

卒業論文 2017 年度（平成 29 年度）

ソーシャルメディア上のコミュニティ形成に
対して個人の性格が与える影響の研究

指導教員 坂田 一郎 教授

東京大学 工学部 システム創成学科 C

03-160977 鈴木 凱亜

2018 年 2 月 1 日 提出

ソーシャルメディア上のコミュニティ形成に 対して個人の性格が与える影響の研究

要旨

ソーシャルメディア上の人々の関わり合いは、今後もますます発展し多様化するであろう。そんな時流で、ソーシャルメディア上での人との関わりの生み出す生産性や効用をよりよいものにするためには、人々がどのようなモチベーションでソーシャルメディア上でコミュニティを形成するか、そのメカニズムの把握が重要である。本論文では、SNS 上の人間同士のつながりに影響を及ぼす要因を「環境要因」、「その時点での個人の感情や思想」、「個人の性格」の 3 点に絞った。そのうち、「環境要因」、「個人の感情や思想」の 2 要素がコミュニティ形成に与える影響は、Homophily（民族、人種などで、似た者同士が集まろうとする傾向）や Echo Chamber（思想の偏り）の研究から、小さくないことが分かっている。

本研究では、コミュニティの形成やそのふるまいにおいて、構成員の性格がどのような影響を与えるかを分析するために、「環境要因」と「個人の感情や思想」を統制した共通のトピックのネットワークを作成した。このネットワークを用い、性格推定モデルによって定量的に推定した SNS ユーザの性格指標が、Homophily な傾向を示すか、そして SNS 上のコミュニティの規模や活性化に、構成員の性格の平均値やばらつきが影響を与えるのか相関分析を行った。その結果、たとえば Conscientiousness（誠実性）という性格は、似た人の方が仲良くなりやすい、という Homophily な作用があることが分かった。また、コミュニティの活性化のための性格的な相性として、Emotionality（情動性）などの人との接し方に関わる性格傾向はなるべく似た者同士で、Self-efficacy（自己効力感）などの自分のあり方に関わる性格傾向はなるべく多様性があるように集まることが重要であることを示した。

以上より、性格の与えるソーシャルメディアのネットワークダイナミクスへの影響を調べる上で、本研究の分析のフレームワークの有効性を示した。

目次

| | | |
|------------|-------------------------------------|-----------|
| 第1章 | 序論 | 1 |
| 1.1 | 研究背景 | 1 |
| 1.1.1 | ソーシャルメディアの普及と活用 | 1 |
| 1.1.2 | ソーシャルメディアのデータを用いた個人の特徴抽出 | 2 |
| 1.1.3 | ソーシャルネットワーク上のコミュニティ形成 | 3 |
| 1.1.4 | 問題提起 | 4 |
| 1.2 | 本論文の目的 | 5 |
| 1.3 | 本論文の貢献 | 6 |
| 1.4 | 本論文の構成 | 6 |
| 第2章 | 関連研究 | 7 |
| 2.1 | 心理学における性格の評価基準 | 7 |
| 2.1.1 | ビッグファイブ性格指標 | 7 |
| 2.1.2 | ビッグファイブを利用した研究；性格とチームのパフォーマンスの関係など | 7 |
| 2.2 | ソーシャルメディアの分析 | 8 |
| 2.2.1 | ビッグファイブと SNS 上のユーザの行動の関係 | 8 |
| 2.2.2 | SNS ユーザの性格の推定 | 8 |
| 2.2.3 | コミュニティ形成のメカニズムに関する研究 | 9 |
| 2.3 | 本論文の位置づけ | 9 |
| 第3章 | 背景となる技術 | 11 |
| 3.1 | IBM Watson Personality Insights API | 11 |
| 3.1.1 | 概要 | 11 |
| 3.1.2 | 背景技術 | 12 |
| 3.1.3 | Personality Insights 使用例 | 12 |
| 3.2 | ネットワーク上でリンクを結ぶノードの属性の相関を測る指標 | 14 |
| 3.3 | ネットワークの分割手法 | 15 |
| 3.3.1 | モジュラリティについて | 15 |
| 第4章 | 提案手法 | 17 |
| 4.1 | 手法の手順 | 17 |
| 4.2 | 共通のトピックネットワークに所属するユーザの抽出 | 17 |

| | | |
|--------------|---|-----------|
| 4.3 | 抽出したユーザの性格推定 | 18 |
| 4.4 | ネットワークの構築 | 19 |
| 4.5 | ユーザの分析指標の作成とネットワーク指標の抽出 | 19 |
| 4.5.1 | 作成するユーザ行動の分析指標 | 19 |
| 4.5.2 | 抽出するユーザのネットワーク指標 | 19 |
| 4.6 | クラスタリングとコミュニティの分析指標の作成 | 20 |
| 4.7 | ユーザの性格と行動・ネットワーク指標の相関分析 | 21 |
| 4.8 | ネットワークの Matching Assortativity による分析 | 21 |
| 4.9 | コミュニティの性格とふるまいの相関分析 | 21 |
| 第 5 章 | 実験 | 23 |
| 5.1 | データの収集 | 23 |
| 5.1.1 | データ収集に用いるソーシャルメディア | 23 |
| 5.1.2 | 特定のトピックネットワークを抽出 | 23 |
| 5.1.3 | コミュニティの分析指標作成のためのデータ収集 | 25 |
| 5.2 | ユーザの性格の計算 | 26 |
| 5.3 | ネットワークの構築とコミュニティ分割 | 26 |
| 5.4 | 相関分析用の分析指標の抽出 | 28 |
| 5.5 | リサーチクエスション 1 の結果の予想 | 28 |
| 5.6 | Homophily な作用が起こる性格についての仮説 | 28 |
| 第 6 章 | 結果 | 31 |
| 6.1 | リサーチクエスション 1 : 外部モデルを使用する分析の妥当性の確認 | 31 |
| 6.1.1 | ツイートの頻度 | 31 |
| 6.1.2 | リプライ率 | 32 |
| 6.1.3 | リプライ傾向 | 33 |
| 6.2 | リサーチクエスション 2 : 集団の人間関係における個人の位置づけに性格が 与える影響の分析 | 34 |
| 6.2.1 | 次数 | 34 |
| 6.2.2 | 近接中心性 | 35 |
| 6.2.3 | 媒介中心性 | 36 |
| 6.3 | リサーチクエスション 3 : Homophily な傾向を示す性格はあるか | 37 |
| 6.4 | リサーチクエスション 4 : コミュニティ内の性格の分布がコミュニティのふ るまいに与える影響の分析 | 39 |

| | | |
|--------------|--|-----------|
| 6.4.1 | コミュニティの人数に対する寄与 | 39 |
| 6.4.2 | コミュニティの活性化に対する寄与 | 40 |
| 第 7 章 | 考察 | 41 |
| 7.1 | 外部モデルを使用した提案手法の妥当性の確認 | 41 |
| 7.2 | 集団の人間関係における個人の位置づけに性格が与える影響の分析 | 42 |
| 7.3 | Homophily な傾向を示す性格はあるか | 43 |
| 7.4 | コミュニティ内の性格の分布がコミュニティのふるまいに与える影響の分析 | 44 |
| 第 8 章 | 結論 | 45 |
| 8.1 | 本研究の結論 | 45 |
| 8.2 | 課題と今後の展望 | 46 |
| 8.2.1 | より多様なトピックネットワークによる分析 | 46 |
| 8.2.2 | スパースでないトピックネットワークより多く抽出する方法の考案 | 47 |
| 8.2.3 | 異なる性格の組み合わせの評価 | 47 |
| 8.2.4 | 性格分析モデルの精度の向上 | 47 |
| | 謝辞 | 48 |
| | 参考文献 | 50 |

図 目 次

| | | |
|-----|---|----|
| 3.1 | Personality Insights による個性の計算 | 13 |
| 3.2 | 夫婦の結婚時の年齢 | 15 |
| 4.1 | 性格とコミュニティ形成の関係分析の手順 | 18 |
| 5.1 | Louvain 法によって分割されたコミュニティごとの人数 (5 人以上) | 27 |
| 6.1 | 映画ネットワークのリンクの両端の性格値 (a)Gregariousness (b)Self-consciousness | 37 |

表 目 次

| | | |
|-----|--|----|
| 3.1 | Personality Insights API が計算する全 35 種類の性格 | 12 |
| 3.2 | Personality Insights API への入力に含まれる投稿データ | 13 |
| 3.3 | Personality Insights API による計算結果の例 | 14 |
| 5.1 | トピックネットワーク抽出のためのハッシュタグ検索 | 24 |
| 5.2 | タイムライン検索 | 24 |
| 5.3 | TwitterAPI によりツイートを検索した日時 | 25 |
| 5.4 | Twitter タイムラインと Watson API への入力の対応関係 | 25 |
| 5.5 | 構築したネットワーク | 26 |
| 6.1 | トピックネットワークごとの性格の推定値とツイートの頻度の関係 | 31 |
| 6.2 | トピックネットワークごとの性格の推定値とリプライ率の関係 | 32 |
| 6.3 | トピックネットワークごとの性格の推定値とリプライ傾向の関係 | 33 |
| 6.4 | トピックネットワークごとの性格指標と次数の関係 | 34 |
| 6.5 | トピックネットワークごとのビッグファイブと近接中心性の関係 | 35 |
| 6.6 | トピックネットワークごとの性格指標と媒介中心性の関係 | 36 |
| 6.7 | トピックネットワークごとの各性格についての matching assortativity | 38 |
| 6.8 | トピックネットワークごとのコミュニティの性格の平均や分散とコミュニティ の人数の関係 | 39 |
| 6.9 | トピックネットワークごとのコミュニティの性格の平均や分散とコミュニティ の活性化の関係 | 40 |
| 7.1 | Extraversion（外向性）の小分類の性質の考察 | 44 |

