

数理実習 B （離散手法）レポート

路線グラフに基づく日本の鉄道ネットワークの特徴抽出

03-180604 的矢知樹 kazu.8128.9.6@gmail.com

概要

路線のグラフを用い、日本の鉄道ネットワークの特徴を抽出した。特に、「日本の中心の駅はどこにあるか」という点に様々な指標を用いて中心の駅を求めた。今回の実験の結果、媒介中心性から東西をつなぐ要所の駅が、固有ベクトル中心性から人口、経済活動の中心を見出すことが出来た。また、頂点を駅ではなく路線とすることで頂点の数を容易に少なくすることができるという結果も得られた。

路線グラフの入手

駅データ.jp (www.ekidata.jp)に駅、路線のデータ、各駅にどの路線が通っているかのデータ、各駅が互いに乗り換え可能かのデータが公開されている。このデータを用いて日本の路線のグラフを構成することが出来る。

今後路線というときはこの駅データ.jpのデータにおいて一本の路線と記録されているものを指す

日本の駅は10000駅以上存在するが、路線の本数は600ほどである

路線グラフの頂点とエッジ

路線をグラフにする方法として次の2つが考えられる

1. 駅が頂点、直接路線によって繋がっている頂点同士を辺で結ぶ
 2. 路線が頂点、ある駅によって乗り換えられる路線同士を辺で結ぶ
- どちらのグラフも路線全体の構造を反映したものとなる。

ある頂点の離心数を「ある頂点からの他の任意の頂点への最短距離の最大値」として定める。
離心数の最大値を直径とよび、離心数の最小値を半径と呼ぶ。

路線が頂点のグラフ

まず、路線を頂点としたグラフを考える。頂点数が600であるため $O(n^3)$ アルゴリズムであるワーシャルフロイド法を容易に実行可能である。

直径は26であった。

離心数最大の路線のペアとして以下のものが得られた。

花咲線 阿波室戸シーサイドライン

花咲線 宿毛線

花咲線 西九州線(有田～伊万里)

花咲線 西九州線(伊万里～佐世保)

花咲線 長崎電軌4系統

花咲線 長崎電軌5系統

花咲線は根室本線のうち、釧路-根室間につけられた愛称である。宿毛線は高知と愛媛を結ぶ路線である。このように日本列島の南北に細長い形を反映して、北海道から九州、四国の間が乗り換え回数の観点からも最長であると言える。

実際に花咲線と阿波室戸シーサイドラインの間でどのような乗り換えを行うと、最も乗り換え回数が少ないかを出力すると次のものが得られる。

花咲線 JR 根室本線(新得～釧路) JR 根室本線(滝川～新得) JR 函館本線(小樽～旭川) JR 函館本線(長万部～小樽) JR 函館本線(函館～長万部) 道南いさりび鉄道線 JR 海峡線 JR 津軽線 JR 奥羽本線(新庄～青森) JR 羽越本線 JR 信越本線(直江津～新潟) 日本海ひすいライン あいの風とやま鉄道線 JR 高山本線 JR 東海道本線(岐阜～美濃赤坂・米原) 琵琶湖線 京都線 JR 神戸線(大阪～神戸) JR 神戸線(神戸～姫路) JR 山陽本線(姫路～岡山) JR 宇野線 瀬戸大橋線 JR 予讃線 JR 高德線 JR 牟岐線 阿波室戸シーサイドライン

離心数最小の路線は「あいの風とやま鉄道線」だった。これは花咲線と阿波室戸シーサイドラインを結ぶ最短路の中心に位置している。

半径 * 2 = 直径

という方程式を満たすため、路線網は非常に細長く出来ているということが分かる。

頂点数が600のためワーシャルフロイド法が動き、非常に解析が楽であった。

駅が頂点のグラフ

駅を頂点として同様に半径と直径を求める。
駅は 10000 ほど存在するため、ワーシャルフロイド法は動かないが、幅優先探索を用いることで、任意の二点間の距離を $O(n^2)$ で求めることが出来る。

グラフの直径は 666 であり、半径は 333 であった。

直径を与える 2 つの駅は枕崎と稚内であった。枕崎は鹿児島島の先端であり稚内は北海道の先端である。路線を頂点とした場合と同様に北海道と九州が離れていることを反映している。

また、半径 * 2 = 直径であるため、非常に細長いグラフであることが想定される。

以下日本の中心の駅について考察する

離心数最小

半径を与える駅は岐阜県の坂本駅であった。
右に離心数が小さい駅の表を載せる。なお二列目は最も離れた駅を表している。

この坂本駅は中心であると言えるが、稚内と枕崎のみに依存しているため、稚内付近と枕崎付近で少しグラフが変化することにより大きく中心が動いてしまう。

坂下	稚内	333
落合川	稚内	334
田立	枕崎	334
中津川	稚内	335
南木曾	枕崎	335
美乃坂本	稚内	336
十二兼	枕崎	336
恵那	稚内	337
野尻	枕崎	337
武並	稚内	338

