

ネットワーク解析のための TwitterAPI の試用

谷 研究室 岡田 沙季

概要

Twitter ネットワークの相関関係を解析する準備として、TwitterAPI を用いたアプリケーションを 2 つ試作した。

1 はじめに

近年、ブログや SNS(Social Network Service) をはじめとするソーシャルメディアサービスが大変注目を集めている。ソーシャルメディアとは、一般のユーザが情報を作り出し、発信していくメディアのことであり、その中でも特に Twitter が世界的に急成長を遂げている。現在 Twitter のユーザ数は 1 億 4 5 0 0 万に達し、大変多くの人々に支持されていることが分かる。

Twitter に限らず、近年ウェブの発展により、ネットワークが数理科学の研究の対象となっており、様々な性質が発見されている。その 1 つとして、スケールフリー性を持つネットワーク構造などが至るところで発見されている (例えば [1] を参照)。

定義

あるネットワークの頂点の次数分布 $P(k)$ がある定数によるべき分布に従うとき、このネットワークはスケールフリーであるという。

ここで、次数分布がべき分布であるとは、

$$P(k) \propto k^{-\gamma} \quad (\gamma > 0)$$

を満たしていることである。

スケールフリーの性質としては、一部のノードが他のたくさんのノードとエッジで繋がっており、大きな次数を持っている一方で、大多数のノードはごくわずかなノードとしか繋がっておらず、次数は小さいという点が挙げられる (図 1 を参照)。

Twitter のフォローの関係も一種のネットワークと見なせる。それらを可視的、定量的に捉えることが、本研究の動機である。本研究では、その準備として、Twitter API(Application Program Interface) を利用し、Twitter ネットワーク間の相関をネットワークグラフにより視覚的に表現するアプリケーションを開発した。

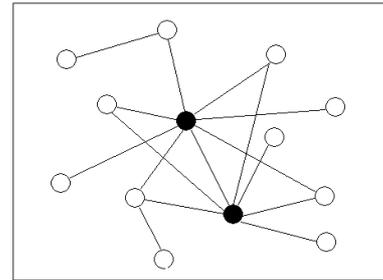


図 1 スケールフリーネットワーク

2 ソーシャルメディアについて

2.1 ソーシャルメディア

ソーシャルメディアとは、ユーザが情報を発信し、形成していくメディアのことで、以下の 3 つの特徴を持つ ([2])。

1. ユーザのコミュニケーションがコンテンツの主役である。
2. 登録制により、ユーザの同一性が保たれている。
3. ソーシャルグラフを持っている。

2.2 ソーシャルグラフ

ソーシャルグラフとは、ウェブ上における人間の相関関係、SNS でいえばユーザ同士のつながりの情報を図式化したものを言う ([3])。ソーシャルグラフは、構成要素のノード (個人) と関係性のエッジ (つながり) で説明される (図 2 を参照)。

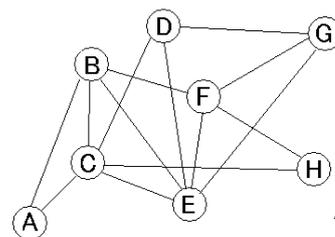


図 2 ソーシャルグラフ

2. 訪問時にキューから取り出し、次に行くことが出来る未訪問の頂点を全てキューに入れる。
3. キューが空になるまで 2 を繰り返す。

しかし純粋な幅優先探索すると、Twitter のネットワークにおいて、相互フォローしているユーザが存在した場合ループしてしまう。

そのため本研究で用いた実際の探索方法としては、Follow テーブルに、root(起点となるユーザ) を 1 人指定し、root のフォロー、フォロワー、root がフォローしているユーザのフォロー、フォロワーといった要領でデータベースにデータを格納していく方法である。

