan. This study uses 9,008,709 data collected during the 15 months from December 1st, 2013 to February 28th, 2015, and creates Origin-Destination network. Then, this study clarifies station groups and examines the usefulness of graph polishing to Origin-Destination network clustering.

**Keywords:** Traffic Big Data, Smart Card, Origin-Destination Table, Similarity

交通ビッグデータ, 交通系 IC カード, OD 表, 類似性

#### 1. はじめに

情報通信技術の進歩に伴い、長期的かつ継続的にデータが収集・蓄積されるようになったことから、ビッグデータと呼ばれる巨大で複雑なデータが利用可能となった。交通分野では、携帯電話やカーナビに搭載されている GPS のログデータ、ETC データ、交通系 IC カードの乗降履歴データ、プローブカーデータなどのビッグデータが注目されている。これらビッグデータには人や車の移動軌跡が詳細に記録されており、従来の紙面調査では得られなかった交通行動を把握できると期待されている。

特に、交通系 IC カードの乗降履歴データにはカード利用者が、いつ、どこから乗車し、どこで降車したのかという時空間的な移動軌跡が記録されていることから、公共交通の利用実態を把握する上で活用されている。例えば、乗降履歴データに記録される OD (Origin-Destination) 情報の分析によって、どの地点間を移動したかを把握できる。

地点間の交通量把握のための代表的な方法の 1 つとしてクロス集計がある。クロス集計では乗車駅と降車駅の組合せごとにカード利用者数を集計し、OD 表を作成することで、駅間の移動の多寡を明らかにできる。さらに、OD 表はネットワーク図によって可視化し、視覚的に捉えやすくなる。しかし、対象とする起終点の組合せ、すなわち交通機関に存在する駅数が多い場合には、OD 表あるいはネットワーク図は複雑になる。

ビッグデータのようにデータ量が膨大で、データ項目の要素数が多い(次元が高い)場合には、一般的にデータマイニングが有効となる。データマイニングには、データ項目間の相互関係に基づき項目を縮約する情報縮約と、同一のパターンや規則性を持つサンプルをグループごとに束ねるクラスタリング手法などがある。前者の情報縮約におい

ては、トピックモデル<sup>1)</sup>や非負値行列因子分解<sup>2)</sup>、非負値テンソル因子分解<sup>3)</sup>が近年盛んに使用されているが、本研究では後者のクラスタリング手法に着目する。なお、ビッグデータに対して前者の情報縮約を行う場合、データ項目やデータ項目内の要素数が多くなると、因子数が一意に定まらない場合や、あるデータ項目あるいはデータ項目内の要素がどの因子に属するのかの判断が難しくなるなどの問題があることが知られている<sup>4)</sup>。このような場合、情報縮約を行なった後にサンプルをクラスタリングし、情報縮約で得られた結果の解釈や妥当性の検証がなされることが多い。例えば、因子分析では因子が得られた後に、サンプルをクラスタリングし、グループごとに各因子の因子得点の平均値を算出し比較することで、サンプルグループと因子の関係を推測し、因子の解釈が妥当であるか否かの検討が行われる。このように、情報縮約を目的とした分析においても、結果の妥当性を判断するためにクラスタリング手法は重要な意味を持つ。

本研究では、乗降客の行動パターンや規則性に基づき、駅をグループごとに束ねるクラスタリング手法について検討することを目的とする。ここで、分析対象とする駅の数が多い場合、1 つ 1 つの駅を個別に分析しようとする、その分の数のモデル分析が必要となり、分析が煩雑になるだけでなく、全体としての特徴が把握できなくなる。一方、多数の駅を同時に分析しようとする、全体としての平均的な特性のみが抽出されることになり、例えば周辺に住宅地が多い、あるいは事業所が多いなどで異なると考えられる駅間の異質性が考慮できない。潜在クラスモデリングなど利用特性と異質性を同時に考慮する分析方法も存在する<sup>5)</sup>が、利用特性が異なる駅が多数存在する場合、膨大な数の潜在クラスを仮定した上でモデル分析を行い、その上で最

\* 学生会員 鳥取大学大学院工学研究科 (Tottori University)

\*\* 正会員 鳥取大学大学院工学研究科 (Tottori University)

適なクラス数を決定する必要があるため計算負荷が高い。よって、全体としての特徴を保持しながら異質性を持つ駅ごとにグルーピングするクラスタリングを予め行うことが公共交通計画にとって重要な意味を持ち、その後利用特性を分析するためにも有益であると考えられる。

ところで、内々交通量がゼロまたは極めて小さい値であるとき、OD 表を対角成分が欠損している行列とみなすことができる。特に、乗車駅と降車駅の組合せで表現される鉄道の OD 表を考えると、内々交通量は存在しない。つまり、乗車駅と降車駅が同じ駅の組合せは情報を持たない。これにより、異なる乗車駅同士は異なる降車駅の組合せを持つことになる。この場合、時間帯ごとの利用者数や利用率、駅周辺の施設立地情報などを用いることで駅のクラスタリングは可能であるが、OD パターンから駅のクラスタリングを行う場合には、従来のクラスタリング手法の適用は困難である。

