

ハッピーキャリア SNS を利用したユーザと 言葉のネットワーク分析

中原 孝信, 大内 章子, 羽室 行信

ハッピーキャリア SNS は、われわれが開発・運用している SNS である。本稿では当該 SNS を利用しているユーザの関係性に基づく社会ネットワークの構築方法を示し、SNS を活性化させるためのキーパーソンをネットワークの中心性とコミュニティの観点から特定する。また、投稿された内容を把握するために、テキストマイニングの技術を利用し、言葉ネットワークを構築することで投稿内容の概要を把握する方法を提案する。本稿で扱うネットワークは小規模な社会ネットワークであり、そこからコミュニティやハブとなるユーザが捉えられることを示す。そして、最終的には得られた知見を実際の SNS の運用に活用することを目指している。

キーワード：社会ネットワーク分析, コミュニティ抽出, 言葉ネットワーク, クラスタリング, SNS

1. はじめに

世の中のあらゆるものは何らかの関係性をもち相互につながりをもつことが多い。インターネットの WWW, 送電網, 企業間取引, 知人関係はもとより、同時に購入される商品を対象にした同時購買関係を表すつながりもある。これらはすべてネットワークで表現することができ、行為者がどのようなものかによって対象とするネットワークが異なる。たとえば WWW では、節点がページで枝がハイパーリンクになり、知人関係は節点の人が、枝が友達関係になる。また、先の同時購買の場合は、節点が商品で枝が同時購入のされやすさになる。本稿で対象とするネットワークは、送電網のような物理ネットワークではなく、われわれが開発し運用しているハッピーキャリア SNS と呼ぶソーシャル・ネットワーキング上の利用者間のつながりである。

ハッピーキャリア SNS¹は女性の就業継続や管理職昇進など女性のキャリア形成を継続的に調査研究するために開発・運用している SNS であり、関西学院大学のハッピーキャリアプログラムで利用されている。ハッピーキャリアプログラムは、女性の仕事復帰、起業、そしてリーダー育成を目的とした社会人学び直しのプログラムで、当該 SNS のユーザはこれらのプログ

ラムの受講生とプログラムに関わる教員などの関係者である。

われわれが SNS を開発・運用している大きな理由は、利用者のログデータを分析し研究と教育の改善につなげていくためである。利用者からはユーザー登録時に研究利用でログデータを利用する旨の承諾は取っている。Facebook などの SNS は安定した機能を多くのユーザに提供しているが、第三者が運営する SNS ではログデータの取得は難しく、完全にデータを自由に扱えないという問題がある。また、教育の利用を考えると、Facebook では PDF ファイルや PPT フォーマットが利用できない点、スレッド構造として議論を蓄積できない点、同一アカウントで教育利用とプライベート利用を分けて使えないなどいくつかの問題がある。当該 SNS でも三つ目の点には対応していないが、ハッピーキャリアプログラムで利用していることから、教育用を前提に運用している。

本稿では、ハッピーキャリア SNS を利用し、次の三つの目的で社会的ネットワーク分析を実施した。

- 1) ネットワーク指標を利用した影響力の強いユーザの特定。
- 2) クラスタリングを利用したユーザの類型化によるコミュニティ抽出。
- 3) 言葉ネットワークによる意味抽出。

次節では、ハッピーキャリア SNS の説明を行い、3 節でネットワーク分析を実施し影響力の強いユーザの特定とコミュニティ抽出を行う。そして、4 節で言葉ネットワークを対象にした意味の抽出を行い、5 節で

