

# SNS 上のフォロワー構造に基づく実際の支持度推定

小林 周平<sup>†</sup> 田島 敬史<sup>††</sup>

<sup>†</sup> 京都大学工学部情報学科 〒606-8501 京都府京都市左京区吉田本町

<sup>††</sup> 京都大学情報学研究科 〒606-8501 京都府京都市左京区吉田本町

E-mail: <sup>†</sup>skobayashi@dl.soc.i.kyoto-u.ac.jp, <sup>††</sup>tajima@i.kyoto-u.ac.jp

**あらまし** この論文では、単純なフォロワー数が支持を表す指標に直結するとは限らないことを示し、フォロワー同士の辺の量と、フォロワーのフォロワー・フレンド数を考慮することでより正確に支持度を推定する手法を提案する。提案手法を用いて分析を行ったところ、Twitter と Instagram の両方の情報を考慮して行った場合は単純なフォロワー数のほうが提案手法よりもうまく支持度の順位を表すコンテストの結果と一致していたが、Twitter の情報だけで行った場合は、提案手法のほうが単純なフォロワー数よりもコンテストの結果と高い一致を示すことができた。その実験結果と、Twitter のフォロワー構造が取得できた一方で、Instagram のフォロワー構造は制限により取得できなかったことを考慮に入れると、支持度は単純なフォロワー数だけでなく、そのフォロワーがどれだけ散らばっているのかということ、またフォロワーの価値にも影響を受ける可能性があるということが分かった。そして、Instagram のフォロワー構造も取得できれば、Twitter と Instagram の両方を使う場合にも、提案手法が単純にフォロワー数を用いる手法に比べて精度向上できる可能性があると考えられる。

**キーワード** SNS, Twitter, Instagram, 人気度, ソーシャルネットワーク分析

## 1 はじめに

近年、自己発信の容易化や自己発信を行える場の多様化、さらには新型コロナウイルスによる活動制限という社会的状況も相まって、多くのネットワークユーザが自分の趣味や嗜好などを発信するようになってきている。それに伴い、その発信された情報を受け取るユーザも多く存在し、いわゆるフォロワーとして、日々それらのコンテンツを楽しんでいる。そして、コンテンツの発信者がどれくらい人気があるのかを表す指標として、その発信者のフォロワー数が最も一般的な指標となっており、そのフォロワー同士のつながりの情報にはあまり触れられていない。

そのため、フォロワーの数が多いほうが支持度が高いと単純に判断することには疑問が生じる。例えば、あるユーザ A のフォロワーとユーザ B のフォロワーがともに 1,000 人であったとする。この時、A のフォロワーの 1,000 人は、実は実際の知り合いばかりで、フォロワー同士のつながりが密であり、一方で B のフォロワーは全くの他人でフォロワー同士のつながりが疎であったとする。この時、果たして A と B のどちらのほうが広く認知および支持されているかと考えると、B のほうが広く認知および支持されていると予想される。

本研究では、あるユーザの支持度をフォロワー同士のつながりの情報も考慮して推定する手法を提案する。つまり、フォロワー数がそのまま支持度に直結するのではなく、フォロワーのリンク構造を考慮することによって、より実情に沿った支持度を推定でき、フォロワー数が少ないユーザのほうが、支持度は高いと推定する場合がありうる。

提案手法では、あるユーザの 2 ホップフォロワーのグラフを

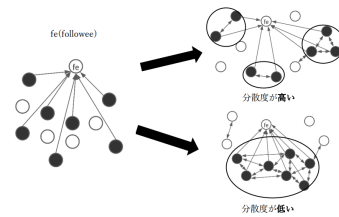


図1 **研究目的**. 本研究の目的は、フォロワー構造の違いに基づく支持度の推定である。図は、同じフォロワー数に対する散らばりの例を表すものであり、フォロワーの散らばりが右上のグラフのほうが大きく（疎）、右下のグラフのほうが小さい（密）ことを表している。

利用する。2 ホップフォロワーとは、ここではフォロワー、フォロワーのフォロワー・フレンドを意味すると定義する。すなわち、フォロワー、フォロワーのフォロワー、フォロワーのフレンド ID を取得し、それらを用いてあるユーザの 2 ホップフォロワーグラフを作成し、グラフの特徴を用いて支持度を推定する。ここで 2 ホップフォロワーを取得する理由としては、ユーザのフォロワー同士のつながりやフォロワーのフォロワー・フレンド数を提案手法では考慮するが、2 ホップフォロワーを取得することでそれらの情報を取得することができるからである。

具体的には、対象とするユーザについて、利用率が高い SNS である Twitter と Instagram のフォロワー情報を取得し、2 ホップグラフを作成した後、直接のフォロワーのクラスター係数、クラスター係数に類似する指標、およびつながりの多さを考慮した指標を計算し、それらをもとに個々のフォロワーのユーザに対する支持度合いをスコア化する。その後、全フォロワーについてのスコアの総和を対象ユーザのスコアとすることにより、

支持度を数値化する。

実験においては、対象とするユーザを9大学のミスコンテストエントリー者とした。その理由としては、各大学において、グランプリ、準グランプリという明確な順位が発表されるためである。また、ミスコンテストの順位は審査員等が独自に決めるのではなく、毎日のインターネット投票や当日投票などで、一般の支持に基づいて決定される。各大学においてそれぞれ単純なフォロワー数と前述したスコアを比較し、どちらのほうがより順位を当てているかを評価基準とすることにより、提案手法の評価を行う。結果として、TwitterとInstagram両方を考慮して行った提案手法では単純な合計フォロワー数より支持度の順位を正しく表すことができなかったが、Twitterのみで行った場合、Twitterのみの単純なフォロワー数よりも順位を正しく表すことができた。これは、Twitterのフォロワー構造は取得することができた一方で、Instagramのフォロワー構造は制限により取得することができなかったことを考慮すると、Instagramのフォロワー構造を取得することができれば、TwitterとInstagram両方を考慮して提案手法を実施した場合にも順位の予測精度を向上できる可能性があることが分かった。よって、提案手法を用いることにより、フォロワーの支持度をより正確に表すことができる可能性があることが分かった。