そこで本研究では、グラフ理論の枠組みの中で、駅のクラスタリングを試みる。具体的には、各駅を頂点(ノード)、駅間の利用件数を辺(リンク)として、OD 交通量をグラフによって記述する。そして、条件付き確率の考え方を応用した類似度指標を用いることで、乗降客数に偏りが存在し、相対的に利用件数が少ない駅(頂点)であっても分析の過程で特徴が失われにくいクラスタリング手法であるグラフ研磨を適用し、OD 表から乗降パターンが似た駅を抽出し、分析手法の有用性を検証することを目的とする。グラフ研磨はマーケティング分野や自然言語処理分野で開発され、商品の購買履歴から同時に購入される商品の組合せを明らかにする研究<sup>6)7)</sup>や SNS 上で発信された文章を要約する研究<sup>8)</sup>に用いられている。なお、クラスタリング手法には大別して1つの要素が1つのグループに属するハードクラスタリングと1つの要素が1つ以上のグループに属するソフトクラスタリングがある。ソフトクラスタリングは、ハードクラスタリングを包含可能であり、より柔軟にデータの傾向を捉えることができることから、適用事例が増えている手法である。本研究で用いるグラフ研磨もソフトクラスタリングである。

## 2. 使用データの概要

本研究では香川県において利用されている交通系 IC カード「IruCa」の乗降履歴データを用いる。IruCa は高松琴平電気鉄道株式会社によって運営されている鉄道路線「ことでん」、バス路線「ことでんバス」および県内バス会社路線で利用可能である。本研究では「ことでん」における乗降履歴データのみを用いる。

図-1 に示すように、ことでんには琴平線、長尾線、志度線の3路線全52駅が存在する。高松築港駅から瓦町駅の区間は琴平線と長尾線の2路線が共有しており、「ことでん」の中で最も運行本数が多い区間である。この区間は高松市の中心市街地に位置する。高松築港駅の近くにはJR高松駅や高松港が位置しており、他交通機関への乗換えが可

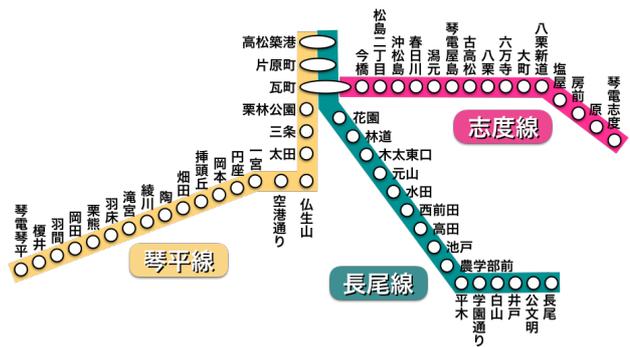


図-1 ことでん路線図<sup>9)</sup>

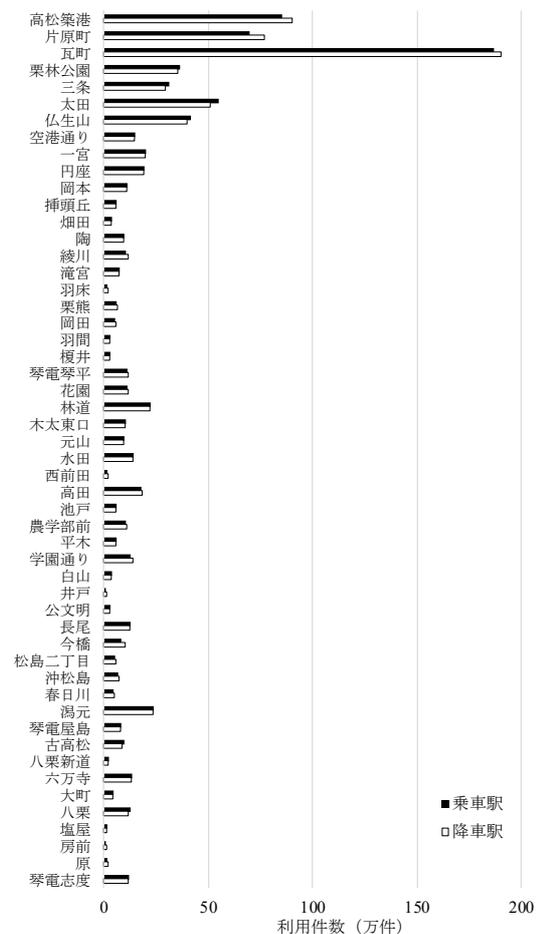


図-2 乗車駅および降車駅別利用状況

能である。また、片原町駅周辺には百貨店や丸亀町商店街などが存在し、高松市の商業的中心地である。中心部から離れた駅の周辺には、住宅地や観光地が広がり、特に、琴平線上の駅周辺には主要な観光地が存在する。例えば、栗林公園駅近くには国の特別名勝に指定された回遊式日本庭園が、琴電琴平駅近くには「こんぴらさん」の呼び名で有名な金刀比羅宮がある。