なかはら たかのぶ

専修大学商学部

〒 214-8580 神奈川県川崎市多摩区東三田 2-1-1

nakapara@isc.senshu-u.ac.jp

おおうち あきこ, はむろ ゆきのぶ

関西学院大学経営戦略研究科

〒 662-8501 兵庫県西宮市上ヶ原一番町 1-155

ouchi@kwansei.ac.jp

y.hamuro@kwansei.ac.jp

¹ <https://happycareer-sns.jp>



図 1 フォーラムに作成されたトピックの例

は当該 SNS を利用したネットワーク分析のまとめを述べる。

2. ハッピーキャリア SNS

われわれが開発・運用している SNS は非常にシンプルで、「イベント情報」、「みんなの投稿」、「フォーラム」という三つのコンテンツからなる。SNS ユーザは会員登録することでこれらのコンテンツをすべて利用できる。一方で、会員登録していないユーザは一部のコンテンツの閲覧だけが可能である。

「イベント情報」はイベントを告知する仕組みで、開催地の地図表示や、出席者の管理などが可能である。「みんなの投稿」はブログ機能で登録ユーザは当該 SNS サイト上にブログを投稿することができる。「フォーラム」は、SNS に会員登録したユーザの中で編集権限をもつユーザが作成できるコンテンツであり、そのフォーラム内にトピックを作成し利用する。

トピックは一種の掲示板で登録ユーザは誰でもトピックを作成することができ、そのトピックに関するコメントを投稿することで意見を述べる。たとえば、図 1 は、「SNS サイト Q&A」というフォーラムに作られた三つのトピックを示している。「返信コメントでのリンクや画像貼り付けについて」というトピックを見ると、トピック作成者は大内章子氏で三つのコメントがあり、いいね数は 0、閲覧数が 96 回であることが確認できる。また図 2 は、コメントの一例でありアクセスログ解析というフォーラムのトピック「2017 年 6 月のアクセスログ解析結果」に投稿されたコメントである。コメントの掲載順は新規投稿が上になるように降順で並んでいる。

SNS の本格的な運用を開始したのは 2017 年 1 月からで、本稿で利用するデータは 2019 年 6 月までの 2 年半を対象にしている。2019 年 6 月の段階で登録ユー



図 2 コメントの例

表 1 当該 SNS の基礎集計の結果

項目	値	標準偏差
月間平均アクティブ ユニークユーザ	26.60	11.28
月間平均アクティブ ユーザ	607.36	298.34
月間平均ページ閲覧数	8296.80	6098.93
総フォーラム数	96	
総トピック数	474	
総コメント数	1633	

ザ数は 328 人であった。ハッピーキャリアプログラムの受講者からは、SNS 開設年度以降に入校した受講者全員と開設以前の修了生からは希望者の計約 230 名が登録している。それ以外に、社会人大学院のいくつかのクラスで SNS を利用した授業を実施しており、それらの学生が約 60 名で、残りは教員や関係者である。

表 1 は当該 SNS を対象にした基礎集計値を示している。各値は表のとおりであるが、いずれも標準偏差は高くなっている。これは 2018 年 6 月、7 月、8 月に SNS の活性化を目的に投稿祭りというイベントを実施したことで投稿数が大幅に増加し、それに関連してアクティブユーザやページ閲覧数などもそれ以前に比べて約 7 倍増えたことが関係している。

次節では、フォーラムに作成されたトピックとコメントを分析対象としネットワーク分析について説明する。

3. ネットワーク分析

ハッピーキャリア SNS は教育と研究の両輪を目指



図3 トピックとコメントの関係から構築したネットワーク

して運用している SNS であり、教育の面では、ソーシャル・ラーニングの実践として、SNS を通じた意見交換の場を提供し、発言を助長することで経験の共有と知識の共創を推進し学習効果を高めることを目標にしている。当該 SNS を利用した教育効果については、Nakahara et al. [1] で報告している。また研究の面では、この場でより活発な意見交換を生み出し、議論を通じて最終的には女性の働き方に関する施策の提言につなげることを目標にしている。

これらを具体的に実践していくためには、第一にネットワーク構造を把握し人同士のつながりをもとに活性化につながる施策を発見することが重要となる。そのために当該 SNS のトピック作成ユーザとそのトピックにコメントを投稿するユーザのつながりから SNS のネットワーク構造を把握する。そして、分析目的の一つ目である 1) ネットワーク指標を利用した影響力の強いユーザの特定を行う。

