

Twitter を用いた相互作用の非対称性に関する分析

○石井充 (金沢工業大学)

An Analysis of the Asymmetry of the Interaction between Agents through Twitter

* M. Ishii (Kanazawa Institute of Technology)

概要— 組織や個人による意思決定等を分析するためにエージェントモデルが用いられることが多い。既存のエージェントモデルの多くにおいては、各エージェントが他のエージェントに与える影響は少なくとも1対1の関連性で見ると対称的であり、同じ強さの影響を相互に及ぼしあう。本論においては、この妥当性を検証するため、Twitterにおけるフォロワーとフレンドの非対称性に注目し、これがフォロワー数すなわち、個々の利用者の影響力の大きさにどのように依存しているかを分析した結果を提示する。

キーワード: 非対称性, 相互作用, 局所性, Twitter, ソーシャルネットワーキングサービス, 冪乗則

1 背景

組織や個人の意思決定は、その周囲との様々な相互作用の結果として生じている。この様相を分析するために、エージェントモデルを用いた数値的な分析が行われることがある。

一般に、各エージェントが全ての他のエージェントと相互作用を行う場合には、一般均衡理論により、一様な定常状態が存在し得るが、エージェント間の相互作用が局所的である場合^{1),2)}には、クラスター構造が準安定状態になることが少なくない³⁾。

相互作用が局所的なモデルのうちで、最も単純なもの、エージェントが格子状に配置され、最近接エージェントとのみ相互作用し得る、いわゆる Ising 型のものである。このようなモデルでは、すべてのエージェントが同じ役割を果たす。

しかしながら、現実には、多くの人とつながり影響力の大きい人物や組織と、周囲に及ぼす影響力が小さい人物や組織とが共存しているため、単純な Ising 型のモデルには限界がある。

このような問題を解決するため、Ising 型のモデルを拡張する試みもなされている。一般に組織や個人の周囲への影響力は一樣ではなく、冪乗則に従うことが知られている^{4),5)}。このことを利用して、Oh 等⁶⁾はオープンソースコミュニティの安定性を調べるためにエージェント間のつながりの大きさが冪乗則に従う場合を調べ、一定規模以下のコミュニティが不安定で停止しやすいことを示している。

しかし、これはあくまでも統計的性質として非対称性がマクロ的に取り込まれたモデルである。影響力の強いエージェント A と影響力の弱いエージェント B があるときに、その影響力の大きさは、つながるエージェントの数によって与えられる。この場合、B が A に与える影響は他の多数のエージェントの中の 1 人の与える影響にすぎず、結果として B が A に与える影響力は小さいというモデルになっている (Fig.1 参照)。

こういったモデルにおいては、エージェント間の 1 対 1 の相互作用という観点から見た場合、その相互作用の大きさは対称的である。すなわち、

双方ともに同じ大きさの影響を与え合っている。影響力の強いエージェントは、多くのエージェントとつながっているため、1 つのエージェントから受ける影響は、他の多くのエージェントからの影響に埋もれ、結果として相対的に小さくなるが、1 対 1 の関連性という意味では平等な形でモデル化されている。

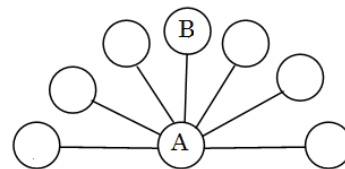


Fig.1: マクロ的に生じる非対称性

しかし、現実には、影響を及ぼす意思決定主体は相互の直接の面識がない場合も少なくなく、このようなモデル化は必ずしも妥当ではないと考えられる。有向グラフによる分析なども可能ではあるが、その場合には影響力は完全に一方だけに偏り、やはり妥当とは言えない。

そこで相互作用の非対称性を取り込むようにモデルを拡張しようとした場合、今度はどのような非対称性をモデルに導入するのが妥当であるかという問題が生じる。

この非対称性の問題に対処するため、現実の社会においてどのような非対称性が観測されるかを、代表的なソーシャルネットワーキングサービス (SNS) である Twitter を用いて調査したものが本研究である。以下では、2 章において、Twitter の特性と、それを用いた非対称性の取得方法に関する概略を述べ、3 章において解析の具体的な内容について説明する。4 章では解析結果をふまえ、得られた非対称性が、エージェントモデルを用いた各種のシミュレーションに対してどのような示唆を与えるかを述べた後、本研究の限界と今後の展望について述べる。

