

SNS 上の一方的な選好をもとにしたコミュニケーションの解析

Communication based on unidirectional preference in Twitter

浅谷 公威^{*1}
Kimitaka Asatani川畑 泰子^{*2}
Yasuko Kawahata鳥海 不二夫^{*1}
Fujio Toriumi坂田 一郎^{*1}
Ichiro Sakata^{*1}東京大学
The University of Tokyo^{*2}群馬大学
Gunma University

Homophily and preferential attachment explain a large part of the formation of online social network(OSN). However, a unidirectional preference for specific kind of person is not regarded to be an important factor in network formation. Such preference assumes to have important roles in OSN where no geographical restriction and high searchability exist. To observe such preference in the social network, we analyzed the user network constructed by interactions between who tweets about "leaving home" and who react to them. The network has a similar structure to a bipartite network which composed of the targeted group(who tweets "leaving home") and reaction group. Furthermore, about 20% of users in the latter group take one-to-many communication with the users in former group. We can assume a clear unidirectional preferences exist in them. In addition, we found that a large part of the tweets from users taking such one-to-many communication have the intention of invitation.

1. はじめに

オンラインソーシャルネットワーク (OSN) 上におけるコミュニケーションには時空間的な制約がなく、人物や投稿の閲覧/検索性も高い。そのような性質をもつ OSN 上では、マイノリティがコミュニティとしてつながり [3]、マスメディアが報じない重要な情報の拡散されている [4]。しかし、ソーシャルネットワーク上では従来にないリスクが表面化しつつある [8]。例えば、2017 年神奈川県で起きた自殺希望表明者の殺害事件や、未成年の誘い出し事例の一部は Twitter 上でのコミュニケーションが発端となっている。このようなコミュニケーションがどのような規模でどのような人々の間でどのように行われているかを解明することは、リスクの早期発見、教育による周知徹底、リスクを防止するシステムの設計に必要である。

OSN における会話のネットワークや友人ネットワークの構造は、Preferential Attachment(一部の人への人気の集中)と Homophily (同類選好) によりその多くが説明されてきた。Preferential Attachment とはネットワーク上の度数に比例した確率で該当するノードにリンクが貼られる性質である [1]。この性質は、Justin Bieber などの一部の人物が膨大な数のフォロワーを有していることを説明している。Homophily は似た人同士は好きになるという有向性のない概念であり、SNS 上の趣味・性的嗜好・職業などが同一である人々の結びつきを説明している。同類選好の強さは Assortative mixing[7] として指標化されており、多くの事例でその存在が確認されている。

しかしながら、これらのメカニズムには特定の属性を持つ人間への別の属性を持つ人間からの一方的な選好は想定されていない。例えば、家具を売りたい人々と買いたい人々の間や、家出希望を表明している(主に)女性達と家に泊めたい(主に)男性達との間には、特定の属性への別の属性からの一方的な選好が想定される。OSN はコミュニケーションに地理的な制約がなく、さらにハッシュタグ等で検索する機能が備わっているため、一方的な選好によるコミュニケーションが容易であると考えられる。Twitter であれば、特定のハッシュタグを持つ人を検索できるし、そのような人を起点として友人関係をたどる

ことも可能である。このように、オンラインコミュニティは一方的な選好によるコミュニケーションが行われやすい場所であると想定される。

近年、このような一方的な選好により起きたと考えられる OSN 上の事件が表面化しつつある。このような事例を理解するため、未成年の誘い出しの事例の分析や [2]、未成年がさらされているリスクのヒアリングによる調査 [5, 6] が行われている。これらの研究により、ミクロな視点での個別の事例および、それら個別のケースの集合により個人がリスクにさらされる頻度が分かりつつある、しかし、誘い出しなどのリスクが内在する OSN におけるコミュニケーションの全体像(どのような人同士でどのようにどのように行われているか、どれくらいの数の潜在的な加害者や被害者がいるか)は明らかではない。リスクの早期発見、教育による周知徹底、リスクを防止するシステムの設計のためにも、そのコミュニケーションの全体像を解明することは重要である。