3.1 影響力の強いユーザの特定

全体で 474 トピックがあり、それらのトピック作成ユーザ数は 112 人、それらのトピックにコメントしたユーザが 221 人でユニークなユーザは 192 人である。これが節点数である。そして各トピックの作成ユーザとそのトピックに一度でもコメントしたユーザを枝で結び図 3 のネットワークを構築した。

この図 3 は有向グラフであり、細い枝はトピック作

表 2 当該 SNS の基礎集計の結果

節点番号	次数	クラスタ係数	媒介中心性
56	90	0.0294	2259.8302
133	26	0.0996	271.5222
13	20	0.1170	224.2976
152	20	0.1838	255.5060
101	20	0.0526	149.0000
144	19	0.2426	122.6321
64	19	0.2206	136.3111
162	17	0.2095	16.9167
178	16	0.3077	0.0000
124	16	0.1667	54.7690

次数の多い上位 10 節点を選択している。

成ユーザに入る枝で、太い枝はトピック作成ユーザから出ていく枝を表している。たとえば、一番左にある離れた節点群は太い枝がないので、トピック作成ユーザはほかのトピックに一度もコメントを書いていないことがわかる。また多くの節点は入る枝をもたない節点である。これはコメントを投稿しているだけで、トピックを作成していないユーザが多いことを示している。グラフの中心に多くの枝が入り出している節点があるが、これはハッピーキャリアプログラムを立ち上げた教員の節点であり、表 2 の節点番号 56 の次数を確認すると 90 本の枝とつながっていることがわかる。

この表 2 は、次数の多い上位 10 節点を選択し、次数、クラスタ係数、媒介中心性の値を示している。ク

クラスタ係数は、隣接する節点同士が互いにつながりをもつ程度を表している。たとえば、4 個の節点と接続する場合、その 4 個の節点が完全にお互いにつながっていると 6 本の枝がある。しかし実際にはすべての節点同士がつながっているとは限らないので、それら節点間の枝数は 6 本より少なくなる場合がある。クラスタ係数は実際に接続されている枝と完全グラフの枝の割合を示した値であり、仮に実際の枝が 4 本あったとすると、クラスタ係数は $4/6$ で 0.67 になる。対象は有向グラフであるが、クラスタ係数は無向グラフとして計算しており、双方向の枝がある場合は 1 本の枝として扱っている。

媒介中心性は、ある節点がほかの節点間の最短経路上に位置する程度を中心性指標とした値である。ある二つの節点間の最短経路数の中で、対象節点を通る最短経路の数の割合を計算し、それをすべての節点間で合計したものである。媒介中心性の高い節点は節点間の移動において中継地点になる重要な節点と考えることができる。

節点 56 の次数と媒介中心性は、ほかに比べてかなり高い値になっているが、クラスタ係数はほかの節点のほうが高いものもある。これは、節点 56 と隣接する節点が多いが、隣接節点同士はそれほどつながっていないことを示している。一方で節点 178 は、クラスタ係数の値が 0.3077 で、隣接節点間のつながりは完全グラフのおおよそ 3 割程度つながっており、節点 178 を中心に密な枝の関係がある。

この SNS を活性化させるためには、ネットワークのハブになるこれら 10 人のユーザを中心にして、隣接節点のつながりをさらに増やしていくことが効率的である。また、図 3 から明らかになった問題点として、多くのユーザがトピックを作成していないこと、そして、既存のトピック作成者がほかのトピックにコメントを投稿していない点がある。

これらの点を改善していくためには、トピック作成に対するハードルを下げて、コメントを投稿しやすい場を提供していく必要がある。そのためにはコミュニティを特定し、誰と誰が同じコミュニティに属するのかを見つけ出し、その関係性から発言を助長するような施策を行うことが有効であると考えられる。自分の意見を伝える際に、見知らぬ人に対して発言するよりも、仲間内で発言するほうが抵抗感は低くなるであろう。