2 Twitterを用いた分析手法

Twitterは、少ない文字数で書き込み（ツイート）を行い、それをネットワーク上の多数のユーザに対して公開できるSNSである。

Twitterにおいては、アカウント間の関連性を表すものに、フォローというものがある。これは他者のツイートを自身の画面に反映させる行動である(Fig.2参照)。フォローをする際に、相手の許可を得る必要はない。すなわち、フォローは一方方向の行動である。

あるアカウントがフォローされている数、すなわち、フォロワーの数は、そのアカウントの持ち主が影響を与えられるアカウントの数を表す。フォロワー数が多ければ、そのアカウントの持ち主のツイートが多くの人に読まれており、その人物の発信力が大きいことを意味する。

これに対し、あるアカウントがフォローしている数すなわちフレンドの数は、個々のアカウントの持ち主が影響を受けるアカウントの数を表している。

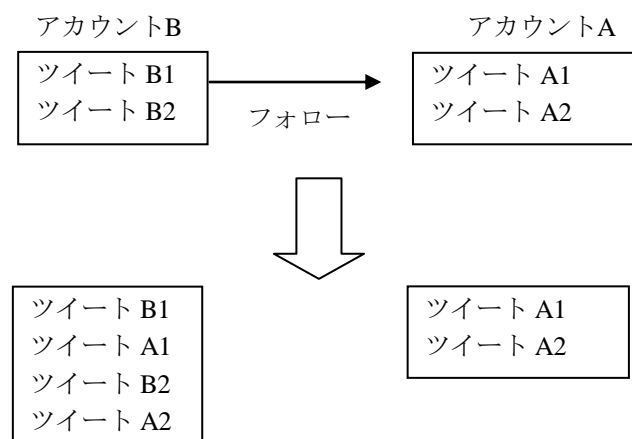


Fig.2: アカウントBがアカウントAをフォローしたとき、アカウントB上の画面（タイムライン）にはB自身のツイートとAのツイートの両方が表示される。

一般にフォロワーとフレンドの数は同じではない。例えば著名人や有名企業などが所有するTwitterアカウントは、フォロワーの数は多いがフレンドの数は少ない。このようなアカウントは多くの人に影響を与える反面、他者から受ける影響はほとんどないものとみられる。これは極端な事例ではあるが、基本的にフォロワーの数が多くフレンドの数が少ないアカウントほど、他者に与える影響が大きく、他者から受ける影響は小さいものと言える。

このような特徴を考えると、Twitterにおける多数のアカウントについて、

フレンド数/フォロワー数

の値を取得し、これがフォロワー数とどのような関係にあるかを調べれば、現実の世界における相互作用の非対称性に関して一定の見解が得られるものと期待される。

これまでに、SNSのユーザ特性に関する研究や、Twitterのユーザ特性に関する研究は存在したが^{7),8)}、フ

レンド数/フォロワー数がフォロワー数とどのような相関を有するかを調査した研究は存在しなかった。また、既存の研究はSNSの性質を調査したものであり、エージェントモデルにおける相互作用の非対称性を調べるという視点からとらえたものではない。

3 分析の詳細と結果

Twitterにおいては、各アカウントには固有の番号であるIDが割り振られている。そこで、我々は、乱数を用いてIDを生成し、多数のアカウントに関する情報を取得した。具体的には248万個の乱数を生成した。総じて、新しく作られたアカウントほど大きなIDを有しているが、IDは常に連続する数値になっているわけではない。このため、248万個の乱数のうちで、実際に存在するアカウントのIDは100万個（正確には1,000,1837個）であった。

この100万個のアカウントの中には、1回も書き込みを行っていないものが相当数存在する。これらは、まったく利用されていないアカウントであるとみなして、今回の調査対象から除外した。その結果、555,538個のアカウントが得られた。これを母集団として分析を行った。

あるフォロワー数を持つアカウントの数は、フォロワー数が大きくなるとともに減少する。このため、フォロワー数が大きくなると、そのフォロワー数を有するアカウントの数が1個か0個かのどちらかである場合がほとんどになる。例えば、フォロワー数が2000のアカウントが1個あり、フォロワー数が2001および2002のアカウントは0個であり、2003のアカウントは1個であるといったケースが出てくる。