IruCa の乗降履歴データには、乗降車日時、乗降車駅、移動所要時間、カード利用区分などが記録されているが、本研究では乗降車日時、乗降車駅、移動所要時間に関する情

報を用いる。データの収集期間は 2013 年 12 月 1 日から 2015 年 2 月 28 日までの 15 ヶ月間で、この期間に収集されたデータ数は 9,033,748 件である。このうち、ことでの運行時間帯である 5 時台から 24 時台に利用され、移動所要時間が 60 秒以上のデータ 9,008,709 件を有効データとして用いる。

図-2 に乗車駅および降車駅別利用件数を示す。図-2 より、各駅は乗車駅と降車駅で利用件数に差は見られない。駅ごとで利用件数を比較すると、琴平線と長尾線の共有駅である高松築港駅や片原町駅、3 路線の結節駅である瓦町駅は利用件数が 50 万件を超えているが、その他多くの駅は 25 万件を下回っていることがわかる。図-2 より、「ことでん」の駅間の利用件数には大きな偏りがあることが確認できた。中心部では利用件数が多く、郊外に行くにつれて利用件数が著しく減少する傾向は「ことでん」に限ったことではなく、都市における公共交通利用件数の分布の一般的な傾向であると考えられる。

OD 表から、駅の延利用件数を頂点、駅間の延利用件数を辺で表現したネットワーク図を図-3 に示す。ただし、グラフの複雑さを回避するために、分析対象期間における延利用件数が 2,500 件以下の駅の組合せは削除した。削除後に残ったデータ数は全利用件数の約 95%にあたる。同図において、各駅は、高松築港駅を上部に配置し、そこから右回りに片原町駅、瓦町駅、琴平線駅、長尾線駅、志度線駅の順に配置している。各路線上の駅の配置の順は図-1 と対応するように、琴平線駅は栗林公園駅、三条駅、...、榎井駅、琴電琴平駅、長尾線駅は花園駅、林道駅、...、公文明駅、長尾駅、志度線駅は今橋駅、松島二丁目駅、...、原駅、琴電志度駅である。また、琴平線上に位置する駅の円を黄色、長尾線上に位置する駅の円を緑色、志度線上に位置する駅の円を桃色で色分けしている。白色の円は共有駅である。図-3 内の頂点数は 52、辺数は駅の組合せ(52 × 52 - 52) / 2 = 1,326 から延利用件数 0 件の 44 組合せと、さらに、スクリーニングによって削除した延利用件数 2,500 件以下を除いた 350 である。全組合せから延利用件数が少ない組合せを削除することで、約 4 分の 1 の組合せ数になったものの、OD の特徴を把握するにはまだ複雑である。そこで、延利用件数が 5 万件以上の駅の組合せのみを抽出し、再度描いたネットワーク図を図-4 に示す。なお、円の大きさと線の太さは図-3 の凡例と同様である。図-4 では頂点数は 28、辺数は 45 となった。3 路線の結節点である瓦町駅と結ばれる駅数が最も多く、次いで琴平線と長尾線の共有駅である高松築港駅と片原町駅が多い。また、栗林公園駅、三条駅、太田駅、仏生山駅、一宮駅、円座駅、林道駅は 2 つ以上の辺数がある。図-4 からは延利用件数が多い OD ペアの把握は可能であるが、延利用件数が少ない駅は無視されており、類似した乗降パターンを持つ駅を抽出することはできない。そこで、次章以降ではグラフ研磨を適用することで、図-3 に示したネットワーク図から乗降パターンが類似した駅グループの抽出を試みる。