次節では、コミュニティを見つけ出し、誰と誰が同じコミュニティに属しているのかを分析する。そ

して、異なるコミュニティ間をつなぐためにブリッジの役割をしているキーパーソンを発見する。

3.2 コミュニティー抽出

次に二つ目の目的である 2) クラスタリングを利用したユーザの類型化を目的にコミュニティ抽出を行う。

図 3 のネットワークに Newman and Girvan のクラスタリングアルゴリズム [2] を適用しコミュニティ抽出を行った²。コミュニティは、辺の媒介中心性を計算し最もそれが高い辺を取り除くことでネットワークを分割し生成される。ある辺の媒介中心性とは、その辺を通る 2 頂点間の最短経路数のことである。その際の分割数は、modularity が最大となるように決定される。modularity は、分割されたグループ内に含まれる節点間の枝数とネットワーク全体の枝数に対する割合から枝がランダムに配置された場合の期待値を引いた値である。

図 4 は、コミュニティ抽出の結果であり、九つのコミュニティが抽出できている。各コミュニティの内容を確認すると、ハッピーキャリアプログラムの期生ごとのコミュニティがいくつか構築されていた。これは各期生によって作成されたフォーラムがあり、そのフォーラム内だけで投稿が行われていることが関係している。

Granovetter [3] は、弱い紐帯（枝・リンク）の強さの重要性を述べており、職を見つけたり、情報を得たり、流行を生み出したりする場合は、強い社会的ネットワークよりも弱い社会的ネットワークが重要になると主張している。一般的にコミュニティ内でのメンバー同士は別のコミュニティのメンバーよりも強いつながりをもっており、同質なメンバーが集まっている。そのために、コミュニティのメンバー内だけのつながりでは、新しい情報が入手しづらくなったり、意見の均質化が起こったりするであろう。そこで、コミュニティ間を結ぶような弱いつながりが重要になる。そして、そのようなコミュニティ間を橋渡しする弱い紐帯（枝）は、ネットワークの局所間をつなぐブリッジの一種になる。

表 3 は、異なるコミュニティとつながっている枝をもつ節点の中から上位五つの節点を選んでいる。図 4 にもこれら 5 節点を示している。節点番号 56 は、次数が最も多くネットワークの中心となっている節点であり、ほかのコミュニティとの枝数も多くなって

² 計算は NYSOL (<https://www.nysol.jp>) の `mnewman` を使っており、これは R の `igraph` に含まれる関数 `cluster_edge_betweenness` が内部で動いている。



図4 Newman クラスタリングによるコミュニティ抽出

いる。また節点番号 46 は、10 期生コミュニティのメンバーであるが、ほかの期のメンバーともつながりがある。

当該 SNS 活性化のためには、コミュニティ内のメンバー間で発言を促す施策をまずは実施し、より密なコミュニティに構造を変化させていく。そして、コミュニティ間のつながりをより強固にしていくために、ブリッジの役割を担う 5 人のユーザにアプローチをしていくことが重要であろう。

4. 言葉ネットワークによる意見抽出

当該 SNS で議論されている内容から施策に活用できるような情報を抽出するためにテキストマイニングを利用し言葉ネットワークを生成した。言葉ネットワークは文書内に出現する格フレームからクラスタを生成し、そのクラスタの共起関係を基に生成する。

これまでテキストマイニングなどで文章を表現するために数多く利用されてきた方法の一つは bag-of-words (BOW) [4] であり、それは単語の出現だけを考慮したベクトルで文章を表現する方法である。BOW による表現は非常にシンプルで、ときに有用な結果をもたらすが、単語の出現順序や文章の構造を無視しているた