第1章 序論

人にはそれぞれ独自の性格がある。陽気な人と無口な人。神経質な人と鈍感な人。これらの性格は簡単には変化することはない。人がそれぞれもつ個性となっている。他人と人間関係を築き上げる際、性格的な相性は少なからずその関係性の発展に影響を与える。ソーシャルメディア上でも、密につながっているコミュニティにネットワークを分割すると、その分かれ方や各々のコミュニティの特徴は個人の性格の影響を受けるのではないだろうか。

Twitter や Facebook を始めとするソーシャルメディアは瞬く間に世界中に普及し、そのネットワーク上の人と人の関わり方も多様化している。ソーシャルメディアは人々の取得する全ての情報の中でも大きな割合を占めるようになり、そのネットワーク上の人と人のつながり方、コミュニティの分かれ方は流行の伝播や世論の形成などに大きな影響を与えるようになった。

また、ソーシャルメディアは人間のパーソナルな情報を集積した膨大なデータベースである。それを活用して、社会現象の予測や、マーケティングへの応用、個人の行動予測や特徴抽出などを旨とする研究は広くなされている。そして、ソーシャルメディアから取得可能なデータから、ユーザの性格を推定する手法も開発された。

本研究では、ソーシャルメディアの情報から個人の性格が計算できることに着目した。そして、短期的な現象に左右されない、参加者の性格が、ソーシャルネットワークのコミュニティ形成にどのような影響を与えるかを分析するフレームワークを提案する。

本章では、以上の背景について詳しい説明を行ったのち、現状における問題意識と本研究の目的を述べる。そして、行った実験と分析結果の概略をもとに、本研究の貢献を説明する。最後に本論文の全体構成について書く。

1.1 研究背景

1.1.1 ソーシャルメディアの普及と活用

そもそもソーシャルメディアとは、SNS、マイクロブログ、動画共有サイト、CtoC サービスのアプリケーションといった、利用者が情報を受信すると同時に発信もできるような情報の共有媒体のことをさす。ソーシャルメディアの登場により、それまで大衆に向けて一方方向的だった社会における情報の流れは、全てのユーザが発信する側にも受信する側にもなりうる双方向的なものとなった。一般大衆でも日常の些細な出来事や感情を手軽に世界中に発信することが可能となったことで、人々の「身の回りの情報を発信したい」という新たな需

要が刺激され、ソーシャルメディアは数年間で爆発的に普及した。

(1) ソーシャルメディアのもたらす社会の変化

ソーシャルメディアの普及は、社会を流れる情報の性質を変化させた。一つ目に、情報の流れが双方向的になり誰もが発信する側になれるようになったことで、それまではマスメディアによって選択され、加工されていた情報が、より大量で、無責任で、パーソナルなものになった。たとえば、災害時のデマや風評被害の問題、差別問題や労働問題に関して見られるインターネット上の憎悪などの感情のうねりは、このような情報の性質の変化がもたらした現象の一例であろう。二つ目に、情報発信において編集や校閲、権利の確認といった手間が少ない一般大衆による投稿は、情報の拡散の速さ、リアルタイム性の高さをもたらし、たとえばテレビ番組やスポーツ観戦において、現在進行形で大勢と感想を共有することができるようになった。つまり、ソーシャルメディアの普及により、人々は大量のパーソナルな情報にリアルタイムでアクセスできるようになった。

(2) ソーシャルメディアのデータの活用

ソーシャルメディアによってリアルタイムに取得できる膨大な量の情報は、社会のさまざまな動向や実態を予測、評価するためのビッグデータとして注目されている。ソーシャルメディアのリアルタイム性や、パーソナル性は、社会の変化をより敏感に捉えることを可能にする。また、以前よりも低コストで多種多様な人々の意見にアクセスできるようになったため、分析におけるデータの偏りによるバイアスの発生の心配が軽減される。

したがって、さまざまな立場にとってソーシャルメディアのデータ分析を行うことは有益である。たとえば、企業は社会の潜在的ニーズの抽出やトレンド予測により新規事業参入などにむけたヒントを得られる。政治家は選挙の際、取り込めそうな浮動票が多くある地域を分析することで、最適な選挙戦略を決定することができる。実際に、SNS 分析によって企業活動や選挙活動をサポートするコンサルティング会社も多数存在しており、こういった分析の需要の高さがうかがえる。

1.1.2 ソーシャルメディアのデータを用いた個人の特徴抽出

ソーシャルメディアのデータは社会のマクロな現象やトレンドの分析のみならず、ユーザ個人の特徴を抽出することにも応用される。具体的には、ソーシャルメディアのデータから特定のユーザの興味や性別、職業といった属性の推定、ユーザの移動や購買といった行動を予測する研究がなされている。これらの研究は、人それぞれにパーソナライズされたサービスやユーザインタフェース、レコメンドの提供を可能としている。同様に、ソーシャルメディア上のユーザの投稿から、そのユーザの性格を推定する研究が進んでいる。

性格の評価方法として、心理学の分野ではビッグファイブと呼ばれる 5 軸の評価指標によって性格傾向を表すことがスタンダードとなっている。各指標の値は心理テストを通して

算出できるものであり、ビッグファイブは幅広い性格傾向を定量的に捉えるものとして広く受け入れられ、様々な人間行動との関係が研究されている。詳しくは第2章で述べる。

当初は心理テストなどのアンケートを実施しないと算出できなかったビッグファイブは、ソーシャルメディアのユーザとの紐付けが難しく、性格とソーシャルメディア上の行動の関係を調べるには向いていなかった。そこで、ソーシャルメディアから取得可能なデータから、直接ビッグファイブを推定する様々な研究がなされた。その中で IBM の Watson は、テキストデータやマイクロブログの投稿データを入力として、著者のビッグファイブを計算するモデルを API として 2015 年に公開しており、現在は英語だけではなく、日本語のテキストや SNS アカун トにも対応している。日本語の Twitter データを用いて計算した値は心理テストを通して得た実スコアと高い相関を示した [1]。この API はビッグファイブの性格因子に、それぞれのグループをより細かく分類した指標を加え、計 35 種類の性格指標を算出できる。本研究ではなるべく高い精度で性格を推定する手段として、この API を使用した。Watson Personality Insights API については詳しく第3章で説明する。

1.1.3 ソーシャルネットワーク上のコミュニティ形成

ソーシャルメディア上では、利用者が互いに情報の送受信を行い、巨大なつながりのネットワークを形成する。ネットワークにおけるユーザ同士のつながりは、決して一様に発生するのではなく、ユーザの組み合わせによってつながりの生じやすさは大きく異なる。このようなネットワークにおいて、グループ内でのつながりがグループ外とのつながりと比べて密であるような集団をコミュニティと呼び、それぞれ独自の興味や文化が観測される。コミュニティはさまざまな要因で形成され、さまざまな現象をもたらす。

(1) ネットワークのコミュニティ形成のメカニズムの例

社会ネットワークにおけるコミュニティ形成とそのもたらす現象の例として、Echo Chamber が挙げられる。この現象は、人が自らと似た価値観を持つ人に対し好感を抱きやすいことに起因する。人はついつい自らと似た感情や意見を持つ人を選択してつながりを持つようとする。それが繰り返されていくうちに、人々はそれぞれ似た思想の人間同士が次々とつながるようになる。結果、偏った主流派の意見のみがこたまし、部外者に排他的な思想を持つコミュニティにネットワークが分断されてしまうという現象である。

この考え自体はソーシャルメディアの普及以前から存在したが、注目されるようになったのはソーシャルメディアの普及によるところが大きい。ソーシャルメディアの登場により、自分にとって似ている人間や、都合のいい情報を見つけるのが簡単になった。それによりつながる対象の選別が容易になったため、実際に Echo Chamber が観測されるようになったのである [2]。特に、2016 年アメリカ大統領選挙におけるオルタナ右翼の Echo chamber の形成は、独自の情報の共有と拡散を起こし、選挙結果にも大きな影響を与えたと言われている。

(2) ソーシャルネットワーク上のコミュニティ形成を研究する意義

このように、ソーシャルメディア上の人と人の繋がり方に起因して、時に大きな社会現象が引き起こされる。ソーシャルメディアの情報が現代人の取得する情報の大きな割合を占めているため、ソーシャルメディアの情報は人々の意思決定や意見形成に大きな影響を与えるのである。

したがって、ソーシャルメディア上の人々がどのようにお互いと繋がっているか、そのネットワークの形状を把握することには、様々なメリットがある。どのような人々がコミュニティを形成するか分かれば、先の例のような大きなスケールのネットワークでは、社会でどのように情報が拡散され人の意思決定に影響を与えるかを分析することができる。小さなスケールのネットワークにおいても、どのような人間同士が相性がよいか、最適なチームビルディングやパートナー推薦システムの構築に対する知見を得ることができる。

(3) ソーシャルネットワーク上のユーザ間のつながり方に影響を与える要因

本研究では、ソーシャルメディアのネットワークで接触したユーザ同士がつながるか否かを決定する要因は、以下の3要素で評価できると仮定する。

- ユーザ同士の接触した場（その場の環境的要因）
ユーザ同士がどういった状況で接触したのか、そのトピックや文脈など。たとえば、同じSNS上でも議論を交わす目的の場と、人間同士が知り合いになり近づくため設けられた場では、その後のユーザ同士の関係の発展の仕方は異なるであろう。
- ユーザ同士の接触時の感情や思想（個人のその場での性質）
それぞれのユーザの互いに対する感情や、その場のトピックに対する思想など。たとえば、Twitterで同じ政権について話し合う場においても、政権に対して同一の感情を抱いている人同士であるか、異なる感情を抱いている人同士であるかによって、その後の関係は異なるであろう。
- 接触したユーザ同士の生まれながらの性格（個人の不変な性質）
接触したユーザ同士がそれぞれどのような性格の人か。場面やトピックによらず、性格的な相性がいい場合と悪い場合では、その後の関係性は異なるであろう。