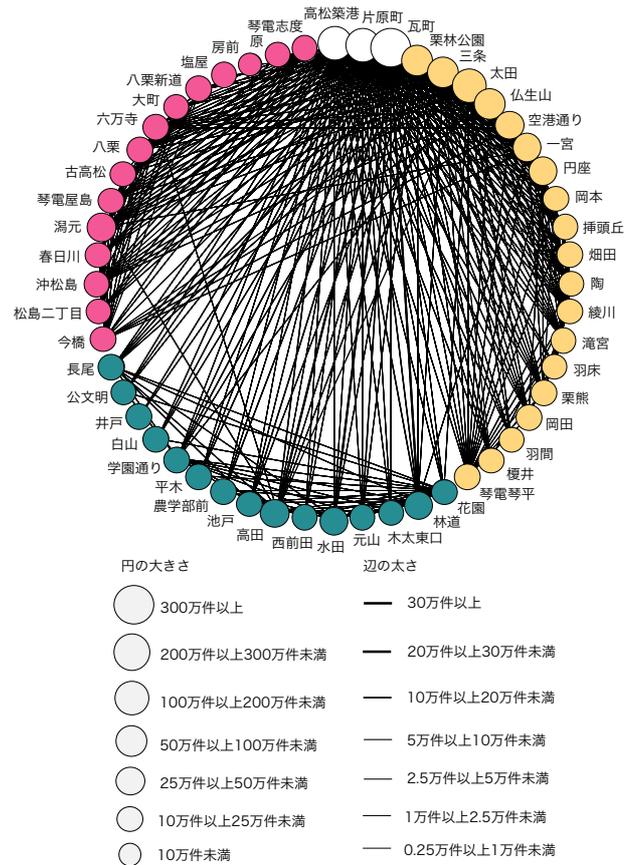


図-3 駅の組合せ別の利用状況

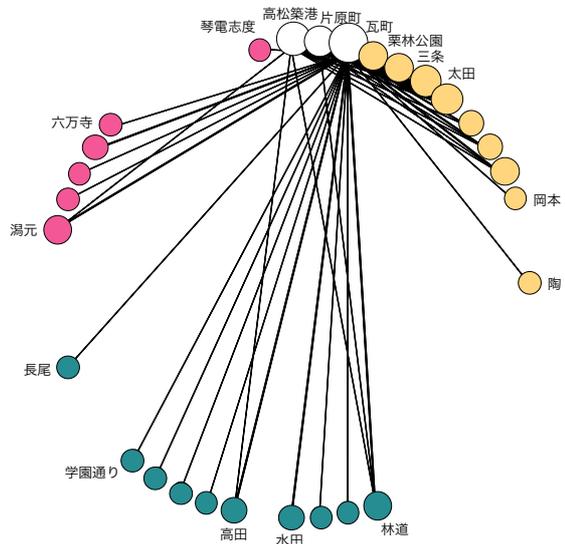


図-4 延利用件数 5 万件以上の駅の組合せ

### 3. グラフ研磨を用いた駅のクラスタリング方法

#### (1) 記法

$V$  を頂点集合、 $E$  を辺集合とし、2 つの集合のペア  $G = (V, E)$  をグラフという。ただし、無向グラフとする。

任意の 2 つの頂点  $v_m, v_n$  について、 $v_m$  と  $v_n$  の間に辺が張られているならば、 $v_m$  と  $v_n$  は隣接しているという。また、

頂点 $v_m$ に隣接する頂点 $v_n$ を頂点 $v_m$ の近傍といい、頂点 $v_m$ ,  $v_n$ に共通して隣接する頂点を $v_m, v_n$ の共通近傍という。

$N(v_m)$ を頂点 $v_m$ に隣接する頂点集合とし、頂点 $v_m$ の閉近傍を $N[v_m] (=N(v_m) \cup \{v_m\})$ と表現する。ここで、 $\{v_m\}$ は頂点 $v_m$ を元を持つ要素数 1 の集合である。頂点 $v_m$ の閉近傍の個数は $|N[v_m]|$ で表す。

## (2) 駅グループ抽出のための手順

グラフ研磨を用いた駅のクラスタリング方法は 1) 共起グラフの構築, 2) 共起グラフへのグラフ研磨の適用, 3) グループの抽出の 3 つの手順で構成される。各手順について順に述べる。

### (a) 共起グラフの構築

まず、前章の図3のグラフを基に共起グラフを構築する。共起グラフとは頂点間の共起関係を表するグラフであり、駅の各組合せのうち共起性のある組合せのみを残すことで構築される。ここで、共起が偶発的であるかを考慮するために、 $t$ 検定を用いて共起の統計的有意性を確認する。 $t$ 値は(1)式を用いて算出する。

$$occ(x_m, x_n) = \frac{|x_m \cap x_n| - \frac{|x_m| \times |x_n|}{W}}{\sqrt{|x_m \cap x_n|}} \quad (1)$$

$|x_m|$ は駅 $m$ の延利用件数を、 $|x_n|$ は駅 $n$ の延利用件数を表し、 $|x_m \cap x_n|$ は $x_m$ と $x_n$ の積集合の要素数、すなわち共通延利用件数を表す。 $W$ は延利用件数の総数を表す。なお、本研究では $t$ 値の絶対値が 1.65 以上であれば有意であるとす<sup>10)</sup>。そして、有意な駅の組合せの延利用件数を 1 に、有意でない組合せの延利用件数を 0 に置き換える。これにより、駅間の利用件数や選択確率などの情報は失われることになるが、2 値変数に変換することによって、相対的に延利用件数が少ない組合せであっても、延利用件数が多い組合せと同じように乗降パターンの検出が可能となる。これが他のクラスタリング手法と比較したときのグラフ研磨手法の特徴である。