本論文では、Twitter 上の一方的な選好によるおこるコミュニケーションのネットワーク構造を分析する。さらに、そのネットワーク構造から典型的な行動を行うユーザーを抽出する提案し、その方法により想定される典型的なコミュニケーションが抽出できることを示す。具体的には、家出の希望を表明する人々と何らかのリアクションを行う人々の間のインタラクションのネットワークを観察し、その中に一方的な選好があると想定されることを明らかにする。さらに、家出希望表明者の誘い出しにつながるような行動を取りやすいユーザーを抽出するため、典型的な行動をとるユーザーを会話ネットワークから抽出し誘い出しに繋がる行動をとっているかを分析する。

2. 手法

2.1 コミュニケーション構造の解析

はじめに、ハッシュタグに紐づくツイートの集合へのリアクションより、ユーザー間ネットワークを構成する。具体的には、特定のハッシュタグを用いて検索したツイートをシードツイートとし、そのツイートへのリアクション(リプライ、リツイート、ライク)を収集する。そのリアクションの元となる人から先となる人への関係性を有効リンクとしたネットワークを構成

する。

そのように構成されたネットワークはどのような構造になるだろうか。はじめに、ハッシュタグが#理科○類△組のようにコミュニティを代表する場合について考える。シードツイートをを行ったユーザー自身もシードツイートへのリアクションを行うことがあり、入次数も出次数も1以上のノードがある程度の割合で存在すると考えられる。次に、一般的なハッシュタグである「#○△を食べたい」というタグに対するリアクションを観察した場合、入次数のみが1以上と、出次数のみが1以上の2つのグループからなる2部グラフになると想定される。その2部グラフにおいて、「#○△を食べたい」という人を探し出してリアクションを行う人は少ないと考えられるので、出次数が2以上のユーザーは想定しにくい。上記とは異なり、「#家具を売ります」もしくは「#家出」などのツイートが一方的な選好を惹きつけた場合、あるユーザーからの不特定多数へのアプローチが想定される。このような場合、ユーザー間ネットワークは2部グラフになるとともに、出次数が2以上のユーザーが少なくない割合で存在していると考えられる。

以上の考察より、以下の構造のユーザー間ネットワークは一方的な選好を想定することでその形成過程を理解することが可能である。

- 内部のリンクが疎な2つの集合により大部分が構成され、片方の集合から他方の集合へ一方的に有向リンクが張られる
- 前者の集合のなかには、出次数が2以上のユーザー（後者の複数のユーザーにリンクを貼る）が少なくない割合で存在する

ハッシュタグをもとにしたリアクションより構成されたユーザー間ネットワークが以上の構造を持つかを測るため2つの指標を定義した。1つめの指標はリンクの非対称性である。ユーザーを、ターゲットのみのユーザー集合 $T = \{u \mid k_u^{in} > 0, k_u^{out} = 0\}$, リアクションのみをユーザーの集合 $R = \{u \mid k_u^{in} = 0, k_u^{out} > 0\}$, その他のユーザー O に分類する。 k_u^{in} は入次数, k_u^{out} は出次数を表す。リンクの非対称性指数 $A = \frac{N(R)+N(T)}{N(R)+N(T)+N(O)}$ となる。もう一つの指標として、リアクションのみのユーザー R からのエッジの平均本数を $ka_R = N(E_{ij} \{i \in R\})/N(R)$ と定義した。検出されたネットワークのリンクの非対称性指数 A が大きく、 R に属するユーザーからのエッジの平均本数 ka_R が大きい場合に、一方的な選好により構成されたネットワークであると想定可能である。

2.2 典型的な行動を行うユーザーの抽出

検出したユーザーのネットワークでリアクションを行ったユーザーの集合 R の中から、典型的な行動を行うユーザーを以下のルールベースのアルゴリズムによる抽出する。

まずはじめにアクティビティの高いユーザー群を抽出する。具体的には、複数のターゲットにリアクションしている、つまり出次数が k 以上のユーザーを抽出する。

さらに、そのユーザー群の中で典型的な行動を行っているユーザー群の抽出する。そのような人は、リアクションを行う先が他の人のリアクションを行う先とよく似ている。まず初めに、ユーザーのペア同士でリンク先ノードの Jaccart 係数を計算する。次に、その Jaccart 係数を重みとしたユーザー間のネットワークを構成し、そのグラフ内でのページランクの上位のユーザーを典型的な行動を行うユーザーとして抽出する。このようにして抽出されたユーザーは、周囲の人たちの平均的な行動特性を有している代表的なユーザーと考えられる。