たとえば、大統領選挙時にSNS上でEcho Chamberが見られた背景には、政治的な思想の差によるコミュニティ形成、すなわち上記の3要素のうち「ユーザ同士の接触時の感情」の影響を受けたコミュニティ形成があると考えられる。

1.1.4 問題提起

ユーザの性格と、そのユーザのSNS上の行動の関係に関する研究は世界で様々ある。しかし、第一に、複数のユーザが集まってコミュニティを形成するときに、性格がどのような影響を与えるか議論した研究は少ない。これは、上述したように、コミュニティ形成のメカ

ニズムに関わる要因として、環境要因や個人の思想や感情といった要素が寄与する部分が大いからと思われる。また、第二に、日本語圏の SNS ユーザに限っては、そもそもユーザの性格と行動の関係に関する研究もあまり見られない。これは、ビッグファイブの枠組みや測定方法が英語圏で生み出されたものであり、これまで日本語への対応が十分には進められていなかったためと思われる。

ネットワークの環境条件や、その場での個人の感情や思想と関係なく、純粋に性格的な因子だけが与えるコミュニティ形成に対する影響はどのようなものだろうか。場面に応じて変わるものではなく、ユーザごとに定義される性格という不変量は、コミュニティの形成に影響を与えるのだろうか。これを解明することで、たとえば出会ったことのない人々をマッチングするときなど、未知の状況において人々がどのように関係性を築き上げるか、そのポテンシャルを捉えることができると思う。

1.2 本論文の目的

本研究では個人の性格がソーシャルネットワーク上のコミュニティ形成にどのような影響を与えるかを分析し、その分析のフレームワークを提案する。実験に必要な性格の推定には、今回はコストの制約上実現可能な選択肢の中で最も精度の信頼性が高いと思われる、IBM Watson の Personality Insights API を用いる。

以上の目的を達成するために定めた 4 点のリサーチクエストと、提案する分析のフレームワークにおけるそれぞれへのアプローチの概略を以下に示す。

(1) 外部モデルを使用する分析の妥当性の確認

API を使用して計算した日本語の SNS ユーザの性格と、ユーザの SNS 上の行動の相関関係が、第 2 章で紹介する関連研究と矛盾せずに観測されるかを分析する。

(2) 集団の人間関係における個人の位置づけに性格が与える影響の分析

SNS ユーザの性格の指標と、会話ネットワークにおけるユーザのネットワーク指標に関係が見られるか、相関分析を行う。

(3) Homophily な作用が起こる性格の存在の確認

Echo Chamber が似た思想の人同士が集まる結果起こる現象であるように、コミュニティ形成過程において、性格も似た者同士があつまる傾向がないか、Matching Assortativity と呼ばれるネットワーク指標の計算により分析する。

(4) 個人の性格が所属するコミュニティの性質に与える影響を分析

SNS の会話ネットワークから抽出したコミュニティのふるまいと、コミュニティに所属する個人の性格の統計量に関係が見られるか、相関分析を行う。

以上の点について、実験を行い考察することを本研究の目的とする。

1.3 本論文の貢献

ここでは本論文の貢献について述べる。

- ソーシャルメディアのネットワークダイナミクスと、ユーザの性格の関係を分析するフレームワークの提案を行った
- Twitter のメンションネットワーク上で形成されたコミュニティが、所属するユーザの性格によってどのような影響を受けるか分析した
- 性格計算ツールを用いた分析のもつ可能性を考察した。

また、性格的な相性が非常に重要となるのが、面識のない人同士のマッチングである。したがって実用的な貢献として、たとえば、リモートでのチームワークのメンバー構成をマネージャーが考える際、もしくはマッチングサービスなどでユーザにとって相性の良い相手をレコメンドする際、性格的な要素を考慮することでより大きな効用を生み出せる可能性を提示した。

人同士のやりとりの情報化が著しく進んでいる現代において、本研究は個人の性格的な相性をシステムが考慮し、「空気の読める」誘導ができるようなシステム設計に向けた一つのアプローチであると考える。

1.4 本論文の構成

第2章では序論で述べたことに関連する研究と本研究の位置付けについて述べる。第3章では本研究で使用した性格分析 API と、ネットワークのコミュニティ分割法について述べる。第4章では性格とコミュニティ形成の関係を分析する手法について述べる。第5章では、第4章で述べた手法を用いた本研究の実験を説明し、リサーチクエスションに対する仮説を述べる。第6章では実験の結果について述べる。第7章では実験結果に関する考察を行い、分析、提案手法の有用性・課題について議論する。第8章では本研究の結論を述べる。

第2章 関連研究

本章では、人間の性格の捉え方、そしてソーシャルメディアの分析とネットワークのコミュニティ形成についての関連研究を概説し、本研究の位置づけについて述べる。

2.1 心理学における性格の評価基準

2.1.1 ビッグファイブ性格指標

心理学において、全ての人が固有にもつ性格を分類する基準の作成は長い間課題とされていた。さまざまな性格の評価基準とその測定方法が考案され、データが蓄積されていった。すると、次第にどの性格の評価基準もビッグファイブと呼ばれるおおまかなカテゴリに分けることができると判明した。Goldberg らは、どのような性格の評価基準も「神経症傾向 (Neuroticism)」、「外向性 (Extraversion)」、「経験への開放性 (Openness)」、「協調性 (Agreeableness)」、「誠実性 (Conscientiousness)」の五つの性格 (の評価基準) のいずれかとある程度の相関があることを示した [3]。Costa らの考案した NEO PI-R と呼ばれる 240 問の心理テストにより定量的に性格を指標化できることもあり [4]、この分類は心理学におけるスタンダードとなった [5]。McCrae らは、性格が言語により規定される以上、言語によって特有の性格が存在する可能性を認めつつ、ビッグファイブの傾向は日本語を含めユニバーサルに見られることを示した [6]。そして Schmitt らはこのビッグファイブの枠組みを用いて各言語圏における性格の分布の比較や考察を行った [7]。また、Costa らは通常の性格だけではなく、さまざまなパーソナリティ障害もビッグファイブの指標により表せる範囲にあることを示し [8]、性格診断におけるビッグファイブの汎用性の高さが定着していった。

2.1.2 ビッグファイブを利用した研究；性格とチームのパフォーマンスの関係など

人のビッグファイブの指標とその人の属性や行動の関係を分析することは、人の資質や適正の判断、精神病理の発展など、さまざまなことに応用される。たとえば Barrick らはビッグファイブと仕事能力の関係を分析し、誠実性 (Conscientiousness) の高さは多くの職業で能力の高さと正の相関があることを示した [9]。また、性格とチームの生産性の関係を分析するために、チームの構成員のビッグファイブの平均と分散、最小値、最大値について、それぞれチームのパフォーマンスとの相関を調べた。その結果、個人の場合同様、誠実性の平均が高いチームは低いチームと比べ高いパフォーマンスを発揮するが、同時に誠実性のばらつきが大きいとパフォーマンスは低下してしまうことを発見した [10]。同様に、Driskell ら

はビッグファイブを細分化した小分類について、チームの効率性に影響する因子を分析した [11]. このような研究は 2018 年現在も行われており、たとえば Steca らはどのような性格因子がスポーツ選手に多くアスリートとしての成功に関わるかを相関分析により研究している. [12]

2.2 ソーシャルメディアの分析

2.2.1 ビッグファイブと SNS 上のユーザの行動の関係

ビッグファイブの性格因子はソーシャルメディア上の人の行動にも影響を与える. たとえば Ryan らは Facebook のユーザには経験への開放性 (Openness) と孤独さ (Loneliness) が高い人が多いことを示し [13], Skues らは中でもその傾向が強い人ほど、友達を多く持つ傾向があることを示した [14]. また, Seidman は協調性 (Agreeableness) と神経症傾向 (Neuroticism) が高いほど、他人との接触を求める傾向が強く、神経症傾向が高く、誠実性が低いほど、投稿やプロフィールの更新などの自己表現傾向が強くなることを示した [15].

同様の分析が、Twitter や中国の QQzone など、様々なソーシャルメディアにおいてなされている. Correa らは性格と行動の関係は性別や年齢によっても変化することを示した [16]. ただし、これらのような性格と SNS 上のユーザの行動の相関を分析した研究は、日本ではあまり盛んには行われていない.

以上の分析は、全てビッグファイブのデータを被験者の協力のもと心理テストを実施することによって集めている. 次は、ソーシャルメディア上で取得可能なデータによって性格の値を算出する研究を紹介する.

2.2.2 SNS ユーザの性格の推定

SNS 上のユーザの性格と行動の関係を調べる上で、心理テストを通してでないと性格が抽出できないというのは、被験者の協力の元でデータを収集する必要があることを意味し、サンプルの多様性を損なってしまう. したがって、SNS から取得可能な情報のみで性格を推定する研究が広く行われている. Kosinski らは Facebook でユーザと、そのユーザが「いいね」を押した投稿の対応行列を作成し、次元削減とロジスティック回帰を用いて、人種や性別などの属性、さらにはビッグファイブ性格因子を予測するモデルを提案した [17]. このモデルは「いいね」の情報のみを使用しているにもかかわらず、特に Openness (経験への開放性) の予測においては心理テストと変わらない精度を示すことが分かった.