### (b) 共起グラフへのグラフ研磨の適用

共起グラフから、共通した特徴を持つ駅のグループを抽出するために、共起グラフにグラフ研磨を適用する。

グラフ研磨は宇野<sup>11)</sup>によって提案されたグラフから共通した特徴を持つグループを抽出するためのソフトクラスタリング手法である。グラフ研磨では「2 つの頂点に共通して隣接する頂点の個数がある程度以上多ければ、それらの頂点は一定以上の密度の部分グラフ $G'(G' \subseteq G)$ に含まれる」という仮定に基づいて、共通隣接頂点の個数に閾値を与え、任意の頂点对に対して閾値以上であれば頂点間を辺で結び、そうでなければ辺を削除する<sup>11)</sup>。

任意の 2 つの頂点が  $k$  個以上の共通隣接頂点を持つことを  $k$ -共通近傍条件という。 $k$ -共通近傍条件はグラフに含まれる各部分グラフが比較的独立して存在し、頂点が複数の部分グラフ(グループ)で重ならないときに有効な条件となる<sup>11)</sup>。しかし、隣接する頂点の個数が多い頂点が複数存

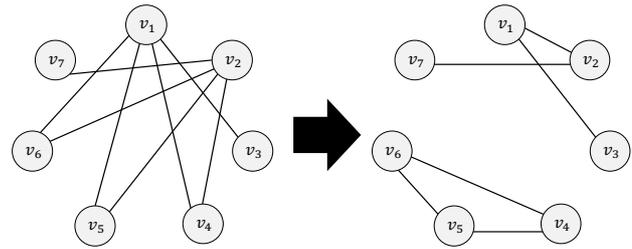


図-5 共起グラフおよび研磨グラフの例

在するとき、それらに隣接する頂点がすべて 1 つのグループに含まれることがある。これは共通近傍のみに着目し、共通しない近傍がどの程度存在するかを考慮していないために発生する<sup>11)</sup>。そこで、任意の 2 つの頂点の類似度を求める。集合の類似指標には Simpson 係数や Dice 係数などがあるが、本研究では既往研究<sup>6,8), 12)</sup>と同様に Jaccard 係数を用いる。Jaccard 係数による頂点 $v_m, v_n$ の類似度 $sim(v_m, v_n)$ は(2)式で求められる。 $|N[v_m] \cap N[v_n]|$ は $N[v_m], N[v_n]$ の積集合の要素数、 $|N[v_m] \cup N[v_n]|$ は $N[v_m], N[v_n]$ の和集合の要素数であり、(2)式は任意の頂点 $v_m, v_n$ の閉近傍の類似度を表している。

$$sim(v_m, v_n) = \frac{|N[v_m] \cap N[v_n]|}{|N[v_m] \cup N[v_n]|} \quad (2)$$

Jaccard 係数を用いたグラフ研磨の計算手順について図-5の左図に示す共起グラフを例に述べる。まず、(2)式を用いてすべての頂点对の類似度を計算する。例えば、頂点 $v_1, v_2$ の類似度を算出する場合、頂点 $v_1$ の閉近傍は $N[v_1] = \{v_1, v_3, v_4, v_5, v_6\}$ 、頂点 $v_2$ の閉近傍は $N[v_2] = \{v_2, v_4, v_5, v_6, v_7\}$ であるから、 $|N[v_1] \cap N[v_2]| = 3$ 、 $|N[v_1] \cup N[v_2]| = 7$ となるので、類似度は $sim(v_1, v_2) = 3/7 \approx 0.43$ となる。

次に、求めた類似度が閾値以上であれば頂点間を辺で結び、そうでなければ辺を削除する。閾値は分析者が任意で設定する。図-5の場合、例として閾値を 0.4 とすると、図-5の右図のようにグラフは変形する。この処理をグラフの変形が収束するまで行う。図-5の共起グラフの場合は、グラフ研磨を再度適用してもグラフは図-5の右図から変化しないため、1 度の適用でグラフ研磨は終了する。グラフ研磨によって最終的に得られるグラフは研磨グラフと呼ばれる。

### (c) グループの抽出

グラフ研磨において頂点が辺で結ばれている場合、それらは同一のグループに属する頂点对と解釈し、すべての頂点对が結ばれている頂点集合を 1 つのグループと考える。そのため、図-5の研磨グラフにおいてグループ数は、 $\{v_1, v_2\}, \{v_1, v_3\}, \{v_2, v_7\}, \{v_4, v_5, v_6\}$ の 4 つとされる。

ただし、抽出されたグループの解釈は a) グラフ研磨前後で頂点間に辺がある場合と、b) 新たに辺が追加された場合とで異なると考えられる。例えば、 $v_1$ はグラフ研磨前は $v_3$ 、

$v_4, v_5, v_6$  と繋がっていたが、グラフ研磨後は  $v_3$  のみと繋がっている。これは  $v_1$  と  $v_3$  の共起性が高く、さらに  $v_1$  と  $v_3$  の類似度が  $v_1$  と  $v_4, v_5, v_6$  との類似度よりも高いために、辺が残ったといえる。一方で、 $v_1$  と  $v_2$  には新たに辺が追加されている。これはこれら 2 つの頂点が  $v_4, v_5, v_6$  と共通して辺で結ばれていたことが特徴となり、辺で結ばれたと考えられる。つまり、 $v_1$  と  $v_2$  は他頂点との接続が類似するという特徴を持つ。