本研究においては、フォロワー数と、そのフォロワー数を有するアカウントの数やフレンド数/フォロワー数の間に冪乗則が成立するかどうかを調べることが目的である。仮にフォロワー数 N_f とアカウント数 y_f との間に冪乗則が成立するならば、ある冪指数 α を用いて

$$y_f \propto N_f^\alpha \quad (1)$$

と書ける。この両辺の対数をとれば、

$$\log(y_f) = \text{const} + \alpha \log(N_f) \quad (2)$$

となり、両対数グラフの傾きから冪指数が得られる。

この両対数グラフを描くにあたっては、フォロワー数が多い領域と小さい領域とで同じ精度でデータを集計する必要はない。

そこで、フォロワー数が 10^4 以上 10^{n+1} 未満の場合には 10^n 刻みで集計を行うこととする。例えば、フォロワー数が100~999の場合、100~199, 199~200, ... の範囲内にあるアカウント数を数え、その総数を100で除算したものを平均アカウント数とする。

この結果、アカウント数等が1より小さい値となることがあり得る。例えば、アカウント数が0.1であれば、それは今回取得した総アカウント数の10倍の数のアカウントを取得すれば、平均的に1程度のアカウントがあるものと期待されるということの意味する。

Fig.3は、横軸にフォロワー数を、縦軸にそのフォロワー数を持つアカウントの数を両対数で示したものである。おおむね直線に乗っており、両者の相関は冪乗

則に従うと言ってよい。横軸が3の位置における直線の傾きは1.9であり、式(1)における冪指数 α は1.9であると言える。これは、ネットワークサービス分析において得られている結果である2前後の値と符号している⁷⁾。

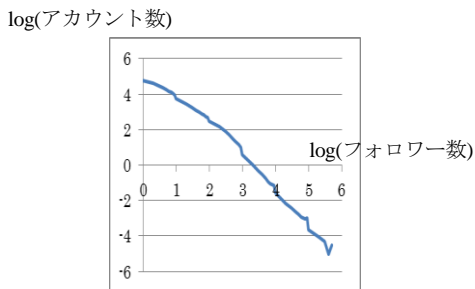


Fig.3: フォロワー数の分布

Fig.3の横軸が3のときの縦軸の値は0.6程度であるため、おおむね

$$\begin{aligned} \text{アカウント数} &\approx 10^{0.6} / (\text{フォロワー数} / 3)^2 \\ &\approx 36 / (\text{フォロワー数})^2 \end{aligned} \quad (3)$$

という関係式が成立する。

Twitterのアカウントには、最近の活動実態がないもの、つまりアクティビティの低いものがある。そこで、個々のアカウントについて、最も新しいツイートの日付を取得し、それが過去1週間以内であるアカウントのみを抽出して、Fig.3と同様の分析を行ったものがFig.4である。Fig.3と比較すると、フォロワー数が100以下のアカウントに関しては顕著にアカウント数が減少していることがわかる。既存の研究で冪乗則に従うと言われる多くの事象においては横軸の値が小さいときには冪乗則から外れることが知られており、Fig.4の事例は先行研究と矛盾するものではない。そもそも、これらのアカウントはツイート数が少なく、実質的に使われていないアカウントである場合が多いので、この領域でグラフが直線から外れても大きな問題ではないであろう。なお、本研究はアカウント間の影響力を調査することが主たる目的であり、アカウントのアクティビティは直接的には関係のない事項であるが、Twitterのユーザ特性に関する既存の調査においてはアクティビティに注目している事例が多いため、それらとの整合性があることを示すデータとしてFig.4を示したものである。

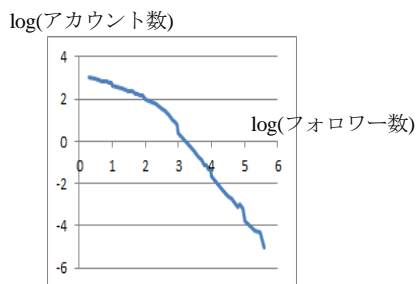


Fig.4: アクティブアカウントに対するフォロワー数の分布

Fig.5は、Fig.3と同じ母集団、すなわち、アクティビティを考慮することなく、ツイートが1回以上行われているアカウントを母集団として、横軸にフォロワー数を縦軸にフレンド数/フォロワー数を取り、両対数で示したものである。直線状の結果は、この振る舞いも冪乗則に従うことを示している。