本研究の貢献としては、(1)単純なフォロワー数だけが支持度を表すのではなく、フォロワー構造も考慮に入れることにより、より正確な支持度を推定することができること、の提案と(2)フォロワー同士が疎であり、フォロワー自体の価値が高いほうがより支持度が高いこととの示唆の二つがあげられる。フォロワー間の辺の作成方法に、共通の趣味や嗜好を持ったユーザ間にエッジを加えるなどの工夫を加えることによって、さらに正確に支持度を推定できるのではないかと考えられる。

本論文の以降の構成としては、まず2章において関連研究について述べる、3章では提案手法の詳細、4章では実験の詳細を述べ、5章でまとめを行う。

## 2 関連研究

ユーザ同士のつながりの強さを表現する手法は、クラスター係数を基本とし、その拡張として様々な手法が提案されてきた。また、SNSアカウントの人気度予測についての研究もおこなわれている。本章ではそれらの関連研究を示す。

### 2.1 重み付きグラフに対するクラスター係数

隣人同士がどの程度つながりがあるかを示すクラスター係数は、ソーシャルネットワークを分析する上で使用される指標として代表的なもののひとつである。しかし、実際のソーシャルネットワークを分析する上では、ノード間のつながりの強さにはばらつきがある(メールのやり取りの量など)ことを考慮に入れる必要がある。そこで、クラスター係数の様々な拡張が研究され提案されており、Saramäkiら[1]によっていくつかがまとめられている。

それらのうち、Barratら[2]の提案したものは、頂点から出

る全ての辺の重みの合計(strength)に対して隣接する辺がどれだけ関連を持っているかを考慮したものとなっている。

Onnelaら[3]の提案したものは、重みの最大値で各辺の重みを正規化したのち、クラスター係数における三角形のすべての辺の重みの幾何平均をとったものを考えるという手法となっている。式では以下のように定義される。

$$\tilde{C}_i = \frac{1}{k_i(k_i - 1)} \sum_{j,k} (\hat{w}_{ij}\hat{w}_{ik}\hat{w}_{jk})^{1/3}$$

ここで、 $k_i$ は頂点*i*の次数を表す。 $w_{ij}$ はノード*i*とノード*j*の間の辺の重さを表し、 $\hat{w}_{ij}$ は、 $\hat{w}_{ij} = w_{ij} / \max(w)$ で定義される。 $\sum_{j,k}$ を用いることにより、隣接ノードと作成可能なすべての三角形について考慮する。式からわかるように、非常に小さい重みをもつ辺が存在する場合は、その三角形のクラスター係数に対する貢献度が非常に小さいものとなるようになっている。また、すべての隣接ノード間がつながっており、かつ辺の重さが重さの最大値と等しければ、クラスター係数は1となる。

これら二つ以外にも、Saramäkiらの論文には重み付きへと拡張したクラスター係数がまとめられている。

これらクラスター係数の拡張は、対象とする目的に応じて選択されるもので、どれが優れているということはない。本研究では、既存のクラスター係数とその拡張を用いており、各ノードについてクラスター係数を求めた後、その値を1から引き、得られた値を各ノードのスコアとしている。すなわち、各ノードのクラスター係数の値を求める段階でそれらを利用している。

### 2.2 有向グラフに対するクラスター係数

2.1章で述べた、クラスター係数の重み付きネットワークへの拡張に関する研究のほか、重み付き有向グラフに対するクラスター係数の拡張の研究を、Fagiolo[4]が行っている。FagioloはOnnelaら[3]が提案している重みの幾何平均の計算方法をもとに、有向グラフのクラスター係数と組み合わせ、有向重み付きネットワークに対するクラスター係数を定義している。有向グラフのクラスター係数は、重み無しの場合、作成可能な有向三角形の総数に対する、実際の有向三角形の数で定義される。それを重み付きに拡張する場合は、重み無しの際の実際の有向三角形の数の部分を、重み付きの代わりとなるものに置き換えるだけで実現することができる。

本研究では、Fagioloの定義に基づく有向クラスター係数をスコアを導出する過程で利用する。

### 2.3 意味的関係を利用した SNS 上のエンティティの人気度予測

SNSにおいて作成したアカウントが、将来的にどれだけの人気を得る可能性があるかを考えることは、企業戦略の観点などから重要な意味を持つ。大澤と松尾[5]が行った研究では、あるエンティティに対応するアカウントがSNS上に作成された際に、将来的にどの程度の人気を得る可能性があるかを知ることができるモデルを構築している。それを実現するために、大澤と松尾は意味的に関連するアカウントの特徴を用いて、調査の対象とするアカウントの将来的な人気度を予測している。

しかし、大澤と松尾の研究では、あるユーザの既存のアカウントを用いてそのユーザの現状の人気度を求めようとしているわけではない。したがって、既存のアカウントを用いて現状の人気度を求めようとしている本研究とは扱っている問題が異なっている。

### 3 提案手法

本章では、本論文で提案する支持度の推定方法に関して記述する。提案手法は、あるユーザについて

- (1) 2 ホップグラフの作成
- (2) 指標に基づくフォロワーのスコア化
- (3) ユーザの支持度のスコア化

から構成される。ユーザ ID、ユーザのフォロワー ID、フォロワーのフォロワー・フレンド ID を取得して 2 ホップグラフを作成し、指標を用いてユーザのフォロワーを 1 人ずつスコア化したのち、それらを合計してユーザの支持度をスコア化する。以下の章では、それぞれの構成要素について記述する。