#### 4. グラフ研磨による駅のクラスタリング

図-3 のネットワーク図から得られた共起グラフを図-7 に、共起グラフにグラフ研磨を適用した結果を図-8 に示す。ここで、比較のために、延利用件数が 5 万件以上の駅の組合せのみを残した図-4 と同規模の研磨グラフが得られるように類似度の閾値 0.61 と設定した。その結果、研磨グラフの頂点数は 27、辺数は 38 となった。

図-8 より、駅グループは、

- G1) 仏生山駅, 一宮駅
  - G2) 三条駅, 太田駅
  - G3) 高松築港駅, 片原町駅, 瓦町駅
  - G4-1) 沖松島駅, 渦元駅, 六万寺駅
  - G4-2) 沖松島駅, 六万寺駅, 琴電志度駅
  - G5-1) 林道駅, 水田駅, 高田駅, 長尾駅
  - G5-2) 林道駅, 高田駅, 学園通り駅, 長尾駅
  - G6) 元山駅, 池戸駅
  - G7) 西前田駅, 平木駅, 白山駅
  - G8) 円座駅, 岡本駅, 陶駅, 栗熊駅, 滝宮駅, 琴電琴平駅
- の 10 グループが抽出された。グループを構成する駅に着目すると、延利用件数の多い高松築港駅、片原町駅、瓦町駅などだけでなく、延利用件数が相対的に少ない西前田駅や栗熊駅なども抽出されている。これより、駅間の延利用件数に差があるような OD 表であっても、共通した特徴を

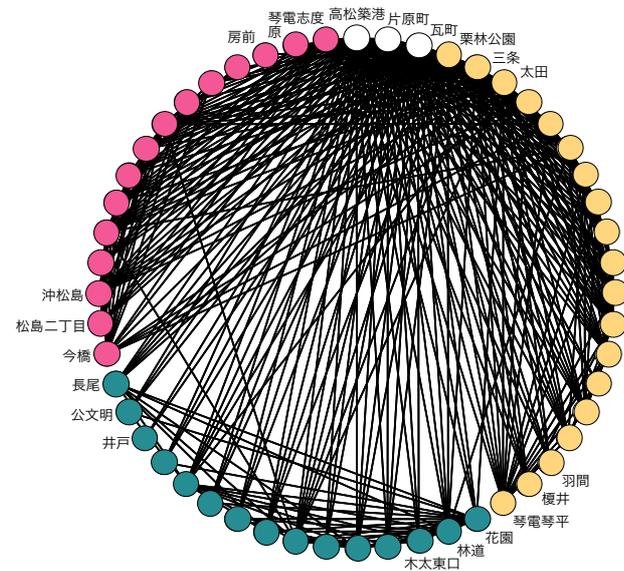


図-7 共起グラフ

持つ駅のグループが抽出されていることがわかる。

続いて、前章で述べた通り、グラフ研磨前後で駅間が辺で結ばれているか否かによって 2 つの解釈ができることに留意し、研磨グラフの考察を行う。

a) グラフ研磨前後で頂点間に辺が残ったグループは (G1), (G2), (G3), (G4-1), (G4-2), (G5-1), (G5-2) であった。グラフ研磨前に辺が存在するとは、2 つの駅間で各駅の延利用件数の合計に占める共通延利用件数が統計的に有意に多く、両駅間の利用が相対的に多いことを表している。さらに、グラフ研磨後も辺が残っているのは、その 2 駅間の利用が多いことに加え、それ以外の駅との乗降パターンも類似していることを表している。

例えば、G1) 仏生山駅、一宮駅に着目する。仏生山駅と一宮駅の延利用件数はそれぞれ 792,378 件、374,775 件で、2 駅間の延利用件数は 5,075 件であることから、3. (2) (a) に記す(1)式を用いて両駅間は統計的に有意な共起性を持つと確認できる。さらに、仏生山駅と一宮駅それぞれの乗降パターン (図-3 から当該駅のみを抽出したもの) を示した図-9 より、両駅の乗降パターンは類似していることがわかる。

次に、G4-1), (G4-2) に着目する。G4-1) と G4-2) は、4 駅で構成される部分グラフであるが、渦元駅と琴電志度駅間には辺がない。これら 4 駅の乗降パターンを示した図-10 より、「渦元駅と沖松島駅・六万寺駅」、「琴電志度駅と沖松島駅・六万寺駅」はそれぞれ乗降パターンが類似しているが、「渦元駅と琴電志度駅」には少し違いが見られる。