一度調べたユーザは調べないようにするため、Checked テーブルを用意し、既にフォロー、フォロワーを調べたユーザを格納していった。また Q テーブルという、Follow テーブルに格納されたユーザで、Checked テーブルに存在しないユーザを格納していくテーブルを用意した。root のフォロー、フォロワーを格納し終わったのち、Q テーブルから 1 番若い ID のものを取り出す。同様の方法で格納していき、Q が空になるまで探索を続け Follow テーブルにデータを格納していく。以上のような方法でデータを格納したデータベースから、任意のユーザ、取得したいネットワークの深さを指定し、可視化を行う。

また、JavaScript によるグラフの描画には、JavaScript ライブラリの canvasXpress のネットワークグラフ描画の機能を使用した ([4])。このライブラリを用いると、ノードとエッジのデータを引数に指定することで、ネットワークグラフを描画することができる。下記の画像は岡田の TwitterID からフォロー、フォロワーを取得し、深さ 1 までのネットワークを描画したものである (図 4 を参照)。

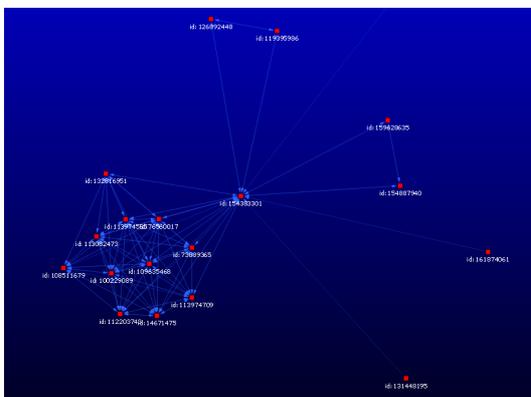


図 4 深さ 1 のネットワークグラフ

5 今後の課題

実際に可視化してみたものの、ブラウザに表示出来るデータの数には限界があり、Twitter のような大規模ネットワークの解析には、数値的に検証する必要があると分かった。

数値的に解析する方法の一つとして、Twitter のネットワークにおける同類選択性について解析することを考えた。

5.1 同類選択性

Twitter のフォローの仕方には大きく分けて以下の 2 つが挙げられる ([5])。

1. 友人、同業者、同じ趣味の人など比較的似た相手を見つけてフォローする。
2. 有名人、ニュース関係、企業の公式アカウントなど自分と異なる優れた、または小まめに情報を提供してくれる相手を見つけてフォローする。

以上の 2 つのフォローの仕方では、Twitter のネットワーク上では、どちらの方が優勢であるか検証する。

5.2 次数総相関係数

同類選択性を検証するにあたってネットワークの次数相関係数を求める必要がある ([6])。

$$r = \frac{\sum_i j_i k_i - M^{-1} \sum_i j_i \sum_{i'} k_{i'}}{\sqrt{\left[\sum_i j_i^2 - M^{-1} (\sum_i j_i)^2 \right] \left[\sum_i k_i^2 - M^{-1} (\sum_i k_i)^2 \right]}}$$

この式において、 j_i は入次数、 k_i は出次数、 M はノードの総数である。

r の値が正であると、同程度の次数の頂点同士が結ばれているようなネットワークとなり、同類選択性は高い。また r の値が負であると、大きい次数の頂点と小さい次数の頂点が結ばれているようなネットワークとなり、同類選択性は低くなる。

以上の検証方法で Twitter ネットワークを解析することが今後の課題である。

参考文献

- [1] 複雑ネットワーク入門 今野紀雄, 井手勇介著
- [2] ハチヨン dialog
<http://84dialog.blogspot.com/>
- [3] kotobank
<http://kotobank.jp/>

- [4] canvasXpress <http://canvasxpress.org/> mat/pdf/0209/0209450v2.pdf
- [5] cafebabe <http://d.hatena.ne.jp/cafebabe/20090930/1254341275>
- [6] Mixing patterns in networks http://arxiv.org/PS_cache/cond-
- [7] Twitter API プログラミング 辻村浩著
- [8] Twitter API リファレンスガイドブック 池田成樹著