なお、今後 2 ホップのグラフについての話が展開されるが、あるユーザの直接のフォロワーか、もしくはあるユーザのフォロワーのフォロワーかという点で、フォロワーという言葉が紛らわしいため、あるユーザの直接のフォロワーのことをダイレクトフォロワーと呼ぶことにする。また、今後フォロワーという場合は、相対的な 1 ホップ先のフォロワーノードを指すこととする。

#### 3.1 2 ホップグラフの作成

本研究では、Twitter と Instagram の 2 ホップフォロワー情報を利用すると 1 章で述べた。しかし、2 ホップグラフを作成するにあたって、Instagram の 2 ホップフォロワー情報は制限により取得が困難であるため、Twitter の 2 ホップグラフのみを作成した。インスタグラムの 2 ホップフォロワー情報をスコアに組み込む手法については、4 章で述べる。

本研究で注目しているのは、ダイレクトフォロワーの散らばり具合は支持度に影響を与えるのか、という点である。散らばり具合とは、ダイレクトフォロワー同士がどれだけつながっているのか、ということの意味する。多くのダイレクトフォロワー同士がつながっている場合、そのダイレクトフォロワー同士はグループを作っている可能性が高いため散らばり具合は小さくなり、一方でダイレクトフォロワー同士があまりつながっていない場合は、グループを作っている可能性が低いため、散らばり具合は大きくなる。さらに、個々のダイレクトフォロワーについて、フォロワーの数が大きいほど影響力が大きく、フレンドの数が多きほど興味の範囲が広くフォローしている対象ユーザと個人的な知り合いである可能性が低い。すなわち、ダイレクトフォロワーのフォロワー・フレンド数が多ければ多いほど、ダイレクトフォロワーの価値は大きくなる。したがってこれらの情報を考慮するために、ダイレクトフォロワー同士のつながり、および各ダイレクトフォロワーのフォロワー・フレンド数がどの程度存在するのかという情報が必要であり、2 ホップ先

のユーザ、すなわちあるユーザのダイレクトフォロワーのフォロワー・フレンドまでの情報を取得することによってそれを実現することができる。この時、Twitter の API を利用することによってフォロワー・フレンドのユーザ ID を取得でき、またユーザ ID は各ユーザに固有の値であるため、それをノードとすることによって 2 ホップグラフを作成する。有向辺は、あるユーザ  $u$  のフォロワーを  $v$  としたとき、 $v$  から  $u$  に対して張られる。この時、鍵アカウントについてはフォロワー情報を取得することができないため、鍵アカウントのフォロワーとフレンドは取得していない。

#### 3.2 指標に基づくフォロワーのスコア化

ダイレクトフォロワーの散らばり具合、すなわちダイレクトフォロワーが疎であるか密であるかを考慮する方法として、既存のクラスター係数およびその拡張を用いた手法、ダイレクトフォロワー同士の辺の量とダイレクトフォロワーのフォロワー・フレンド数を用いた手法を考案した。以下では、それぞれについて説明する。

##### 3.2.1 無向クラスター係数

ソーシャルネットワークにおけるクラスター係数 [6] は、あるユーザのフレンド同士がつながっている割合を図る指標であり、このクラスター係数が高いという性質は、ソーシャルネットワークにおいてよく見受けられる特性である [7]。しかし、ネットワーク全体の平均として高いとしても、個々のユーザに着目してみた場合、ばらつきがみられる。ここで、クラスター係数が高いユーザは同じグループ内でフォロー・フォロワー構造が形成されている傾向にあり、一方でクラスター係数が低いユーザは特定のグループに特化してフォロー・フォロワー構造が形成されているわけではない傾向にあるということが考えられる。したがって、ダイレクトフォロワーそれぞれのクラスター係数を 2 ホップフォロワーのグラフを用いて計算し、得られた値を見ることでそのダイレクトフォロワーがどの程度グループを形成しているのかを判別することができる。このクラスター係数のダイレクトフォロワーごとの違いをもとに、ユーザの支持度のスコア化を行う。

##### 3.2.2 有向クラスター係数

有向クラスター係数は、Fagiolo [4] が定義したものに従う。すなわち、すべての作成可能な有向トライアングルの数に対する、実際の有向トライアングルの数の割合で表す。これを用いることにより、有向クラスター係数の値が一方向的なフォローの場合よりも相互フォローであるほうが大きくなるため、無向の場合よりもより詳しく 2 ホップフォロワー構造の情報を扱うことができる。すなわち、よりつながりが強いダイレクトフォロワーがいるほどクラスター係数が高くなり、より強いグループが形成されていると考えることができる。ここでも、ダイレクトフォロワーごとに有向クラスター係数を導出し、それをもとにユーザの支持度のスコア化を行う。

##### 3.2.3 重み付きクラスター係数

重み付きクラスター係数は、Onnela ら [3] が定義したものに従う。すなわち、無向クラスター係数のうち、実際の三角形の

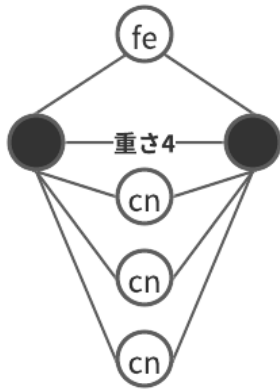


図 2 重み付きクラスター係数. fe が支持度を推定したい対象ユーザ、黒のノードがダイレクトフォロワー、cn がコモンネイバーを表す。コモンネイバーが 3 人存在し、かつダイレクトフォロワー間に辺が存在するため、辺の重さは 4 になる。