同じ部分グラフを構成しているため、G4-1) と G4-2) を

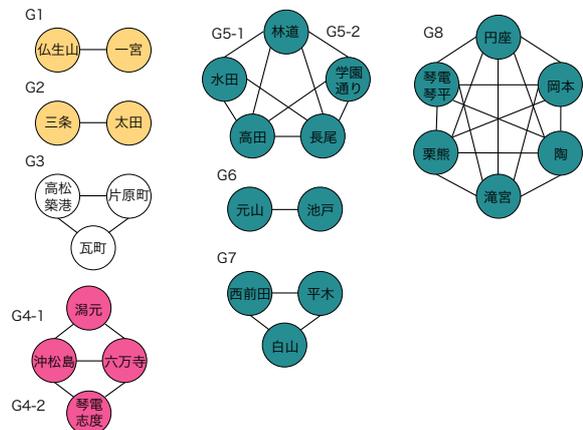


図-8 研磨グラフ (閾値 0.61)

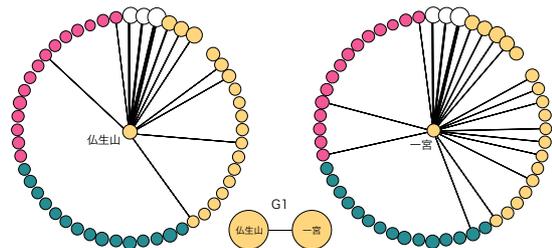


図-9 G1 に属す駅別乗降パターン

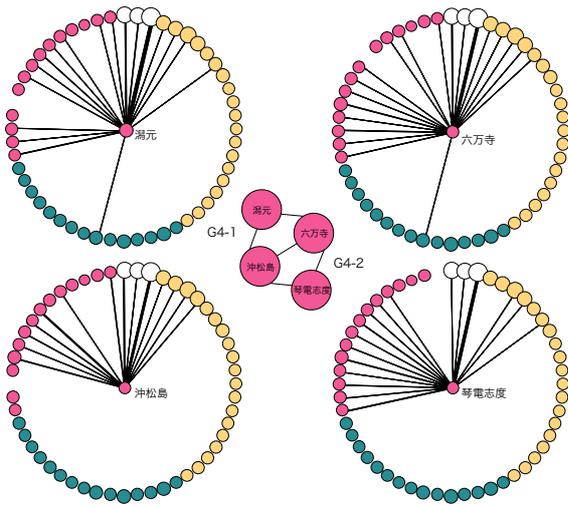


図-10 G4-1 および G4-2 に属す駅別乗降パターン

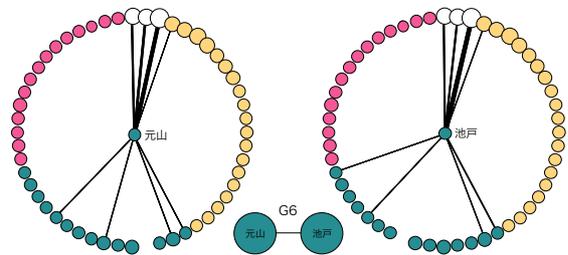


図-11 G6 に属す駅別乗降パターン

1 つのグループと考えることもできるが、辺で結ばれた完全グラフによって 2 グループと見ることもできる。分析目的に応じて、サンプルを大きく分類するときには前者の考え方を、細かく分類するときには後者の考え方をを用いればよく、柔軟にクラスタリングできる点もグラフ研磨の利点の 1 つといえる。

一方、b) グラフ研磨後に新たに辺が張られたことで類似駅として抽出されたグループは G6、G7 であった。新たに辺が追加されることは、両駅間の利用は多いとはいえないが、他駅の利用パターンが類似していることを表している。G6) 元山駅、池戸駅の乗降パターンを図-11 に示す。図-11 より、元山駅と池戸駅の乗降パターンが類似していることが確認できる。

さらに、G8) はグラフ研磨前後で頂点間に辺が残った頂点对と新たに辺が張られた頂点对で構成されるグループである。駅ペアの数だけ繰り返し共起分析を行うことによって共起性が高い駅ペアを抽出する、あるいは繰り返し相関分析を行い乗降パターンが類似した駅ペアを抽出することなく、共起性と類似性を同時に考慮したグループの抽出が抽出されていることも G8) の結果から示された。

## 5. おわりに

本研究では OD 表から乗降パターンが類似する駅グループを抽出するために、都市計画や交通計画の分野では適用

事例がないグラフ研磨というクラスタリング手法に着目し、その有用性を検証した。分析の結果、グラフ研磨によって、延利用件数が比較的少なくても類似した乗降パターンを持つ駅を抽出できることを実証的に明らかにした。公共交通の利用促進施策を検討する際には、高頻度に利用するような定期券の保有者の利用頻度を高めるだけでなく、低頻度に利用される駅間の利用頻度を上げるための施策を実施することも重要である。しかし、低頻度に利用される駅や駅間では、利用件数が少ないことからデータ分析においてその傾向は埋もれやすい。グラフ研磨は、利用件数が多い駅や駅間と利用件数が少ない駅や駅間が混在し、利用件数の分布に偏りがある場合に、利用件数が多い駅や駅間だけでなく、少ない駅や駅間であっても類似した乗降パターンを有していれば、グループとして抽出できることが本分析の結果から確認でき、クラスタリング手法としてのグラフ研磨の特徴を示すことができた。