3. データ

Twitter より 2017 年 8 月 1 日から 2018 年 2 月 20 日まで、「#家出#神待ち #泊めて #泊めてくれる人」それぞれのハッシュタグに関連するツイートと、それへの反応（リプライ、リツイート、ライク）を全て収集した。さらに、そのユーザーの一部に対しフォロー関係を取得した。

ハッシュタグ検索で取得したデータは、2,614 人からの 24,773 件の自殺関連のツイートとなった。1,341 人の 9,250 件のツイートに対し、5,307 人からのライクは 11,245 件、リプライは 5,306 件、リツイートは 894 件のレスポンスを取得した。

「#家出#神待ち #泊めて #泊めてくれる人」のツイートを行っているアカウントは女性と推定されるものが多く、それらに対してレスポンスを行うアカウントは男性と考えられるものが多かった。ネット上で人格を変える人や、複数のアカウントを使う人も想定できるが、本研究ではそのような人は少ないと想定し補正することなく分析する。

4. 分析

4.1 コミュニケーション構造の解析

入次数の分布

抽出したネットワークの出次数を横軸と入次数を縦軸にヒストグラムを作成した (図 1 左)。ヒストグラムの色はログスケールで表示している。出次数・入次数がともに1以上のユーザーの数は少ないことが分かる。このようなユーザー数は 428 人で、全体で 8,121 人のユーザーがいることを考えると、定義したリンクの非対称性指数 A は 0.947 となり、ネットワークは2部グラフに近い構造である。

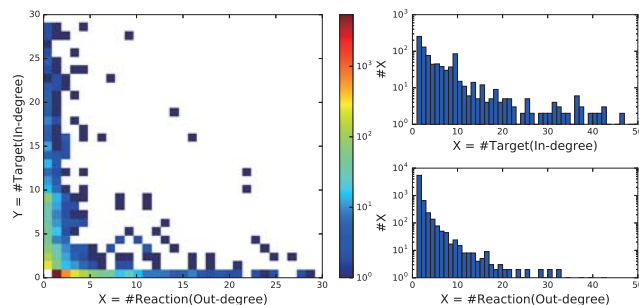


図 1: 左: ネットワークの出次数 (横軸) と入次数 (縦軸) のヒストグラム。右上: 出次数が0のユーザーの集合 R における入次数の分布。右下: 入次数が0のユーザーの集合 T における出次数の分布。

図 1 の右上に出次数0であるターゲットユーザー群 T に属する 913 人の入次数の分布を示している。このグラフから分かるように家出に関するツイートをしているユーザーでリアクションを受けたユーザーの多く (659 人) は、複数人からのアプローチを受けている場合が多いことが分かる。

図 1 の右下に出次数が1以上で入次数0のリアクションのみのユーザー群 R の 6,780 人のユーザーの出次数の分布をプロットである。入次数ほど顕著ではないが2人以上のユーザーにアプローチをしている人が少なくない割合で存在している。このように、複数の#家出ツイートをする人にアプローチをする人は 1,324 人であり、 R 中の 19% を占めている。また、リアクションのみをユーザーの集合 R の出次数の平均値 ka_R は 1.62 本となった。

フォロー関係の密度

ターゲットのみユーザー群 T 、リアクションのみユーザー群 R の間のフォロー関係を図 3 にまとめた。データ取得の関係上、次数 5 以上のユーザーのみを対象とし、7.6% の関係性のみをサンプリングした。

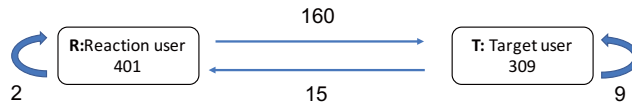


図 2: リアクションのみユーザー群 R とターゲットのみユーザー群 T との間のフォロー関係の数。次数 5 以上のユーザーのみに絞った。リンク数は 7.6% サンプリング。

図 2 から分かるように、リアクションのみのユーザー群 R およびターゲットのみユーザー群 T の間にはほとんどフォロー関係が存在しない。リアクションのみのユーザー群 R の人々はお互いに知り合いではないが、同じ集合に対してリアクションを行っていることが分かる。同様に、ターゲットのみユーザー群 T は R よりも明確ではないものの間でもお互いに知り合いではないと想定される。また、 R と T の間にも、7.6% のサンプリングで 160 の関係性（全体では約 2,100 本と推定）のエッジが存在しており、これは R に属するユーザーが T に属するユーザーを一方向的にフォローしたと考えられる。