数の部分が辺の重みの幾何平均となったものを用いる。

あるダイレクトフォロワー 2 人のつながりの強さは、共通のフォロワー・フォロイーが多ければ多いほど強いと考えることができる。したがって、辺の重みの定義を、ダイレクトフォロワー同士の間の辺は、その 2 つのダイレクトフォロワーのコモンネイバーの数によって決定する。例えば、図 2 のように、コモンネイバーが 3 人の時、ダイレクトフォロワー間にも辺が存在すれば、その辺の重さは、コモンネイバーの数とダイレクトフォロワー間のつながりを合わせて 4 となる。ただし、コモンネイバーが何人いようとも、ダイレクトフォロワー間に枝が存在しなければ、そのダイレクトフォロワー間にはかわりがないため、辺の重さは 0 となり、つまり辺は存在しないとする。よって、

ダイレクトフォロワー間の辺の重さ

$$= \begin{cases} \text{コモンネイバーの数} + 1 & \text{辺が存在する} \\ 0 & \text{辺が存在しない} \end{cases}$$

と定義することができる。ユーザとダイレクトフォロワーの間の辺は、ダイレクトフォロワー間の辺の重さの平均値とする。辺の重さでダイレクトフォロワー間のつながりの強さは表現できるため、この指標ではユーザのノードとダイレクトフォロワーのノード以外のノードについては消去した後に、ダイレクトフォロワーのクラスター係数を計算する。その後、得られたダイレクトフォロワーごとの重み付きクラスター係数をもとに、ユーザの支持度のスコア化を行う。3.2.1 章と 3.2.2 章におけるクラスター係数の範囲と、3.2.3 章のクラスター係数の範囲の違いは、図 3 にまとめてある。

### 3.2.4 ダイレクトフォロワー同士の辺の量とダイレクトフォロワーのフォロワー・フレンド数を考慮した指標

あるダイレクトフォロワーについて、そのフォロワーが多いほど影響力が強く、フレンドが多いほどそのダイレクトフォロ

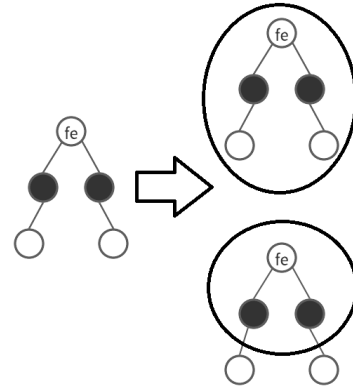


図 3 クラスター係数の範囲. fe が支持度を推定したい対象ユーザ、黒のノードがダイレクトフォロワー、白のノードがダイレクトフォロワーのフォロワーを表す。有向の辺も説明のため、無向にして扱っている。3.2.1 章と 3.2.2 章の定義では図の右上側である、2 ホップグラフ全体でクラスター係数の計算を行う。一方、3.2.3 章の定義では図の右下側である、ユーザとダイレクトフォロワーのみのグラフでクラスター係数を計算する。

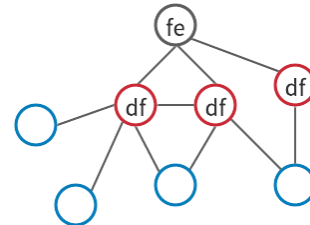


図 4 2 ホップグラフの図. 黒のノードが支持度を推定したいユーザ、赤のノードがダイレクトフォロワー、青のノードがダイレクトフォロワーのフォロワー・フレンドを表している。

ワーの興味の範囲が広く対象とするユーザと個人的な知り合いである可能性が低いと考えられる。したがって、ダイレクトフォロワーはフォロワー・フレンド数が多ければ多いほど支持度の価値が高くなることに貢献していると考えられる。ここで、支持度の価値とはダイレクトフォロワーの存在が持つ価値のことを表し、対象ユーザに対してダイレクトフォロワーが知り合いであり、影響力も小さいほど価値は低くなり、知り合いでなく、影響力が大きいほど価値が高くなるとする。例えば、同じフォロワー数であったとしても、友達ばかりのフォロワーよりも、つながりのない影響力の高い人々がフォロワーであるほうが、支持度は高いと予想されるからである。

ダイレクトフォロワーのフォロワー・フレンド数が多い一方で、同時にその他のダイレクトフォロワーとも密につながっていたとすれば、そのダイレクトフォロワーはグループを形成していると考えられ、ダイレクトフォロワーの散らばりは小さくなってしまったため支持度の価値は低くなってしまふ。例えば、あるダイレクトフォロワーについて、そのフォロワー・フレンドが 10,000 人いて、影響力が大きかったり興味の範囲が広いとする。しかしこの時、そのほかのダイレクトフォロワー 500 人と

つながっていたとすると、あるユーザを支持する上での散らばり具合としては、それほど高くないと考えることができる。図 4 でいえば、青のノードが多ければ多いほど、それとつながっているダイレクトフォロワーの影響力が大きかったり、興味の範囲が広がったりするため、支持度の価値は高くなる。一方で、赤のノード同士のつながりが多ければ大きいほど、グループを形成している可能性が高いため、支持度の価値は低くなる。このことを表現するために、青のノード数と赤のノード数を考慮した指標を用い、ユーザの支持度のスコア化を行う。

### 3.3 ユーザの支持度のスコア化

3.2.1 章から 3.2.3 章で述べた三つのクラスター係数については、ここでは同一の作業を行う。ダイレクトフォロワーについてのクラスター係数は、値が大きいほどグループを作っている傾向にあり、値が小さいほどグループがない傾向にあると以前述べた。そこで、クラスター係数の値が大きいダイレクトフォロワーほど、散らばり具合は小さいため、スコアは小さくなるようにしたい。ここで、3.2 章のいずれのクラスター係数も 0 から 1 の間の値をとるため、最大値は 1、最小値は 0 である。この時、

$$\text{スコア} = 1 - \text{クラスター係数}$$

とすれば、クラスター係数が高いダイレクトフォロワーほどスコアは小さく、クラスター係数が小さいダイレクトフォロワーほどスコアが大きくなる。すなわち、グループに属しているほどスコアを小さくし、グループに属していないほどスコアを大きくすることができ、ダイレクトフォロワーのグループが少ない、つまり散らばっているほどスコアは大きくなる。