投稿テキストを用いた性格推定

性格を推定するために有効な入力データとして注目されたのがユーザが投稿したテキスト情報である. Fast らは人間の言葉遣いは性格を反映するという仮定のもと、言葉と性格の対

応付けを記録した辞書を用いテキスト分析を行った。その結果ビッグファイブをはじめとした多くの性格指標がその人の書く言葉から予測できることが分かった [18]。Hirsh らは被験者に書かせた自らについて書いた作文から、ビッグファイブと非常に相関の強い指標を取り出すことに成功した [19]。Golbeck らは Twitter のデータを処理し、LIWC と呼ばれる分類体系に沿って抽出した性格因子語といくつかの Twitter 上の行動を特徴量としてビッグファイブの値を予測した [20]。Arnoux らは抽出する単語の分散表現化とガウス過程回帰の学習モデルを用いたことで、性格の予測に必要な入力テキストデータ数を大幅に減らした [21]。最新の研究では、Azucar らがユーザの基本属性、Facebook の「いいね」対象、投稿数や友人人数などの活動の統計情報、投稿のテキスト情報、プロフィールや投稿の画像情報など、数多くの種類の情報をすべて特徴量とした性格推定を行った。結果、あらゆる情報を特徴量に含めることで、非常に高精度な性格推定が可能なることを示した [22]。性格推定モデルの日本語化については那須川らが行っており、日本語の SNS データに関しても実スコアと高い相関をもつビッグファイブの推定が可能になったが、そのスコアと SNS 上の行動の関係を分析した研究は少ない。詳しくは第 3 章で説明する。

2.2.3 コミュニティ形成のメカニズムに関する研究

ソーシャルメディアにおけるコミュニティの形成にはさまざまな要因があるが、最もオーソドックスな要因は、共通の趣味や興味のもと、ユーザが自発的にクラスタを形成することであろう。4chan や Reddit といったプラットフォームは、自分が興味を持っている情報のジャンルを最初に自発的に選択しその中で人々と接触し議論を交わす、第 1 章で述べたように、あるトピックに対する思想がコミュニティ形成に影響を与える例のひとつに Echo Chamber があげられる [2]。このようなコミュニティ形成の背後には Homophily と呼ばれる、人間は共通点をもつ人間とつながりやすい、という原理がある。共通点によるコミュニティ形成として、McPherson らは人種と民族の共通点が最も強くネットワークのクラスタ化を引き起こし、年、宗教、職業、性別などの共通点がそれに続くとした [23]。このように、人間の環境要因や感情、興味によるコミュニティ形成のメカニズムは研究されている。

2.3 本論文の位置づけ

既存研究と照らし合わせた本研究の位置づけ・有用性は以下の通りである。

- テキストから推定された性格と、そのユーザの行動の関係を、改めて日本語のデータと性格推定モデルを用いて分析した
- SNS 上のコミュニティ形成のメカニズムを、個人の性格という観点から分析した

第3章 背景となる技術

3.1 IBM Watson Personality Insights API

本節では、SNS ユーザの性格を推定するために用いるモデルである、IBM Watson の Personality Insights API について、概要と背景技術について述べたのち、その使用例を示す。

3.1.1 概要

Watson とは、IBM が開発した自然言語処理や画像処理、音声認識などの機能を搭載したシステムであり、Bluemix という IBM のクラウドサービスから様々な自然言語処理や機械学習のサービスを提供している。この一連のサービス群の一つが Personality Insights であり、これはテキスト情報を入力として、テキストを記述した人の性格を推定するモデルである¹。Fig. 3.1 にブラウザ上で API を使用した画面を表示する。

Personality Insights は出力としてビッグファイブの性格因子の値、そしてそれぞれの性格因子の小分類として定義された 6 種類の性格（ファセット）を計算する。具体的には、Table 3.1 に API が計算する 35 種類の性格を示す。

また、本研究では用いてはいないが、ビッグファイブとは別に、ビジネスの領域に有益な「ニーズ」を 12 種類の評価軸で、何に対して価値を求めるかを表す「価値観」を 5 種類の評価軸で算出するため、対象の消費選好の推定など、ビジネス面への応用も期待されている。

Personality Insights API は Twitter などのマイクロブログのデータを分析することも想定している。単純なテキストの自然言語処理はもちろん、マイクロブログの投稿の順序や日時、テキストの返信、引用関係等を入力情報に加えることもできる。その場合、Personality Insights API はこれらの情報を用いテキスト同士の関連性を考慮した性格の推定を行う。

Personality Insights に限らず、音声認識や画像分類など、Watson はさまざまな機械学習の手法を公開している。今後、機械学習を用いたマーケティング分析や戦略立案を行いたいが、計算リソースや学習データが不足している企業などが、このサービスを活用して低コストでさまざまな知見を得られるようになることが期待されている。

¹<https://www.ibm.com/watson/jp-ja/developercloud/personality-insights.html>

Table 3.1: Personality Insights API が計算する全 35 種類の性格

| ビッグファイブ | 6つの小分類（構成要素の性格） |
|------------------------|--|
| Openness（知的好奇心） | Adventurousness（大胆性）, Artistic interests（芸術的関心度）, Emotionality（情動性）, Imagination（想像力）, Intellect（思考力）, Authority-challenging（現状打破） |
| Conscientiousness（誠実性） | Achievement striving（達成努力）, Cautiousness（注意深さ）, Dutifulness（忠実さ）, Orderliness（秩序性）, Self-discipline（自制力）, Self-efficacy（自己効力感） |
| Extraversion（外向性） | Activity-level（活発度）, Assertiveness（自己主張）, Cheerfulness（明朗性）, Excitement-seeking（刺激希求性）, Outgoing（友好性）, Gregariousness（社交性） |
| Agreeableness（協調性） | Altruism（利他主義）, Cooperation（協同性）, Modesty（謙虚さ）, Uncompromising（強硬さ）, Sympathy（共感度）, Trust（信用度） |
| Emotional range（感情起伏） | Fiery（激情的）, Prone to worry（心配性）, Melancholy（悲観的）, Immoderation（利己的）, Self-consciousness（自意識過剰）, Susceptible to stress（低ストレス耐性） |

3.1.2 背景技術

Personality Insights API に用いられている技術自体は、第 2 章で紹介した、テキストの自然言語処理である。具体的には、LIWC(Linguistic Inquiry and Word Count) と呼ばれる分類体系にならってあらかじめ性格と関連を示すような単語を分類しておき、テキスト上にそれぞれの分類の単語が何回表出するかを性格推定のための特徴量とする手法を用いている。これは英語のテキストの性格推定において実績のある手法であったが、Personality Insights API の日本語対応の際に那須川らによって LIWC の日本語化が行われ、日本語のテキストにおいても実スコアと高い相関をもつ性格指標の予測が可能となった [24]。

3.1.3 Personality Insights 使用例

実際にこの API を使用する際に行う前処理や、API によって出力される値の意味について、説明する。

Personality Insights API への入力

本研究では、SNS のデータを用いるので、マイクロブログの投稿を入力する時について説明する。この場合、分析対象のユーザの全ての投稿データをリスト化したものを JSON オブジェクトとして渡す HTTP リクエストを API に送信する。投稿データとは、マイクロブ

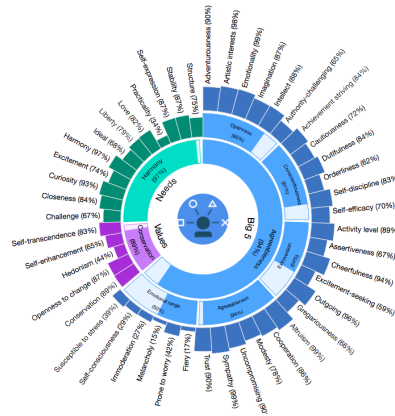


Fig. 3.1: Personality Insights による個性の計算

ログへの投稿の1回分（Twitterの場合1ツイート）に対応し、Table 3.2にあるような変数を持つオブジェクトである。表にある通り、テキスト情報以外の入力は任意である。

Personality Insights API からの出力

APIによる計算結果は、「計算に使用した単語の数」、「パーソナリティ」、「ニーズ」、「価値観」、「エラーメッセージ」などを含むオブジェクトとして返されるが、本研究ではビッグファイブの指標のみに着目したため、「パーソナリティ」についてのみ説明する。「パーソナリティ」には、Table 3.1にある35種類の性格因子について、その値の推定結果などの情報が含まれている。全ての性格の情報は、Table 3.3にあるような変数を持つオブジェクトとして得られる。

パラメータのうち、percentileとraw_scoreの違いについて説明する。raw_scoreはテキストを入力に、モデルが計算した各指標そのままの値、すなわち心理テストによる実スコアの予測値である。対し、percentileはモデルが持つサンプルのraw_scoreの分布に対し、計算

Table 3.2: Personality Insights API への入力に含まれる投稿データ

| パラメータ | 説明 |
|-----------------|-------------------------------|
| content | 投稿のテキスト情報。合計 20MB まで分析可能 |
| id (任意) | 投稿の識別番号 |
| created (任意) | 投稿された日時 |
| updated (任意) | 投稿が更新された日時 |
| contenttype(任意) | content が生のテキストであるか、HTML であるか |
| language (任意) | 言語（アラビア語、英語、スペイン語、日本語、韓国語） |
| parentid (任意) | 投稿が他の投稿と階層関係にある場合の、親の id |
| reply (任意) | 他の投稿に対する返信であるか否か |
| forward (任意) | 他の投稿の引用であるか否か |

Table 3.3: Personality Insights API による計算結果の例

| パラメータ | 説明 | 例 |
|------------|---|---------------------|
| trait_id | ビッグファイブであれば big5, 小分類であれば facet となる | big5_characteristic |
| name | 性格の名前 | Openness |
| percentile | のちに説明する raw_score を性格の分布から 0-1 に標準化したもの | 0.20715016762798294 |
| raw_score | 生の計算スコア | 0.5637599749079132 |

された値がどこに位置するか, 0 から 1 の範囲に正規化した値である. すなわち, Table 3.3 の例のように知的好奇心の percentile が 0.207 の場合, その人の知的好奇心は全体の下から 20% の水準にあることを表す.