コミュニケーション構造の解析より、家出に関するコミュニケーションの大部分が、リアクションを受ける側のユーザーの集合 T とリアクションをする R の 2 部グラフから構成され、 R のユーザーの中には複数人の T に属するユーザーにリアクションを行うユーザーが千人以上存在している。また、リアクションを受ける集合 T の中のフォロー関係も大きく疎であることから、その千人以上のユーザーがリアクションを送る先のユーザー同士が友人でコミュニティ内にいるとも想定し難い。このことより抽出したユーザーネットワークは手法で定義した背後に一方的な選好が想定されており、特に複数人の家出希望表明者へリアクションする千人以上のユーザーの行動原理には一方的な選好が想定される。

中心的なユーザー間の会話構造の可視化

コミュニケーションの構造を可視化するために、全てのノードの次数 k が 6 以上となる k -core[9] ネットワークを図 3 に可視化した。このグラフにおけるエッジの大きさは出自数（リアクションを取った回数）を示しており、色が明るいほど入次数が高い。また、リンク先のノードの近辺のリンクが太く表示されている。この図は、円の外側の出次数が高い（リアクションをする）ユーザーから内側に入次数が高い（家出をツイートする）ユーザーへの一対多の一方的なコミュニケーションがあることを示している。

4.2 危険度の高いと思われるツイートの特定

検出したデータの中に、誘い出しにつながりうるリプライ（誘い出しリプライ）が存在する割合を調べた。誘い出しリプライとは、Twitter のパブリックな場所以外でのコミュニケーションや現実世界での接触を促すツイートである。例えば、「家出したいです」というツイートに対し、「DM ください」（※ DM=ダイレクトメッセージ）や「〇〇（地名）なら大丈夫です」というツイートは誘い出しリプライとなる。それぞれ検出したサンプルから 150 を選び、各々のツイートがどの集合に属しているかをマスキングした上で、人間が読んで誘い出しリプライかどうかを判断した。

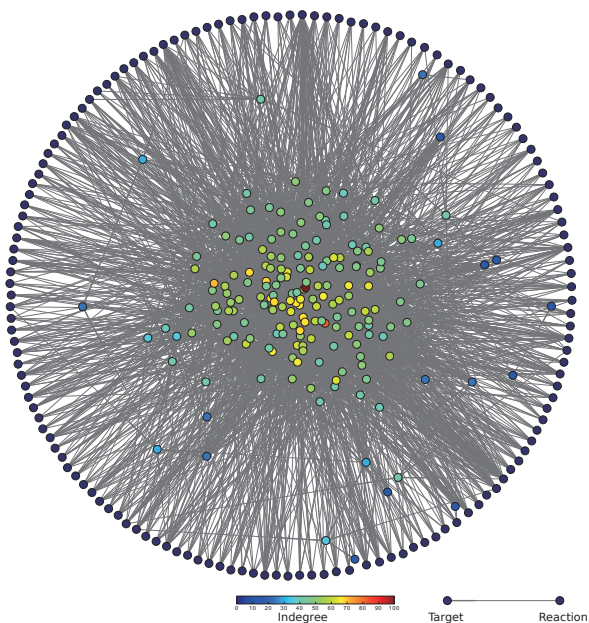


図 3: 家出に関するユーザー間ネットワーク。 $k = 6$ の k -core 部分のみを描画。ターゲットとなるノード側のエッジは太く、ノードの入次数が高いほど明るく描画。この図では中心部に入次数が高い（自らの家出に関するツイートに複数のリアクションを受けている）ユーザー、周辺部に入次数が小さい（リアクションをする）ユーザーが配置されている。また、周辺部のユーザーは複数の中心部のユーザーにリアクションを行っている。

表 1: リプライの検出数と誘い出しリプライの割合

	誘い出し	
	リプライの割合	検出数
全リプライ	0.58	5,306
高頻度ユーザー	0.69	1,996
典型的な行動のユーザー	0.73	668

表 1 は各ツイートの集合に対して誘い出しリプライの割合である。全リプライの対して 58% が誘い出しに関するものであり、単純に #家出などに対するリプライは高い蓋然性で誘い出しの危険を孕んでいることがわかる。また、高頻度ユーザーや本論文で定義した典型的な行動を行うユーザーからのリプライは誘い出しの危険性が高いことが分かった。また、ランダムに数十のユーザーをサンプリングしたところ、性別の判定は主観であるものの、家出したいとツイートしているのは未成年を含む若い女性でありリプライしているのは主に男性と想定される。