3.2.4 章で述べた、ダイレクトフォロワー同士の辺の量とダイレクトフォロワーのフォロワー・フレンド数を考慮した指標については、青のノード数と赤のノード数を考慮した以下の指標を考案した。

スコア

$$= \frac{1}{\log_2(\text{隣接した赤のノード数} + 2)} \times \log_2(\text{隣接した青のノード数} + 2)$$

まず、隣接した赤のノード数 + 2 の対数の逆数をとることにより、隣接した赤のノード数が多ければ多いほどそのダイレクトフォロワーのスコアを小さくしている。次に、隣接した青のノード数 + 2 の対数をかけることにより、青のノード数が多ければ多いほど影響力が大きかったりユーザと個人的な知り合いである可能性が低いいため、支持度の価値も高くなるということを表現している。ここで対数をとっている理由は、隣接ノード数による違いを小さくするためである。例えば、赤ノード数の対数でなく逆数で考えると、隣接ノード数が 5 人の場合と 500 人の場合で 100 倍もスコアが変わってしまうため、望ましくない。このような事態を防ぐために、大きく変化が起こらない関数を利用しようと考え、したがって、対数をとることにした。また、それぞれの対数の中身に +2 をしているのは、隣接ノード数が 0 や 1 であった場合にスコアが発散して

しまったり、負の値をとることを防ぐためである。ここで、対数の底に 2 を選択しているのは、隣接ノード数が 0 の時、対数の値が 1 となるため、わかりやすいからという理由であり、異なる値を用いても問題はない。

それぞれの指標について、スコアをすべてのダイレクトフォロワーについて求めたのち、その和をとったものがあるユーザのスコアとなる。ダイレクトフォロワーのスコアの和をとることによって、単純なフォロワーの数の情報もスコアに組み込むことができる。鍵アカウントおよび頂点の次数が 1 であるアカウント (フォロワーが 0 人で、フォローしているのはユーザのみ) はクラスター係数が定義できない。よってクラスター係数が定義できないダイレクトフォロワーについては、クラスター係数が取得可能なダイレクトフォロワーのクラスター係数の 2 ホップグラフにおける平均をそのダイレクトフォロワーのクラスター係数としてスコアの導出を行う。3.2.4 章の指標についても、鍵アカウントおよび鍵アカウントではないが次数が 1 であるアカウントについては、隣接する赤のノード数と隣接する青のノード数のいずれも 2 ホップグラフにおけるダイレクトフォロワー当たりの平均値を使用する。ここで、鍵アカウントと鍵アカウントではないが次数が 1 であるアカウントを同様に扱っているが、これは鍵アカウントではないが次数が 1 であるアカウントが、フォロワー数と比べて非常に少ないことから、本来は変えるべきだが、今回の実験ではその影響は大きくないと考えられるためである。

## 4 実験

本章では実験の詳細を述べる。実験では、対象とするユーザとして 9 大学のミスコンテストエントリー者を扱った。また、本来は Twitter, Instagram 両方のグラフ構造を取得することが望ましいが、Instagram の 2 ホップフォロワー情報を取得することが制限により困難であったため、Twitter の 2 ホップフォロワー構造のみを取得した。Instagram のデータも考慮に入れるために、Instagram からはエントリー者のフォロワー数のみを取得し利用した。Twitter の 2 ホップフォロワー構造のみ取得されていることを考慮し、指標を計算するうえでは (1)Twitter+Instagram, (2)Twitter のみ、の 2 パターンの実験を行い、それぞれにおいて結果の評価を行った。

### 4.1 実験手順

実験は以下の手順で行う。

(1) Twitter からミスコンテストエントリー者、エントリー者のダイレクトフォロワー、ダイレクトフォロワーのフォロワー・フレンド ID を取得する。

(2) 取得した ID に基づき、2 ホップフォロワーのグラフを作成する。

(3) 3.2 章で述べた指標に基づき、各エントリー者のダイレクトフォロワーのスコアを導出し、Twitter のみのエントリー者のスコアを導出する。(Twitter だけで行う場合はこの時点でスコアの計算は終了し、5 へ移る。)



(4) Twitter のグラフ構造から得たスコアと Instagram のフォロワー数を用いて Instagram のスコアを導出し、Twitter と Instagram のスコアの合計値をエントリー者のスコアとする。

(5) 各大学ごとにグランプリ、準グランプリの正解率を本研究で提案するスコアを用いる方法と、単純なフォロワー数を用いる手法で比較する。

## 4.2 データセット

実験においては、駒澤大学、学習院大学、龍谷大学、中央大学、成蹊大学、東京大学、関西学院大学、同志社大学、関西大学の9大学のミスコンテストを扱う。各大学について、Twitter はミスコンテストエントリー者の ID、エントリー者のダイレクトフォロワーの ID、ダイレクトフォロワーのフォロワー・フレンドの ID を取得し、Instagram はフォロワー数のみを取得する。データの取得日時は、ミスコンテスト開催日当日および前日を基本としているが、一部数日ずれて取得している大学もある。しかし、当日のフォロワー数から大きく変化がないため、精度には影響がないものと思われる。

## 4.3 Twitter と Instagram のスコアの変換

Twitter の2 ホップフォロワー構造と Instagram の2 ホップフォロワー構造については、ある程度の類似性がみられると仮定したうえで、以下の式でスコアの変換を行う。

$$\begin{aligned} \text{Insta スコア} \\ = & \text{Twi スコア} \times \left( \frac{\text{Insta フォロワー数}}{\text{Twi フォロワー数}} \right) \times \left( \frac{0.106}{0.42} \right) \end{aligned}$$