3.2 ネットワーク上でリンクを結ぶノードの属性の相関を測る指標

本研究では, ソーシャルメディアのネットワークにおいて, 性格の似たユーザ同士がつながりやすいのか, それとも異なった性格同士がつながりやすいのかを分析する. そこで, それぞれのノードが性格という属性を持つようなネットワークにおいて, リンクで結ばれるノード同士の属性値の近さを測定する matching assortativity とよばれる指標を計算する [25]. 本節ではこの matching assortativity について説明する.

結婚ネットワークにおける matching assortativity

Fig 3.2 は, 1141 組の夫婦について, 結婚時の互いの年齢をプロットしたものである².

人をノード, 結婚関係をリンクとしたネットワークにおける, age の matching assortativity は, 夫の結婚時の年齢と, 妻の結婚時の年齢の間の Pearson の相関係数と同義である. したがって, Fig 3.2 から, 結婚関係のネットワークにおける age の matching assortativity は正であることが読み取れる.

結婚ネットワークは, リンクの一端が夫, もう一端が妻の有向グラフであり, 有向グラフにおける matching assortativity r は以下のように定義される.

$$r = \frac{\sum_{xy} xy(e_{xy} - a_x b_y)}{\sigma_a \sigma_b} \quad (3.1)$$

x と y はそれぞれリンクの始点と終点のノードの属性値, e_{xy} は全リンクのうち属性値が x の始点と属性値が y の終点の間に存在するリンクの割合である. a_x , b_y はそれぞれ全ノードのうち属性値が x , y のノードの割合であり, σ_a , σ_b はそれぞれ x , y の標準偏差である.

²Mark EJ Newman(2003) Mixing patterns in networks[25] FIG. 1 より引用

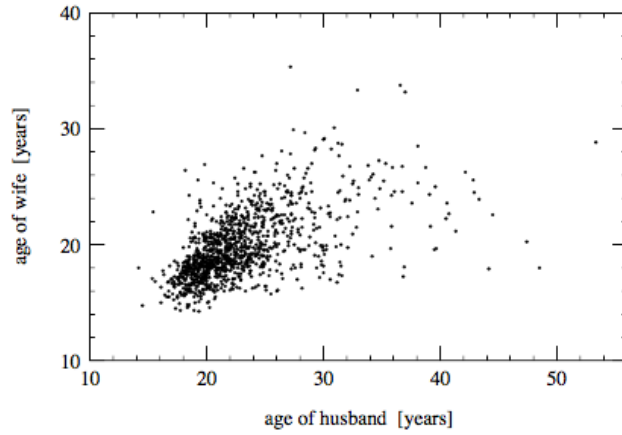


Fig. 3.2: 夫婦の結婚時の年齢

無向グラフの場合、 $a_x = b_x$ であり、また上式は属性値が離散値をとる場合であるが、確率密度関数の積分に直すことにより、連続値にこの概念を拡張することができる。この指標は -1 から 1 の間の値をとり、 1 に近いほど属性値の似たノードがつながりやすく、 -1 に近いほど属性値が離れたノードがつながりやすいことを示す。

3.3 ネットワークの分割手法

本研究では、ソーシャルメディアのネットワーク上のコミュニティのふるまいが、そのメンバーの性格とどう関係するかを分析する。そこで、本節では、ネットワークをコミュニティに分割するためのクラスタリング手法について説明する。

3.3.1 モジュラリティについて

(1) モジュラリティの定義

ネットワークをつながりの強い人同士の集団に分割するとき、その分割を評価する指標として、モジュラリティという概念を紹介する。モジュラリティとは、ネットワークのグループへの分割の質を定量化する評価関数であり、分割されたグループ内のネットワークが、ランダムにリンクを配置した場合と比べてどれだけ密であるかを表す。具体的には以下の通り。

あるグラフにおいて、 M を全エッジ数、 A をグラフの隣接行列、 k_i , k_j をノード i , ノード j の次数としたとき、モジュラリティ Q は

$$Q = \frac{1}{2M} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2M} \right) \delta(g_i, g_j) \quad (3.2)$$

で表される．ただし， $\delta(g_i, g_j)$ はノード i とノード j が同じグループに所属する時は 1，違うグループに所属する時は 0 を返す．これを整理して，

$$Q = \sum_g (e_{gg} - a_g^2) \quad (3.3)$$

ただし， e_{gg} は同じグループ g で接続しているノードの割合、 a_g^2 はリンクをランダムに配置した場合のリンク数の期待値である．モジュラリティは 0 と 1 の間の値を取り，大きいほど良い分割ができていることを表す．

この式は重みのないグラフに対して定義されるモジュラリティであるが，重み付きのグラフに対しては，式 3.2 において次数の代わりにエッジの重みの和を用いて計算する．

(2) モジュラリティ最大化によるネットワーク分割

モジュラリティが大きくなるようなネットワークの分割では，コミュニティ内のつながりが多くなり，コミュニティ間のつながりが少なくなる．したがって，モジュラリティを最大化することによりネットワークをグループ内でのつながりの強いグループに分割することが可能になる．

SNS 上の会話においても，人々は気を許したもの同士で会話する傾向が強く，未知の人との交流は稀である．したがって，ソーシャルメディアの会話ネットワークをコミュニティ分割する方法としてモジュラリティの最大化はふさわしいと考える．モジュラリティ最大化問題は，NP 困難に分類されるため，計算には工夫が必要であり，近似解を求める手法が研究されてきた．

本研究ではそのうちの一つである Louvain 法を用いる．Louvain 法はモジュラリティの局所的な最大化とコミュニティごとのノードの集約を繰り返すことで，高速でモジュラリティの最大化を実現する手法であり，分割数をあらかじめ決める必要なくクラスタリングが可能という特徴がある．

第4章 提案手法

本研究では、ソーシャルネットワーク上でユーザの性格がコミュニティ形成に与える影響を分析する。この章では、分析手法全体のフレームワークを提案手法として説明する。

4.1 手法の手順

本節では、提案手法の概略を Fig 4.1 に沿って述べる。各手順の詳細については次節以降に説明する。

- 共通のトピックネットワークに所属するユーザの抽出
- 抽出したユーザの性格推定
- メンションネットワークの構築
- ユーザの分析指標の作成とネットワーク指標の抽出
- クラスタリングとコミュニティの分析指標の作成
- ユーザの性格と行動・ネットワーク指標の相関分析（リサーチクエスチョン1・2）
- ネットワークの Matching Assortativity による分析（リサーチクエスチョン3）
- コミュニティの性格とふるまいの相関分析（リサーチクエスチョン4）

4.2 共通のトピックネットワークに所属するユーザの抽出

人と人のつながりに対して影響を与える要因として第1章で述べた通り、「その場の環境的要因」, 「ユーザの感情や思想」, 「ユーザの性格」の3要素があると本論文では考える。本手法では、できる限り純粋に性格による影響を観察するために他の要素を取り除くことを提案する。

提案手法では、分析するネットワークのトピックを統一することを考える。共通の興味やモチベーションによって接触しているユーザの集団を集めることで、「その場の環境的要因」, 「ユーザの感情や思想」を統制し、コミュニティ形成に対して性格以外の要素が与える影響を減らすことが狙いである。今後、共通の興味やモチベーションのもと接触しているユーザによって構成された SNS のネットワークを「トピックネットワーク」と呼ぶ。

あるトピックネットワークに所属するユーザを抽出する具体的な方法としては、たとえば、ユーザのプロフィール情報から興味語を抜き出し、それをもとに抽出する方法や、トピックを上手く捉えるキーワードを指定することで投稿のキーワード検索によりユーザを抽出する

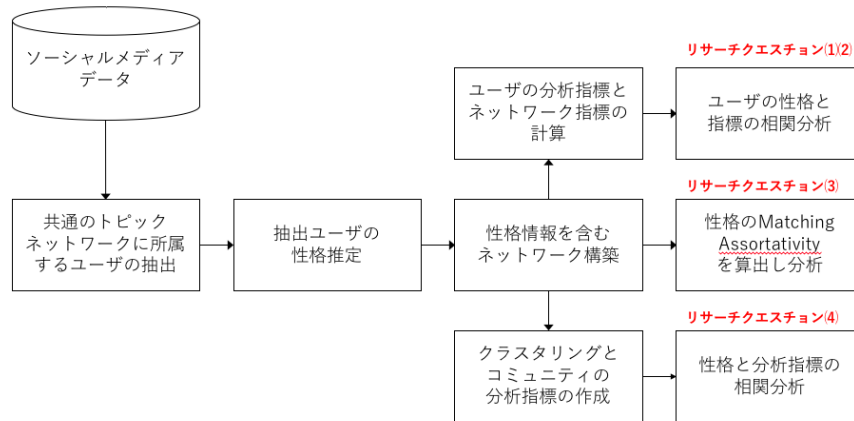


Fig. 4.1: 性格とコミュニティ形成の関係分析の手順

方法などが考えられる。ただ、共通の興味をもとに抽出したユーザが、実際にはあまり接触せず、トピックネットワークが非常にスパースになることがあるため、そのような事態に陥らない工夫する必要がある。本実験では、Twitter のハッシュタグ検索を利用して特定のトピックネットワークのユーザを抽出することでスパース性の問題を解決している。詳細は第 5 章で述べる。