表3 異なるコミュニティとの枝数の高い五つの節点

節点番号	枝数
56	40
64	5
178	4
46	4
96	3

め、文章の意味を表現する場合にはその点が問題となる。一方で、格フレームは、ガ格やヲ格などの格助詞句と、動詞や形容詞などの用言句のペアによる表現で、「仕事を、楽しむ」「父が、経営する」など、格フレームによって文章の意味を表すことができる。

本稿では、当該 SNS で投稿されたコメントから格フレームを抽出するために、日本語の自然言語処理ソフトである KNP [5] を利用し、格解析を実施することで格フレームを抽出した。そして、抽出した格フレームを構成する格助詞句と用言句に枝を張ることで格フレームネットワークを構築した。格フレームネットワークは 20 以上のコメントがあったトピックだけを選択し、対象となる各トピックごとにネットワークを生成した。

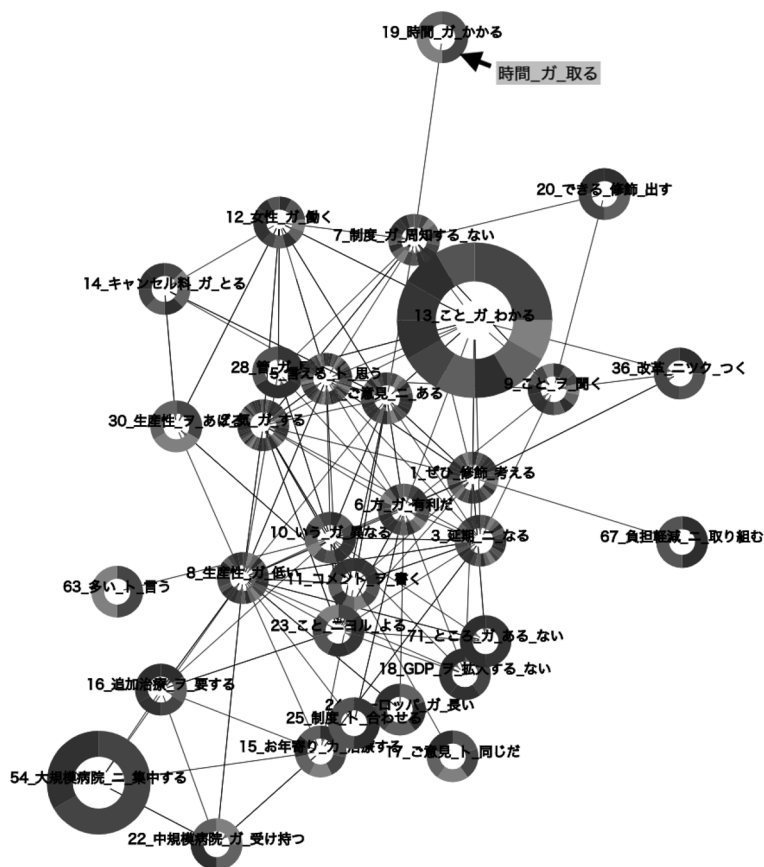


図5 言葉ネットワーク

次にその格フレームネットワークそれぞれに対して、上述の Newman クラスタリングを適用した。たとえば、トピック「働き方改革について」では、167 個のクラスタ³と 1 クラスタ当たり平均 4.5 個の要素ができており、比較的多様な話題が投稿されていた。またクラスタの要素数は多くないため、意味解釈のしやすいクラスタが抽出できた。

最後に、コメントに出現するクラスタの共起関係を Jaccard 係数で計算し、その値が設定した最小 Jaccard 係数以上の場合に枝を張ることで、言葉ネットワークを生成する。ここで、あるコメントに出現するクラスタの集合をトランザクションと呼ぶ。そして、任意の二つのコメントのトランザクションを X, Y とすると、Jaccard 係数は、以下のように定義される。

$$\text{Jaccard}(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} \quad (1)$$