近接中心性

図 1: 離心数の小さい上位 10 駅

離心数最小の欠点としてグラフの端付近の小さな挙動に大きく影響を受けるといものが挙げられる。
その欠点を解消するものとして次の指標を考える。

自分以外の駅への最短経路の長さの総和

この値が最小になる点を近接中心性と呼ぶ。今回の駅グラフで最も近接中心性が高い駅は愛知県の枇杷島駅であった。
右に近接中心性が高い駅の表を載せる。なお、グラフの二列目は最短距離の総和を表している。

枇杷島	1027822
多治見	1028395
岐阜	1029026
尾張星の宮	1029329
土岐市	1029569
清洲	1029658
長森	1029853
小泉	1030231
新那加	1030602
小田井	1030637

これは平均的に近い駅というのを表す指標になっている。中部地方の駅が上位に食い込むことは変わらないが、出現する駅は変わっている。

図 2: 近接中心性が高い上位 10 駅

駅グラフの場合、離心数最小、近接中心性、どちらも中心の駅自体に大きな特徴があるわけではない。これらの指標は駅を見る上では適していないであろう。

媒介中心性

駅自体に中心としての特徴が生じる指標として、媒介中心性が挙げられる。

ある点 i の媒介中心性とは、 i を通る最短経路の本数を、最短経路の本数で割ったものと定義される。

つまりその点を削除すると最短経路が伸びるような二点間の割合を示す。これはグラフの要所を表す指標であり鉄道網に対して適用するのに適している。

媒介中心性は素直に実装すると非常に計算量がかかるが Brandes のアルゴリズムにより $O(V * E)$ で計算することが出来る。今回は V, E ともに 10^4 であるため、十分に計算が可能である。

塩尻	0.440
恵那	0.432
多治見	0.432
土岐市	0.431
瑞浪	0.431
釜戸	0.431
武並	0.431
美乃坂本	0.431
中津川	0.431
落合川	0.431

図 3: 媒介中心性が高い上位 10 駅

前ページの表は媒介中心性の高い駅の表である。塩尻駅というのは非常に重要な駅である。

右に路線図を示す。

塩尻というのは中央本線などの主要な路線が交わる中部の駅であり、媒介中心性最大の駅は実際の駅としても非常に存在感のある駅であった。

また、媒介中心性が0.4以上と非常に大きい。これは西側と東川を最短距離で結ぶとほぼこれらの駅を通るということを意味している

固有ベクトル中心性

今までの指標は、全て中部地方が中心という結果が得られた。

これは人口などの配置を全く考えず、日本列島の形から想像される中心と非常に近い結果である。しかし、実際は日本の中心は東京である。そのような情報を鉄道網から入手できるだろうか



固有ベクトル中心性は、隣接行列の最大固有値の固有ベクトルである。（ペロンフロベニウスの定理より、隣接行列は全て成分が正であるため、最大固有値は唯一存在して、それに対応する固有ベクトルの成分は全て正となる）

行列を繰り返し掛けることで容易に固有値を得ることが出来る（これを少し改良したのがPageRankである）

右に固有ベクトル中心性が高い駅を並べる。最も固有ベクトル中心性が高い駅は東京であった。これは日本の鉄道網というのは東京の周りで非常に密に構成されているということを反映している。鉄道網が密であるということはそれだけ活発な経済活動が行われているということを意味している。

固有ベクトル中心性を見ることで鉄道網から人口、経済活動の中心を見出すことが出来た。

まとめ

様々な中心性の指標で駅グラフを解析したが、媒介中心性からは東西をつなぐ要所の駅が、固有ベクトル中心性からは実際に人口、経済活動が集中している地域を見出すことが出来た。この手法は日本ではない鉄道網や道路網などに対しても適用可能であると考えられる。

東京	0.407
大手町	0.279
品川	0.278
新橋	0.274
三越前	0.217
銀座	0.208
新御茶ノ水	0.204
神田	0.190
四ツ谷	0.179
有楽町	0.144
日本橋	0.135
千葉	0.132
日比谷	0.127
小川町	0.120
上野	0.119
東銀座	0.117
新宿	0.115
渋谷	0.114
秋葉原	0.113
武蔵小杉	0.112

図 4: 固有ベクトル中心性上位の駅

実行環境

Ubuntu 17.10 python3 (anaconda)