4.3 抽出したユーザの性格推定

前節で抽出したユーザを対象に、性格の計算を行う。Twitter などのマイクロブログの投稿データには、しばしば他人の投稿の引用など、そのユーザの性格を捉えるには適切でない投稿が見られる。第 2 章、第 4 章で述べたように、マイクロブログからの性格の抽出には「いいね」の情報をを用いるものや、ユーザの投稿数・投稿頻度などの統計データを用いるものなど様々ある。本研究で注意しなければならないのが、ここで計算する性格は、メンションネットワークのコミュニティの分析指標との比較を行うことを目的に計算しているため、他ユーザとのメンション関係は性格推定の特徴量として使ってはならないということである。同様に、もしメンションネットワークの代わりに友人関係をリンクとする友人ネットワークを用いるのであれば、他ユーザとの友人関係を性格推定の特徴量として使ってはならない。

以上より、他人に対するメンションや他人からの引用を見分けることができ、テキストなどのユーザ内で完結した情報を用いた性格分析のモデルを利用することが望ましい。

本研究では、これらを総合的に判断して、Watson の Personality Insights API を使用することにした。

4.4 ネットワークの構築

抽出したユーザの投稿データをもとに、ユーザ間のネットワークを作成する。本実験では、ただの友人関係ではなく「日常的に会話をする関係」を捉えるために、ユーザ間のメンションの数に着目したメンションネットワークを作成した。具体的には、性格の推定を行ったユーザをネットワークのノードに対応させ、各ノードには推定した35種類の性格指標を属性値として与えた。リンクは、ユーザ間のメンションの回数を重みとして付与した。

4.5 ユーザの分析指標の作成とネットワーク指標の抽出

ユーザの性格との相関を測る指標として、以下のものを使用する。

4.5.1 作成するユーザ行動の分析指標

(1) 投稿の頻度

SNSの利用への積極性を測るために、投稿の頻度を計算した。本研究では返信、引用を含む取得した全投稿に関して、その投稿数を最古の投稿から最新の投稿までの日数で除した、1日あたりの投稿数を指標として用いた。

(2) リプライ率

SNS上の人間関係における、他人への積極性を測るために、取得した投稿について、全投稿数に対する、他のユーザへのメンション付きの投稿数の割合を計算した。本研究ではこれを指標として用いた。

(3) リプライ傾向

相手と会話が続けようとする傾向を測るために、自分に向けられた返信数に対して自分の返信数がどのくらい多いか計算した。具体的には、作成したトピックネットワーク内で、他のユーザによる自分へのメンションを含む全投稿数を集計し、それを、自分からネットワーク内のユーザに向けられたメンション数から引いた値を、指標として用いた。

4.5.2 抽出するユーザのネットワーク指標

(1) 次数

次数とは、ネットワークのノードと接続するリンクの数である。したがってメンションネットワークにおけるノードの次数とは、それぞれのユーザの返信したことのある人数を表す。本研究では、ネットワーク内でユーザが会話をした人数を表す指標として用いた。

(2) 近接中心性

近接中心性ネットワーク内の各ノードへの平均的な距離の近さを表し、ネットワーク内の総ノード数を N 、ノード i とノード j の最短距離を $d(n_i, n_j)$ とすると、ノード i の近接中心性 $C_c(n_i)$ は、

$$C_c(n_i) = \frac{N - 1}{\sum_{j=1}^N d(n_i, n_j)} \quad (4.1)$$

で表される。ただし、 $i \neq j$ 。

メンションネットワーク上では、どのくらい全ユーザに近い位置に位置するかを表すため、ネットワークに対する関与りの深度を表す。本研究では、トピックネットワークに対するユーザの依存の度合いの高さを測るための指標として用いた。

(3) 媒介中心性

媒介中心性とは、ノードが他の2つのノードの最短経路上に存在する度合いを表し、ノード i の媒介中心性 $betweeness(n_i)$ は、

$$betweeness(n_i) = \sum_{j < k} \frac{\text{点 } i \text{ を含む点 } j \text{ と点 } k \text{ の最短経路の数}}{\text{点 } j \text{ と点 } k \text{ の最短経路の数}} \quad (4.2)$$

で表される。ただし、 $i \neq j, k$

メンションネットワーク上では、どのくらい同じトピック内でも、異なるクラスタの橋渡しの役目を担っているか、ネットワークにおけるクリティカルな場所に存在するかを示す。本研究ではユーザのトピックネットワークにおいての重要度の高さを測るための指標として用いた。

4.6 クラスタリングとコミュニティの分析指標の作成

第3章で述べた通り、最適な分割数が分からないことと、適切なローカルコミュニティの検出のため、本手法では、Louvain法を用いて構築したネットワークのクラスタリングを行った。

ユーザの性格との相関を測るコミュニティのふるまいを表す指標として、以下のものを使用する。

(1) コミュニティの人数

性格によって形成されたコミュニティの規模を測るために、本研究ではコミュニティに所属するユーザの数を指標として用いた。

(2) コミュニティ内の会話数の変化

性格によって形成されたコミュニティの活性度を測るために、本研究では、コミュニティ内のメンション数を二つの時点で測り、一人あたりのメンション数の変化率を求めた。データの取得時期について、詳しくは第5章にて説明する。最初に取得したデータにおける、コミュニティ*i*内のメンション投稿の総数を $M_{before}(i)$ 、後で取得したデータにおける、コミュニティ内のメンション投稿の総数を $M_{after}(i)$ 、コミュニティの人数を $n(i)$ とすると、コミュニティの活性度 $V(i)$ は、

$$V(i) = \frac{M_{after}(i) - M_{before}(i)}{n(i)} \quad (4.3)$$

で表され、これをコミュニティが盛り上がるポテンシャルを表す指標として用いた。

4.7 ユーザの性格と行動・ネットワーク指標の相関分析

リサーチクエスチョン1, 2に対応する。

35個の性格因子について、それぞれのユーザの性格指標の値と、作成したユーザの評価指標や、抽出したユーザのネットワーク指標との間の相関係数を計算し、どの性格に相関が見られるか等を分析する。

4.8 ネットワークの Matching Assortativity による分析

リサーチクエスチョン3に対応する

35個の性格因子について、それぞれをネットワークのノードの属性と捉えた時の Matching Assortativity を計算し、どの性格が似た者同士で集まる傾向にあるかを分析する。

4.9 コミュニティの性格とふるまいの相関分析

リサーチクエスチョン4に対応する

35個の性格因子について、それぞれのコミュニティにおける性格指標の平均値や標準偏差と、作成したコミュニティの評価指標との間の相関係数を計算し、どのような特徴が見られるか等を分析する。

第5章 実験

本研究では、第4章で説明した分析のフレームワークのもと、ソーシャルメディア上のユーザの性格とコミュニティの関係を解明するために行った実験について説明する。

5.1 データの収集

5.1.1 データ収集に用いるソーシャルメディア

分析には、Twitterの投稿データを使用した。Twitterは日本で最も普及したSNSプラットフォームの一つで、2017年10月時点で日本だけで4500万¹のアクティブユーザが存在する。大きな特徴として、投稿の長さに140語以内という文字数の制限があるため、些細でとりとめのない出来事も積極的に投稿するという文化が育まれたことがあげられる。それにより、投稿は全体的にパーソナルな意見が表出しやすく、投稿やユーザの行動に性格による差が出やすいと思われる。同時に、投稿の引用（リツイート）や投稿への返信（メンション）に対する敷居も低いため、会話を密に行うコミュニティの観測もしやすいと思われる。以上より、Twitterが性格によるコミュニティ形成を観測するための日本語のトピックネットワークが最も取得できる可能性の高いプラットフォームであると考え、分析対象として選択した。

5.1.2 特定のトピックネットワークを抽出

第4章で述べた通り、まずは共通のトピックネットワークに属するユーザを抽出する。使用するユーザ群は、同じ興味を持つだけでなく、メンションネットワークが作成できるようにユーザ間で交流がなされている必要がある。本研究では、Twitterのハッシュタグ検索により得られるツイートの投稿者を取得する、という方法を用いた。

ハッシュタグとは、#記号とそれに続くトピック名などの文字列のことをさす。ハッシュタグを投稿に含めることで、読み手にその投稿が何についての投稿か容易に伝えることができる。Twitterには、同一のハッシュタグを含むツイートを検索画面などで一覧できる機能がある。このハッシュタグ検索により、同一の話題に関するツイートを容易に取得でき、どのようなハッシュタグが盛り上がっているかを見ることで、その時の社会で人気な話題、旬な話題を把握することが可能となる。そしてなにより、ハッシュタグをつけて投稿することで、同じ話題のTweetを容易に探すことができ、共通の話題で盛り上がるができる。

¹参照元：<https://twitter.com/TwitterJP/status/923671036758958080>

Table 5.1: トピックネットワーク抽出のためのハッシュタグ検索

| ネットワーク名 | 検索したクエリ | ネットワーク名 | 検索したクエリ |
|---------|---|---------|--|
| ファッション | ”#お洒落さんと繋がりたい” ”#おしゃれさんと繋がりたい” ”#お洒落な人と繋がりたい” | イラスト | ”#絵師さんと繋がりたい” ”#絵描きさんと繋がりたい” ”#絵師さんとながりたい” |
| 映画 | ”#映画好きと繋がりたい” ”#映画好きな人と繋がりたい” ”#洋画好きと繋がりたい” | 他者依存 | ”#メンヘラさんと繋がりたい” ”#病み垢さんと繋がりたい” ”#ヤンデレさんと繋がりたい” |

Table 5.2: タイムライン検索

| トピックネットワーク | タイムラインを検索した人数 | 平均取得ツイート数 |
|------------|---------------|-----------|
| ファッション | 923 | 1604.0 |
| 映画 | 490 | 1624.2 |
| イラスト | 947 | 1758.2 |
| 他者依存 | 916 | 1417.6 |

中にはユーザ同士の交流を目的とするハッシュタグも存在し、検索するハッシュタグを適切に選択することで、スパースでない同じ興味をもつネットワークのツイートが取得できると考える。