Insta スコアは Instagram によるスコアを、Twi スコアは Twitter によるスコアを、Insta フォロワー数は Instagram のフォロワー数を、Twi フォロワー数は Twitter のフォロワー数を表す。Instagram のフォロワー数を Twitter のフォロワー数で割り、それを Twitter のスコアにかけることによって、Instagram のフォロワー数の大きさにスコアをスケールアップすることができる。しかし、そのままスケールアップしただけでは Instagram のダイレクトフォロワー同士のつながりの情報をほとんど考慮していないことになるため、0.106/0.42 のレートをかける。この数値は、Java ら [8] の研究と、Manikonda ら [9] の研究に基づくもので、Twitter のクラスター係数が 0.106 であり、Instagram のクラスター係数が 0.42 であることから導いている。クラスター係数が高いということは、ユーザの密度が高いということを示しており、したがってよりグループを形成しているということを意味する。したがって、Instagram のユーザ 1 人当たりの提案手法によるスコアは Twitter のものよりも小さくしたいため、Twitter のクラスター係数 0.106 を Instagram のクラスター係数 0.42 で割ったものを Twitter の提案手法によるスコアにかけることによって、Instagram の提案手法によるスコアと Twitter の提案手法によるスコアの調整を行っている。

## 4.4 評価指標

評価指標としては、正解率の表し方として、1, 2 位の組み合わせ的中、1, 2 位の順番も正しく的中、1 位のみを的中、1 位、2 位を別々に的中、の四つを採用している。1, 2 位を別々に的中とは、1 位、2 位が同時に当たっておらず、片方だけが当たっている場合も、1 つの中しただけとカウントする。すなわち、各大学について正解数の最大が 2 となる。それぞれの評価指標について、Twitter のみのフォロワー数、Instagram のみのフォロワー数、Twitter と Instagram の合計フォロワー数、スコアの四つの値を比較する。Twitter のデータだけで行われた実験の場合は、Twitter のみのフォロワー数とスコアの比較を行う。

## 4.5 実験結果

実験は、Twitter と Instagram の両方を考慮した場合と、Twitter だけを考慮した場合の 2 通りを行った。以下、4.5.1 章では Twitter と Instagram の合計についての結果を、4.5.2 章では Twitter のみで行った結果を述べる。

### 4.5.1 Twitter と Instagram の合計

実験結果は、表 1 から表 10 のとおりである。各表において、Twi の列には各候補者の Twitter フォロワー数を、Ins の列には各候補者の Instagram のフォロワー数を、合計の列には各候補者の Twitter のフォロワー数と Instagram のフォロワー数の合計を記している。また、s1 が無向クラスター係数によるスコアを、s2 が有向クラスター係数によるスコアを、s3 が重み付きクラスター係数によるスコアを、そして s4 がダイレクトフォロワー同士の辺の量とダイレクトフォロワーのフォロワー・フレンド数を考慮した指標によるスコアを表す。表 10 は評価指標について、フォロワー数と各スコアによる結果をまとめたものを表している。各スコアは小数点以下をすべて切り捨てしてある。スコアを表す表中において、黄色で塗りつぶしてある箇所が各手法において最もスコアが高かったことを表し、水色で塗りつぶしてある箇所が 2 番目にスコアが高かったことを表している。表 10 において黄色で塗りつぶしてある箇所は、各評価指標において、最も正解率が高いことを表す。

実験の結果としては、クラスター係数とその拡張を用いた指標の場合、組み合わせと順番では単純な合計フォロワー数と正解率が同じであったが、1 位のみと 1, 2 位別々では合計フォロワー数よりも正解率が劣っているという結果になった。また、ダイレクトフォロワー同士の辺の量とダイレクトフォロワーのフォロワー・フレンド数を考慮した指標の場合、4 つの項目すべてにおいて合計フォロワー数に劣るという結果となった。これより、Instagram の構造をとらずに Twitter と Instagram を同時に考慮して支持度の推定を行った場合、単純なフォロワー数と比べた場合よりもうまく支持度を表現することができないということが判明した。

駒沢	Twi	Ins	合計	s1	s2	s3	s4	結果
No1	1900	3385	5285	2446	2475	2743	5382	準グランプリ
No3	3126	4882	8008	3941	3953	4343	8186	グランプリ
No5	1371	2154	3525	1725	1734	1897	3579	
No7	1320	5172	6492	2246	2303	2604	5471	

表 1 駒沢:Twitter+Instagram

学習院	Tw	Ins	合計	s1	s2	s3	s4	結果
No1	2489	3354	5843	2880	2903	3317	6174	
No2	2577	4920	7497	3534	3540	3806	7740	準グランプリ
No3	2836	2283	5119	3051	3065	3402	5931	
No4	2694	3885	6579	3238	3265	3662	7108	
No5	3212	3912	7124	3832	3848	4197	8333	グランプリ

表 2 学習院:Twitter+Instagram

龍谷	Tw	Ins	合計	s1	s2	s3	s4	結果
No1	1680	1612	3292	1921	1924	2074	3394	
No2	2520	2585	5105	2947	2951	3166	5041	
No3	1114	1740	2854	1417	1419	1545	2290	
No4	1767	3614	5381	2399	2410	2659	4904	グランプリ
No5	1165	1781	2946	1458	1466	1604	2924	準グランプリ
No6	898	1881	2778	1190	1208	1360	3130	

表 3 龍谷:Twitter+Instagram

中央	Tw	Ins	合計	s1	s2	s3	s4	結果
No1	1785	6677	8462	2808	2932	3452	7249	
No2	1797	2500	4297	1882	1996	2415	5085	
No3	2813	14625	17438	5390	5558	6477	16749	グランプリ
No4	1720	1832	3552	1818	1879	2157	5245	準グランプリ
No5	1563	6092	7655	2592	2653	3077	5784	