また、単に類似した駅をグルーピングするだけでなく、分析結果として得られる研磨グラフから、a) 両駅間の延利用件数が統計的に有意に多くかつ他駅との乗降パターンも類似している組合せ、b) 他駅との乗降パターンが類似している組合せの 2 種類により駅間の類似性を考察できることを示した。このような研磨グラフの結果に対する解釈方法は、既往研究では言及されておらず、交通行動分析にグラフ研磨を適用したことによる本研究の新規性であるといえる。

さらに、本研究では、研磨グラフ構築時の閾値を、得られる研磨グラフの頂点数と辺数が概ね図-4 のネットワーク図と同等になるように設定した。グラフ研磨では類似度の閾値を低くすると、得られる研磨グラフの頂点数や辺数は多くなる。つまり、サンプルをどの程度のグループに分類したいかに応じて閾値を設定すれば、柔軟にグループ数を調整可能と考えられる。

本研究では、2 次元グラフに対してグラフ研磨を適用した。3 次元以上で表現されるグラフに対してのグラフ研磨の適用方法については既往研究で提案されておらず、3 次元以上でも適用可能な方法を構築することは今後の重要な課題である。

また、本研究では、乗降パターンが類似した駅のクラスタリングに留まっているため、乗降パターンが類似している要因、例えば駅周辺の人口特性や施設の立地状況との関連性を明らかにする必要がある。その上で、グルーピングした駅の利用者をターゲットにした公共交通の利用促進施策や端末交通手段の整備施策の検討を行うことも今後の課題である。

## <謝辞>

本研究の遂行にあたり、高松琴平電気鉄道株式会社様から、大変貴重なデータをご提供いただいた。ここに、感謝の意を表す。香川高等専門学校の宮崎耕輔教授、東京大学生産技術研究所の伊藤昌毅講師には、データの収集および処理に際して、ご尽力いただいた。記して謝意を表す。

【参考文献】

- 1) 前田侑亮, 金明哲 (2019), 「トピックモデルによる関西私鉄沿線の特徴分析」, 情報知識学会誌, Vol.29, No.1, pp.3-22.
- 2) 西田昌史, 石川勇樹, 山本誠一 (2011), 「非負値行列因子分解に基づく多人数会話における話者分類」, 情報処理学会研究報告, Vol.2011-SLP-85, No.7, pp.1-6.
- 3) 桑野将司, 木下礼央, 福山敬, 谷本圭志, 菅原一孔, 川村尚生 (2017), 「非負値テンソル因子分解を用いた検索ログデータからのバス利用特性の分析」, 土木学会論文集 D3 (土木計画学), Vol.73, No.5, I\_569-I\_578.
- 4) Sun, L. and Axhausen, K. W. (2016), "Understanding urban mobility patterns with a probabilistic tensor factorization framework," *Transport. Res. Part B*, Vol.91, pp.511-524.
- 5) 古屋秀樹, 劉瑜娟 (2016), 「潜在クラス分析を用いた訪日外国人旅行者の訪問パターン分析」, 土木学会論文集 D3 (土木計画学), Vol.72, No.5, I\_571-I-583.
- 6) 中原孝信, 羽室行信, 宇野毅明 (2015), 「グラフ研磨手法を用いた顧客の店舗選択モデル」, オペレーションズ・リサーチ, Vol.60, No.2, pp.89-95.
- 7) 中原孝信, グラフ研磨による考慮集合の予測と推薦システム, [http://www.orsj.or.jp/archive2/or61-11/or61\\_11\\_762.pdf](http://www.orsj.or.jp/archive2/or61-11/or61_11_762.pdf), 2020 年 4 月.
- 8) 前川浩基, 内田将史, 大内章子, 宇野毅明, 羽室行信 (2014), 「データ研磨手法を用いた Twitter ユーザの関係構造変化の検出」, 人工知能学会全国大会論文集, Vol.28, pp.1-4.
- 9) ことでんグループ, ことでん路線図, <http://www.kotoden.co.jp/publichtm/kotoden/time/index.htm>, 2020 年 4 月.
- 10) 石田基広 (2008), 「R によるテキストマイニング入門」, pp.84-85, 森北出版.
- 11) 宇野毅明, 中原孝信, 前川浩基, 羽室行信 (2014), 「データ研磨によるクリーク列挙クラスタリング」, 情報処理学会アルゴリズム研究会報告書, Vol.146, No.2, pp.1-8.
- 12) Uno, T., Maegawa, H., Nakahara, T., Hamuro, Y., Yoshinaka, R. and Tatsuta, M. (2016), "Micro-clustering: finding small clusters in large diversity", arXiv preprint arXiv:1507.03067v2, 24 pages.