スパースでないトピックネットワークを抽出するための検索クエリ設定

本実験では、まず Table 5.1 のようなハッシュタグ検索を行い 4 つの異なるトピックについて、twitterAPI を用いそれぞれ最新 2500 ツイートを取得した。検索したハッシュタグとして Table 5.1 にあるものを選んだのは、「～～について繋がりたい」というハッシュタグが本研究において理想的であると考えたからである。まず、「～～について繋がりたい」というハッシュタグは日本では非常に利用者が多く、多種多様なサンプルが取得できる。そして、これらのハッシュタグでは、ユーザ同士のつながることへのモチベーションが強いためスパースなネットワークになりにくい。最後に、このようなハッシュタグのネットワークで生じたグループは「繋がりたい」のグループであり、知り合ってからあまり長い時間が立っていないと考えられるため、ユーザ間の人間関係もまだ定まっておらず、性格が与えるコミュニティの変化への影響をより敏感に捉えられるのではないか、という期待もある。

Table 5.3: TwitterAPIによりツイートを検索した日時

| ネットワーク | ハッシュタグ検索 | タイムライン検索（1回目） | タイムライン検索（2回目） |
|--------|------------|---------------|---------------|
| ファッション | 2017/12/25 | 2017/12/25 | 2018/01/15 |
| 映画 | 2017/12/25 | 2017/12/25 | 2018/01/15 |
| イラスト | 2018/01/07 | 2018/01/07 | 2018/01/15 |
| 他者依存 | 2018/01/07 | 2018/01/07 | 2018/01/15 |

性格分析を行うのに適さないユーザの除外

ハッシュタグ検索によって得られたツイートのユーザ情報を抽出し、ユーザタイムラインを取得した。ユーザタイムラインとは、そのユーザの過去のツイートが時系列に並んだものをさし、本研究では取得したユーザタイムライン（最大 2000 投稿）のデータを元に性格分析を行った。

取得したユーザのうち、ツイート数が 500 以下のユーザは、Personality Insights API で性格分析を行うにはデータ量が足りないと判断し、構築するトピックネットワークのユーザからは除外した。そこで、最終的に Table 5.2 の人数のユーザについて、Personality Insights API を用いた性格分析を行った。

5.1.3 コミュニティの分析指標作成のためのデータ収集

4.6 節で述べたように、本研究では、コミュニティのふるまいを観察するための評価指標として、コミュニティ内の会話数の変化を用いる。そのため、データの取得を行ったのち、数週間後にコミュニティ内の会話数をあらためて測るため、同じユーザについて二回目のタイムライン検索を行った。

Table 5.3 にそれぞれの Twitter データの取得を行った日時を示す。

Table 5.4: Twitter タイムラインと Watson API への入力の対応関係

| Watson API への入力 | Twitter タイムラインの要素 |
|-----------------|--------------------------------------|
| content | 1 つのツイートのテキスト部分 |
| id | ツイートの status id |
| created | ツイート時間のデータ（取得時は UTC のため、日本時間に直す必要あり） |
| language | ja（日本語） |
| reply | ツイートが誰かに対するメンションであるか否か（ブール値） |
| forward | ツイートが誰かの投稿のリツイートであるか否か（ブール値） |

Table 5.5: 構築したネットワーク

| ネットワーク | ノード数 | リンク数 | 次数が0のノード | 最大のコンポーネント |
|--------|------|------|----------|------------|
| ファッション | 923 | 742 | 451 | 363 |
| 映画 | 490 | 623 | 226 | 223 |
| イラスト | 947 | 662 | 407 | 430 |
| 他者依存 | 916 | 1575 | 261 | 621 |

5.2 ユーザの性格の計算

Personality Insights API に性格の計算をさせるために、Twitter のタイムラインデータを第3章の Table 3.2 にあるような形になるように整えた。twitterAPI により取得したタイムラインのデータと、Personality Insights API の入力変数の対応関係は、Table 5.4 に示す。

第3章で述べたように、Personality Insights API によって計算される性格指標の値には、心理テストによる実スコアの予測である `raw_score` と、Personality Insights API のサンプルデータと照らし合わせた順位を 0 から 1 に正規化した `percentile` の二種類が存在する。本研究では、既存研究と同様の相関分析を行うため、API 独自のスコアである `percentile` ではなく、`raw_score` を用いて相関分析を行う。

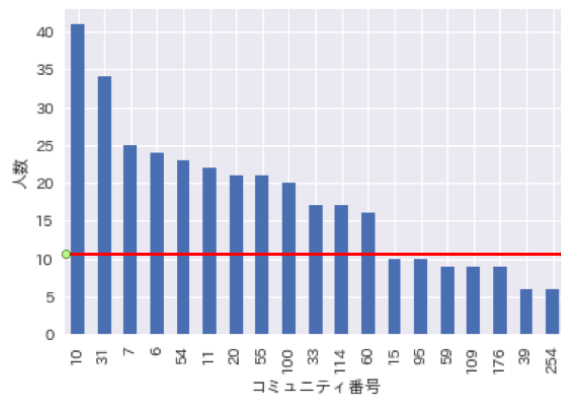
5.3 ネットワークの構築とコミュニティ分割

取得した Twitter ユーザのタイムライン情報から、ユーザ同士のメンション回数をリンクの重みとしたメンションネットワークを作成した。また、それぞれのノードには、前節で計算した 35 種類の性格指標の値を属性値として付与した。

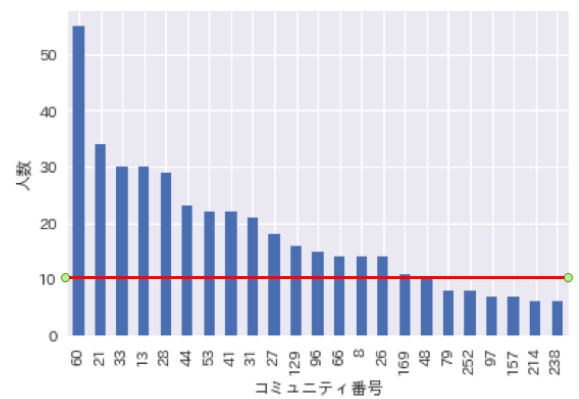
ネットワークの情報を Table 5.5 に示す。ネットワークのノードの総数のおよそ半分ほどが、リンクを 1 つももたないスパースなネットワークになってしまった。しかし、最大のコンポーネント、すなわちネットワークの中で分断されていない最も大きな部分グラフは十分に大きいため、Louvain 法によるクラスタリングを行い、大きなクラスタをコミュニティと捉えた。

クラスタリングした結果得られた各クラスタの所属人数のグラフを Fig 5.1 に示す。このうち、本研究ではある程度の頻度の会話と、人間関係のダイナミクスが期待できることから、所属人数が 10 人以上のクラスタを、分析対象のクラスタとした。

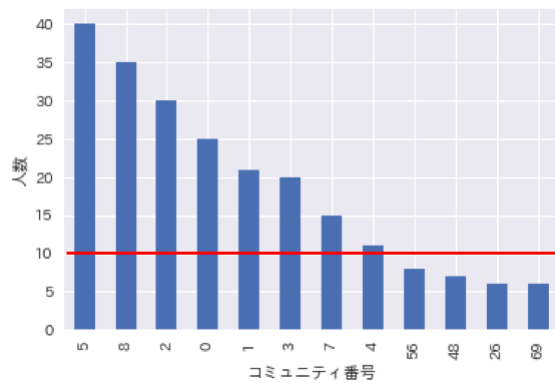
したがって、ファッションネットワークの上位 12 組、映画ネットワークの上位 8 組、イラストネットワークの上位 16 組、他者依存ネットワークの上位 19 組についてコミュニティのふるまいと性格の相関関係を分析する。



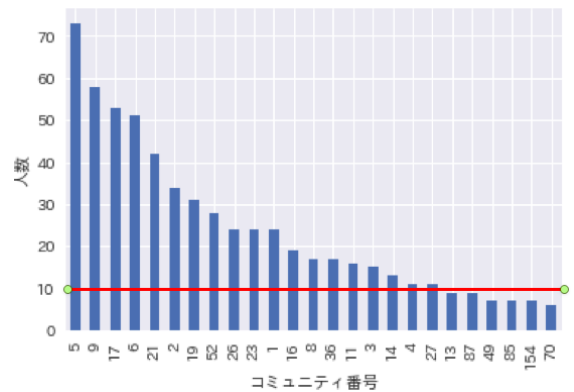
(a) ファッションネットワーク



(b) イラストネットワーク



(a) 映画ネットワーク



(b) 他者依存ネットワーク

Fig. 5.1: Louvain 法によって分割されたコミュニティごとの人数 (5人以上)

5.4 相関分析用の分析指標の抽出

第4章で提案した通り、各々のユーザ・クラスタについて分析指標の計算を行った。

この分析指標と性格の推定値の相関関係を本実験では分析するが、分析にあたり、ネットワークのスパース性を鑑みて、第4章で紹介した分析手法のうち、リプライ傾向、次数、近接中心性、媒介中心性に関してはスパース性の影響を軽減させるための工夫を施した。

これらの分析指標はトピックネットワーク上で孤立したユーザに対し、0の値をとる。ノードの次数が0になる原因、すなわちトピックネットワーク上のユーザが誰とも会話をしない原因は、環境要因などさまざま考えられ、性格的な特徴に起因するとは考えづらい。よって、これらの指標の相関分析を行う際は、次数が0のユーザのデータは排除して、相関分析を行った。

5.5 リサーチクエスション1の結果の予想

本節では、APIの妥当性の検証のための、本研究のリサーチクエスション1について、先行研究をもとにした結果の仮説を説明する。この仮説通りの結果が観測されれば、APIを用いたこの実験手法は妥当なものといえ、その後の分析は有意義なものと言える。