直線の傾き、すなわち、式(1)の α に相当する冪指数は0.4の程度である。

これは、フレンド数すなわち他者から受ける影響は、フォロワー数すなわち他者に与える影響のおよそ平方根程度の大きさであることを示すものである。

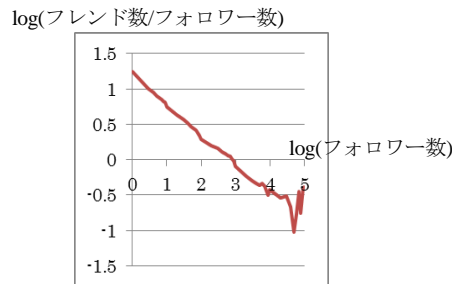


Fig.5: フレンド数/フォロワー数のフォロワー数への依存性

Fig.5において横軸が2の時の縦軸の値は0.3程度であるため、おおむね

$$\begin{aligned} \frac{\text{フレンド数}}{\text{フォロワー数}} &\approx \frac{10^{0.3}}{\sqrt{\frac{\text{フォロワー数}}{2}}} \\ &\approx \sqrt{\frac{8}{\text{フォロワー数}}} \end{aligned} \quad (4)$$

によって与えられると言える。

4 結論

本研究においては、エージェントモデルにおける相互作用の微視的な非対称性に対する知見を得るため、Twitterにおけるフレンド数とフォロワー数の非対称性に注目し、多数のアカウントを用いた調査を行った。その結果、フォロワー数とアカウント数の相関、およびフォロワー数とフレンド数/フォロワー数の相関はいずれも冪乗則に従うことが判明した。

フレンド数/フォロワー数とフォロワー数の相関を表す冪指数は、0.4程度であった。

この結果に基づくと、今後、エージェントモデルにおいて微視的な相互作用をもとにして数値シミュレーションを行う場合には、あるエージェントが n 個のエージェントに1の大きさの影響を及ぼすならば、 $\sqrt{8/n}$ の大きさの影響を個々のエージェントから受けるように非対称性を取り入れればよいことが示唆される。

ただし、本研究には限界も存在する。Twitterにおいて、あるアカウントをフォローしている人が多数いた場合、そのすべてのフォロワーが同じ強さの影響を受けているわけではないであろう。この問題を定量的に把握するためには、ツイートがどれだけ引用されたか、すなわちリツイート数が一定の尺度になると思われる。こういった側面を考慮して、より詳細な調査を行うことが今後に残された課題である。

また、本研究はTwitterに関する調査であり、その結果は、社会におけるすべての集団に対して汎用的に使えるとは期待できない。Twitterのようなネット上でつながる人間関係、すなわち、利害や依存関係が比較的小さい場合には本研究の分析が有効であろうと推測される。しかしながら、小規模な企業などの場合、1人の創業者が全てを決定するような意思決定機構となっている場合も少なくなく、そのような個別の集団に関して本研究の結果がそのまま適用できるわけではないことも予想し得る。

ただ、社会全体に広く行き渡る家電・携帯電話などの製品の購買に関する意思決定などを考える場合には、その意思決定に及ぼす影響は、特定の集団内だけの利害ではなく、個々人をとりまく広範な人間関係によって決定されるとみるのが妥当であり、本研究の結果がある程度有用な示唆を与えるものと期待している。

参考文献

- 1) N.Oomes: Local trade networks and spatially persistent unemployment, *J Economic Dynamics and Control*, **27**, 2115/2149 (2003)
- 2) A.V.Outkin: Cooperation and local interactions in the prisoner's dilemma game, *J Economic Behavior & Organization*, **52**, 481/503 (2003)
- 3) G.Topa, Social interactions, local spillovers and unemployment, *Rev. of Economic Studies*, **68**, 261/295 (2001).
- 4) A.L.Barabasi and R. Albert: Emergence of scaling in random networks, *Science*, **286**, 509/512 (1999)
- 5) M. Newman and J. Park: Why social networks are different from other types of networks, *Phys. Rev.*, **E68**, 036122, (2003)
- 6) W.Oh and S.Joen: Membership herding and network stability in the open source community: the Ising perspective, *Management Science*, **53**, 1086/1101 (2007)
- 7) Y.Ahn, et.al.: Analysis of topological characteristics of huge online social networking services, *Proc.16th international conf. on WWW, ACM*, 835/844 (2007)
- 8) Haewoon Kwak et.al.: What is Twitter, a Social Network or a News Media?, *Proc. 19th international conference on WWW, ACM*, 591/600 (2010)