5. 考察

本論文ではソーシャルネットワーク上に一方的な選好による会話構造が存在していることをデータより裏付けた。そこでは、それぞれの間ではリンクが疎な 2 つの集合があり、1 つの集合から他方の集合への一対多のリンクが多く存在している。ある人から #家出をツイートした複数の人への一対多のリンクは、一方的な選好が背後にあると想定される。また、それぞれの集合のユーザー数はそれぞれ千以上であった。したがって、一方的な選好によるコミュニケーションは、ソーシャルネット

ワーク上のコミュニケーションを議論する上で無視できない規模で存在しているといえる。

また、単純に「家出」関連のツイートへのリプライを収集するだけで半分以上のツイートが誘い出しが想定された。また、家出をしたい人と、誘い出したい人がそれぞれ数千人の単位で存在していることは、Twitter 上での誘い出しリスクの大きさを物語っている。誘い出しの中では、直接的に特定の場所へ呼び出すものと近い頻度で、DM（ダイレクトメッセージ）でのコミュニケーションに誘導するものが多かった。DM はユーザーの権限で OFF にしたり、フォロワー以外からの DM のみ禁止といった設定が可能である。誘い出されるユーザーが任意のユーザーからの DM を許可した場合は、パブリックに閲覧できるリプライではなく DM でやりとりが行われていると考えられる。この DM に関しては本分析のスコープ外であるが、そこでどのようなやり取りが行われているかは分析を行う必要がある。本研究では Twitter の表層部に見える危険なやり取りを検知したが、Twitter 上にはさらに多くの危険が潜んでいると考えられる。

本研究では具体的に家出関連のハッシュタグをシードとして、一方的な選好によるコミュニケーションを検出した。家出以外にも自殺に関するツイートを中心としたネットワークは同様な構造になっていると想定され、このような一方的な選好をネットワーク構造から発見することは新たなリスクの予知につながると考えられる。したがって、将来的にはネットワーク構造のみから一方的な選好によるコミュニケーションが存在する部分集合を発見する技術の開発が必要である。

6. 結論

Twitter における家出に関するツイートとそれに対するリアクションを抽出し、ユーザー間のネットワークを分析することで、OSN 上に数千人単位の一方的な選好によるコミュニケーションが存在することを確認した。そこでは、お互いに関係が疎な 2 つのグループ間（誘い出す側と、家出を表明する側）で一方のコミュニケーションが存在する。さらに、前者のグループのユーザーの 2 割程が後者への一対多のコミュニケーションをとっており、明確な一方的な選好が存在することが想定される。また、そのような一対多のコミュニケーションをとるユーザーからのツイートは誘い出しの意図が想定される割合が非常に高いことが分かった。

7. 謝辞

本研究は RISTEX「未成年者のネットリスクを軽減する社会システムの構築」プロジェクトの支援を受けて実施した。

参考文献

- [1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.
- [2] et al. DeHart. Internet sexual solicitation of children: a proposed typology of offenders based on their chats, e-mails, and social network posts. *Journal of sexual aggression*, 23(1):77–89, 2017.
- [3] Keith N Hampton. Internet use and the concentration of disadvantage: Glocalization and the urban underclass. *american behavioral scientist*, 53(8):1111–1132, 2010.

- [4] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM, 2010.
- [5] Sonia Livingstone, Leslie Haddon, Anke Görzig, and Kjartan Ólafsson. Risks and safety on the internet: the perspective of european children: full findings and policy implications from the eu kids online survey of 9-16 year olds and their parents in 25 countries. 2011.
- [6] Giovanna Mascheroni and Kjartan Ólafsson. Net children go mobile: Risks and opportunities. 2014.
- [7] Mark EJ Newman. Assortative mixing in networks. *Physical review letters*, 89(20):208701, 2002.
- [8] Gwenn Schurgin O’Keeffe, Kathleen Clarke-Pearson, et al. The impact of social media on children, adolescents, and families. *Pediatrics*, 127(4):800–804, 2011.
- [9] Stephen B Seidman. Network structure and minimum degree. *Social networks*, 5(3):269–287, 1983.