この値をもとに、最小 Jaccard 係数と最小支持度を閾値として与え、2 アイテムのクラスタペアを全列挙

し、それらのクラスタ間に枝を張ることで言葉ネットワークを作成した。最小 Jaccard 係数は 0.2 とし、最小支持度を 2 件とした。

図5はトピック「働き方改革について」の言葉ネットワークから密な部分だけを表示している。節点はクラスタを表し円グラフで表現している。また円グラフの要素は格フレームを表しており、格フレームの出現割合によって要素の大きさが異なっている。節点ラベルにはクラスタ内で出現割合が一番大きい格フレームが表示されている。また節点の大きさはクラスタの出現頻度を表しており、多くのコメントに出現するクラスタは大きく表示されている。このネットワークは NYSOL の mnetpie を利用してブラウザ上に描画されており、図5の一番上のクラスタに「時間_ガ_取る」と表示されているが、カーソルをクラスタの要素に合わせるとその要素名も表示できる点が特徴である。

これらのクラスタのつながりを見ていくことで議論されている内容から概要を把握できる。たとえば、左下の大きなクラスタはラベルが「54_大規模病院_ニ_集中する」で、このクラスタにはもう一つの要素「54_患

³ 言葉が節点なのでコミュニティではなくクラスタと呼ぶ。

者_ガ_集中する」がある。

次にこのクラスタとつながりのある「22_中規模病院_ガ_受け持つ」は他に「治療_ヲ_受け持つ」、「治療_ヲ_追加する」、「場合_追加する」からなるクラスタである。そして「15_お年寄り_ガ_治療する」、「16_追加治療_ヲ_要する」と「8_生産性_ガ_低い」というつながりを解釈すると、「大規模病院に患者が集中し、中規模病院がお年寄り患者を受け持ち治療をする。お年寄りの治療や追加治療は生産性が低い」というような意味として捉えることができる。もちろん、クラスタのつながりだけで正確な文章を理解することはできないため、実際の使い方としては、これらのネットワークをトピックごとに用意しておき、このネットワークから興味のあるトピックを見つけてもらい、詳細はそのトピックに投稿されているコメントを読んでもらうことになる。

5. おわりに

本稿は、われわれが開発・運営しているハッピーキャリア SNS を利用し、SNS に登録しているユーザーが作成したトピックとそのトピックへの書き込みコメントの関係からユーザネットワークを構築し、SNS を活性化させるためのキーパーソンを特定した。キーパーソンの特定は、ネットワークの中心性を表す指標を用いて、ハブとなるユーザーの発見と、弱い紐帯の強さの観点からコミュニティ間をつなぐブリッジとなるユーザの発見を行った。

当該ネットワークは節点数と枝の数は少なく小規模

なネットワークであるが、Newman アルゴリズムを適用することでコミュニティ構造をうまく捉えることができていた。また言葉ネットワークでは、文章から内容を表す格フレームをうまく抽出できており、それらの関係性から文章の意味をある程度把握できる点を示した。

今後は、本稿から得られた知見を実際の運用に活用することで当該 SNS の活性化につなげていくように継続して取り組んでいく。

謝辞 本研究は、JST CREST（グラント番号：JP-MJCR1401）および、JSPS 科研費：15H03389 の研究助成を受けている。

参考文献

- [1] T. Nakahara, A. Ouchi and Y. Hamuro, “Using social networking services in education to support learning,” In *Proceedings of the International Symposium on Business and Social Sciences*, pp. 226–234, 2019.
- [2] M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Physical Review E*, **69**, article number: 026113, 2004.
- [3] M. S. Granovetter, “The strength of weak ties,” *The American Journal of Sociology*, **78**, pp. 1360–1380, 1973.
- [4] Z. S. Harris, “Distributional Structure,” *WORD*, **10**, pp. 146–162, 2015, DOI: 10.1080/00437956.1954.11659520
- [5] 黒橋禎夫, 河原大輔, <http://nlp.ist.i.kyoto-u.ac.jp/?KNP> (2019 年 8 月 5 日閲覧)