表 4 中央:Twitter+Instagram

成蹊	Tw	Ins	合計	s1	s2	s3	s4	結果
No1	2921	2208	5129	3171	3182	3470	6456	
No2	2416	3144	5560	2868	2876	3186	5849	準グランプリ
No3	3411	5172	8583	4105	4128	4694	7672	グランプリ
No4	938	948	1886	1017	1025	1153	2306	
No5	1310	1816	3126	1440	1470	1755	3657	

表 5 成蹊:Twitter+Instagram

東大	Tw	Ins	合計	s1	s2	s3	s4	結果
No1	2941	1440	4381	2883	2919	3292	6588	
No2	4611	6315	10926	4758	4912	6186	11779	準グランプリ
No3	2096	1989	4085	2055	2091	2585	4471	
No4	3201	3130	6331	3573	3644	3900	11717	グランプリ
No5	2275	852	3127	2244	2252	2481	4875	

表 6 東大:Twitter+Instagram

関学	Tw	Ins	合計	s1	s2	s3	s4	結果
No1	3856	6937	10793	4988	5010	5591	9467	
No2	3721	5524	9245	4647	4657	5104	9705	
No3	869	1759	2628	1082	1127	1283	3022	
No4	4010	9503	13513	5779	5805	6392	10785	グランプリ
No5	1939	2471	4410	2263	2280	2545	4959	
No6	3660	9336	12996	5513	5529	5995	10381	準グランプリ

表 7 関西学院:Twitter+Instagram

同志	Tw	Ins	合計	s1	s2	s3	s4	結果
No1	4994	22417	27411	8561	8871	10620	25957	グランプリ
No2	1310	3513	4823	1833	1876	2178	4353	
No3	3785	11580	15365	5325	5474	6679	13470	
No5	2400	12889	15289	4747	4821	5623	10879	準グランプリ
No6	2471	5832	8303	3485	3514	3923	7215	

表 8 同志社:Twitter+Instagram

関西	Tw	Ins	合計	s1	s2	s3	s4	結果
No1	2919	4541	7460	3464	3497	4049	7902	グランプリ
No2	2366	2207	4573	2681	2688	2907	4923	
No3	3694	3182	6876	4164	4178	4491	8942	準グランプリ
No4	2378	3481	5859	2940	2959	3249	6395	
No5	1865	2911	4776	2329	2347	2584	4898	
No6	1662	1902	3564	1955	1963	2128	3696	

表 9 関西:Twitter+Instagram

	組み合わせ	順番	1 位のみ	1,2 位別々
twitter のみ	3/9	1/9	6/9	7/18
Instagram のみ	5/9	3/9	6/9	9/18
合計	5/9	3/9	7/9	10/18
s1	5/9	3/9	6/9	9/18
s2	5/9	3/9	6/9	9/18
s3	5/9	3/9	6/9	9/18
s4	4/9	2/9	6/9	8/18

表 10 評価:Twitter+Instagram

#### 4.5.2 Twitter のみ

表 11 から表 20 は、Twitter のみで指標を実行した結果である。以下の表におけるスコア s1 から s4 は Twitter と Instagram の合計に対して行ったものと同じ指標である。表 20 において

は、それぞれの提案手法による Twitter だけのデータを用いて行った際の正解率がまとめてある。

表 20 の結果より、ダイレクトフォロワー同士の辺の量とダイレクトフォロワーのフォロワー・フレンド数を考慮した指標によるスコアは、組み合わせ、順番、1, 2 位別々の 3 つの項目で、Twitter フォロワー数だけと比べて勝っており、1 位のみにおいては同じであるということが分かった。また、三種類のクラスター係数およびその拡張による指標を用いた場合は、Twitter フォロワー数だけの結果と全く同じであるということも分かった。したがって、Twitter だけで支持度の推定を行った場合、単純なフォロワー数よりもダイレクトフォロワー同士の辺の量とダイレクトフォロワーのフォロワー・フレンド数を考慮した指標を用いることにより、より正しくユーザの支持度を推定することができるということが分かった。このことより、今回の実験では性能向上が見られなかった Twitter と Instagram の合計に対する場合でも、Instagram の 2 ホップフォロワー構造さえ取得できればうまくいく可能性があると考えられる。また、指標の特徴である、隣接する青のノード数が多ければ多いほど、また隣接する赤ノード数が少なければ少ないほどスコアが高くなるということから、ダイレクトフォロワー同士が疎であり、ダイレクトフォロワーの価値が高いほうが支持度が高いということが示唆される結果となった。そのほか、表 20 より、クラスター係数およびその拡張による指標では、Twitter のみのフォロワー数と結果が同じであったことから、支持度をうまく表現することができない可能性が高く、3 つの手法にはそれほど差がないということが分かった。

駒沢	Twitter	s1	s2	s3	s4	結果
No1	1900	1687	1707	1892	3713	準グランプリ
No3	3126	2826	2835	3115	5872	グランプリ
No5	1371	1235	1241	1359	2563	
No7	1320	1129	1158	1309	2750	

表 11 駒沢:Twitter

学習院	Twitter	s1	s2	s3	s4	結果
No1	2489	2149	2167	2475	4607	
No2	2577	2385	2389	2568	5223	準グランプリ
No3	2836	2536	2548	2827	4929	
No4	2694	2374	2393	2685	5211	
No5	3212	2931	2943	3210	6373	グランプリ

表 12 学習院:Twitter

龍谷	Twitter	s1	s2	s3	s4	結果
No1	1680	1546	1549	1670	2732	
No2	2520	2341	2344	2515	4004	
No3	1114	1016	1018	1108	1642	
No4	1767	1582	1589	1754	3234	グランプリ
No5	1165	1052	1058	1157	2110	準グランプリ
No6	898	778	790	889	2048	

表 13 龍谷:Twitter

中央	Twitter	s1	s2	s3	s4	結果
No1	1785	1444	1508	1776	3729	
No2	1797	1393	1477	1788	3763	
No3	2813	2331	2403	2801	7244	グランプリ
No4	1720	1433	1481	1700	4134	準グランプリ
No5	1563	1306	1337	1551	2916	

表 14 中央:Twitter

成蹊	Twitter	s1	s2	s3	s4	結果
No1	2921	2662	2672	2914	5422	
No2	2416	2159	2165	2398	4403	準グランプリ
No3	3411	2969	2985	3395	5549	グランプリ
No4	938	810	817	919	1837	
No5	1310	1067	1089	1300	2709	

表 15 成蹊:Twitter

東大	Twitter	s1	s2	s3	s4	結果
No1	2941	2566	2598	2930	5863	
No2	4611	3536	3650	4597	8754	準グランプリ
No3	2096	1658	1687	2085	3607	
No4	3201	2865	2922	3128	9398	グランプリ
No5	2275	2050	2058	2266	4454	

表 16 東大:Twitter

関学	Twitter	s1	s2	s3	s4	結果
No1	3856	3430	3446	3845	6511	
No2	3721	3380	3388	3713	7060	
No3	869	716	746	849	2000	
No4	4010	3616	3632	4000	6748	グランプリ
No5	1939	1712	1725	1926	3752	
No6	3660	3354	3363	3647	6315	準グランプリ

表 17 関西学院:Twitter

同志社	Twitter	s1	s2	s3	s4	結果
No1	4994	4014	4159	4979	12170	グランプリ
No2	1310	1093	1118	1299	2596	
No3	3785	3004	3089	3769	7601	
No5	2400	2015	2047	2387	4618	準グランプリ
No6	2471	2184	2202	2458	4521	

表 18 同志社:Twitter

関西	Twitter	s1	s2	s3	s4	結果
No1	2919	2487	2511	2907	5674	グランプリ
No2	2366	2170	2176	2353	3985	
No3	3694	3420	3432	3689	7345	準グランプリ
No4	2378	2147	2160	2372	4670	
No5	1865	1671	1683	1854	3513	
No6	1662	1517	1523	1651	2867	

表 19 関西:Twitter

	組み合わせ	順番	1 位のみ	1,2 位別々
twitter のみ	3/9	1/9	6/9	7/18
s1	3/9	1/9	6/9	7/18
s2	3/9	1/9	6/9	7/18
s3	3/9	1/9	6/9	7/18
s4	5/9	4/9	6/9	10/18

表 20 評価:Twitter

## 5 ま と め

本研究では、単純なフォロワー数とは異なるユーザの支持度を表すための手法を提案した。ダイレクトフォロワー同士の辺の量とダイレクトフォロワーのフォロワー・フレンド数を考慮した提案手法を用いることにより、ダイレクトフォロワー同士のつながりの強さやダイレクトフォロワーの影響力の大きさ、興味の広さを考慮に入れることができ、それを確かめるために大学ミスコンテストのデータを用いて実験を行った。

実験では、あるユーザの支持度の順位を表すには、Twitter と Instagram の両方を合わせて考える場合は単純なフォロワー数のほうが正しく表現することができるが、Twitter だけで考えた場合は、提案手法を用いたほうが正しく表すことができるということが分かった。その結果と、今回の実験では Instagram の 2 ホップフォロワー構造を制限により取得することができなかったことを合わせて考えると、Twitter と Instagram の両方を合わせて考える場合でも、Instagram の 2 ホップフォロワー構造を直接扱うことができれば提案手法のほうが正しく表すことができる可能性があるということが考えられ、またダイレクトフォロワー同士が疎であり、ダイレクトフォロワーの価値が高いほうが支持度は高いということが示唆された。そのほか、クラスター係数およびその拡張の指標は、それほどうまく支持度を表すことができないということも示した。

今後の研究としては、2 ホップフォロワー構造を取得するこ

とができ、さらには支持度の順位もつくデータで実験を行うことにより、提案手法がより正確であることを示すことを計画している。また、ダイレクトフォロワーの散らばりと価値の影響が、どの程度大きいのかについても調べることも考えている。そのほか、ダイレクトフォロワーの価値の定義があいまいなため、明確に定義することも考えている。

## 謝 辞

本研究は、JST CREST (JPMJCR16E3), JSPS 科研費 21H03446 の支援を受けたものである。

## 文 献

- [1] Jari Saramäki, Mikko Kivelä, Jukka-Pekka Onnela, Kimmo Kaski, and János Kertész. Generalizations of the clustering coefficient to weighted complex networks. *Phys. Rev. E*, Vol. 75, p. 027105, Feb 2007.
- [2] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, Vol. 101, No. 11, pp. 3747–3752, 2004.
- [3] Jukka-Pekka Onnela, Jari Saramäki, János Kertész, and Kimmo Kaski. Intensity and coherence of motifs in weighted complex networks. *Phys. Rev. E*, Vol. 71, p. 065103, Jun 2005.
- [4] Giorgio Fagiolo. Clustering in complex directed networks. *Phys. Rev. E*, Vol. 76, p. 026107, Aug 2007.
- [5] 大澤昇平, 松尾豊. 意味的関係を利用した sns 上のエンティティの人気度予測. 人工知能学会論文誌, Vol. 29, No. 5, pp. 469–482, 2014.
- [6] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, Vol. 393, No. 6684, pp. 440–442, 1998.
- [7] Seth A. Myers, Aneesh Sharma, Pankaj Gupta, and Jimmy Lin. Information network or social network? the structure of the twitter follow graph. In *Proceedings of the 23rd International Conference on World Wide Web, WWW '14 Companion*, p. 493–498, New York, NY, USA, 2014. Association for Computing Machinery.
- [8] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: Understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, WebKDD/SNA-KDD '07*, p. 56–65, New York, NY, USA, 2007. Association for Computing Machinery.
- [9] Lydia Manikonda, Yuheng Hu, and Subbarao Kambhampati. Analyzing user activities, demographics, social network structure and user-generated content on instagram, 2014.