リサーチクエスション1では、第4章で述べた通り、ユーザそれぞれの性格と、投稿の頻度やリプライ率、リプライ傾向の間の相関係数を計算する。

分析指標のうち、投稿の頻度はSNSへの積極性や自己表現への欲求が関わっていると考ええる。対して、リプライ率とリプライ傾向は他人への積極性が関わっていると考ええる。第2章の関連研究より、RyanらはOpenness（知的好奇心）がSNSへの積極性にに関わり [13]、SeidmanはNeuroticism（神経質傾向）（APIのEmotional Range（感情起伏）に対応）の高さとConscientiousness（誠実性）の低さが自己表現欲求に関わることがわかっている [15]。また、SeidmanはAgreeableness（協調性）とNeuroticism（神経質傾向）が他人への積極性に関わることが判明している

以上より、Openness（知的好奇心）とEmotional Range（感情分析）が投稿の頻度に対して正の相関を、Conscientiousness（誠実性）が投稿の頻度に対して負の相関を示すこと。そして、Agreeableness（協調性）、Emotional Range（感情起伏）とリプライ率やリプライ傾向は、正の相関を示すことを、リサーチクエスション1における仮説とした。

5.6 Homophilyな作用が起こる性格についての仮説

第2章で紹介したBarrickらの研究 [10] によると、チームによる仕事においては、conscientiousness（誠実性）やagreeableness（協調性）、emotional stability（感情の安定）がパフォーマンスを向上させる性格因子である。ただ、それと同時にチーム内の誠実性にもしばらつきがあると、それはパフォーマンスを妨げる作用をしてしまう。

もし、誠実性のばらつきによるパフォーマンスの低下が、お互いの物事への取り組み方や自己管理、自制力の差に対するお互いの潜在的なストレスに起因するものであれば、仕事ではない Twitter におけるコミュニティ形成においても、人々は同種のストレスを感じかわりを避ける傾向があると考えられる。したがって、誠実性についてこのような、高い人は高い人同士、低い人は低い人がつながるという Homophily な作用が本当に現実で観測されるのか、実験を行った。

第6章 結果

本章では、実験の結果について述べる。第5章5.5節のリサーチクエスチョンに基づいて、結果を表示する。

6.1 リサーチクエスチョン1：外部モデルを使用する分析の妥当性の確認

6.1.1 ツイートの頻度

第4章の通り、ユーザのツイートの頻度はその人の最初のツイートから最後のツイートまでの間の、1日あたりのツイート回数によって表される。

この指標と、性格の推定値の間の Pearson 相関係数を求め、大きな相関が見られる性格について Table 6.1 にまとめた。

まず、仮説で言われているように、SNS への積極性を示す Openness（知的好奇心）との間には、弱い正の相関を示すことが確認された。しかし、自己表現欲求と関わる Emotional range（感情起伏）との正の相関や、Conscientiousness（誠実性）との負の相関は、確認することはできなかった。

Openness（知的好奇心）についてより深く調べるため、Table 6.1 には Openness（知的好奇心）に属する小分類の性格指標（ファセット）と投稿頻度の相関も同様に示した。Openness（知的好奇心）の小分類のなかでは、Imagination（想像力）と Authority-challenging（現状打破）の性格が投稿頻度に関わることが判明した。

Table 6.1: トピックネットワークごとの性格の推定値とツイートの頻度の関係

| 性格 | ファッション | 映画 | イラスト | 他者依存 | 平均値 |
|---------------------------------|-------------|-------------|-------------|-------------|-------------|
| A Openness | 0.18 | 0.04 | 0.12 | 0.14 | 0.12 |
| A4 Imagination | 0.19 | 0.07 | 0.07 | 0.09 | 0.11 |
| A6 Authority-challenging | 0.18 | 0.05 | 0.09 | 0.13 | 0.11 |
| B Conscientiousness | -0.11 | 0.06 | 0.06 | -0.03 | -0.01 |
| C Extraversion | -0.07 | -0.06 | -0.01 | -0.05 | -0.05 |
| D Agreeableness | -0.19 | -0.04 | -0.04 | -0.14 | -0.10 |
| E Emotional range | 0.07 | 0.02 | 0.01 | 0.08 | 0.04 |

6.1.2 リプライ率

第4章の通り、ユーザのリプライ率はその人のツイートの総数のうち、別のユーザに対するメンションを含むツイートの割合によって表される。

この指標と、性格の推定値の間の Pearson 相関係数を求め、大きな相関が見られる性格について、Table 6.2 にまとめた

まず、仮説で言われている通り、Agreeableness（協調性）がリプライ率と正の相関を示すことがわかった。特に、Agreeableness（協調性）の小分類の中でも、Cooperation（共同性）、Modesty（謙虚さ）、Uncompromising（強硬さ）の3指標とは非常に大きな相関を示した。

また、仮説では Emotional range（感情起伏）がリプライ率と正の相関を示すと予想されていたが、そのような相関は確認できなかった。

他に、Openness（知的好奇心）とリプライ率との間に大きな負の相関が見受けられた。Openness（知的好奇心）に属する小分類の中で、Adventurousness（大胆性）は大きな正の相関を、反対に Imagination（想像力）は大きな負の相関を示すことが明らかとなった。

Table 6.2: トピックネットワークごとの性格の推定値とリプライ率の関係

| | 性格 | ファッション | 映画 | イラスト | 他者依存 | 平均値 |
|-----------|--------------------------|--------------|--------------|--------------|--------------|--------------|
| A | Openness | -0.25 | -0.07 | -0.27 | -0.06 | -0.16 |
| A1 | Adventurousness | 0.16 | 0.20 | 0.23 | 0.32 | 0.23 |
| A4 | Imagination | -0.29 | -0.27 | -0.36 | -0.31 | -0.31 |
| B | Conscientiousness | -0.16 | -0.10 | -0.04 | 0.12 | -0.04 |
| C | Extraversion | -0.10 | 0.13 | -0.06 | -0.06 | -0.02 |
| D | Agreeableness | 0.10 | 0.29 | 0.07 | 0.06 | 0.13 |
| D2 | Cooperation | 0.31 | 0.46 | 0.24 | 0.30 | 0.33 |
| D3 | Modesty | 0.20 | 0.39 | 0.16 | 0.15 | 0.23 |
| D4 | Uncompromising | 0.27 | 0.45 | 0.21 | 0.29 | 0.30 |
| E | Emotional range | -0.17 | 0.07 | -0.06 | -0.03 | -0.05 |

6.1.3 リプライ傾向

第4章の通り，ユーザのリプライ傾向はその人に対して向けられたメンションの数から，その人から発されたメンションの数を引いた値によって表される．

この指標と，性格の推定値の間の Pearson 相関係数を求めたが，性格の推定値とリプライ傾向の間に，特筆すべき相関は見られなかった．もちろん，仮説である Agreeableness（協調性）や Emotional range（感情起伏）に関する正の相関も，確認されなかった．

参考として Table 6.3 に，ビッグファイブと，ある程度の正の相関を見せた Cautiousness（注意深さ），Uncompromising（強硬さ）について紹介する．

Table 6.3: トピックネットワークごとの性格の推定値とリプライ傾向の関係

| 性格 | ファッション | 映画 | イラスト | 他者依存 | 平均値 |
|--------------------------|-------------|-------------|-------------|-------------|-------------|
| A Openness | -0.11 | 0.06 | -0.07 | -0.03 | -0.04 |
| B Conscientiousness | -0.07 | -0.03 | 0.04 | -0.11 | -0.04 |
| B2 Cautiousness | 0.17 | 0.26 | 0.08 | 0.12 | 0.16 |
| C Extraversion | -0.07 | 0.09 | -0.08 | -0.11 | -0.04 |
| D Agreeableness | 0.07 | 0.11 | -0.07 | -0.03 | 0.02 |
| D4 Uncompromising | 0.16 | 0.37 | 0.03 | 0.07 | 0.16 |
| E Emotional range | -0.03 | 0.09 | -0.03 | 0.05 | 0.02 |

6.2 リサーチクエスション2：集団の人間関係における個人の位置づけに性格が与える影響の分析

6.2.1 次数

ユーザの性格の推定値と、メンションネットワーク上の次数の間の Pearson 相関係数を求めたが、性格の推定値とユーザの次数の間に、特筆すべき相関は見られなかった。

参考として、Table 6.4 にいくつかの性格について、計算された相関係数を紹介する。

下表に示すビッグファイブや Artistic interests（芸術的関心度）、Intellect（思考力）、Cheerfulness（明朗性）など、映画のトピックネットワークのみ極端な値をとることが多く見受けられた。

Table 6.4: トピックネットワークごとの性格指標と次数の関係

| | 性格 | ファッション | 映画 | イラスト | 他者依存 | 平均値 |
|----|--------------------|--------|-------------|-------|-------|------|
| A | Openness | 0.04 | 0.27 | -0.04 | 0.02 | 0.07 |
| A2 | Artistic interests | -0.07 | 0.30 | -0.07 | 0.00 | 0.04 |
| A5 | Intellect | -0.02 | 0.34 | 0.06 | 0.04 | 0.10 |
| C3 | Cheerfulness | -0.03 | 0.33 | -0.05 | -0.03 | 0.06 |
| E | Emotional range | -0.06 | 0.21 | -0.11 | 0.05 | 0.02 |

6.2.2 近接中心性

ユーザの性格の推定値と，メンションネットワーク上の近接中心性の間の Pearson 相関係数を求め，大きな相関が見られる性格について Table 6.5 にまとめた．

ビッグファイブの中で Conscientiousness（誠実性）のみが負の相関を示しており、Cautiousness（誠実性）の小分類である Dutifulness（忠実さ），Self-discipline（自制力），Self-efficacy（自己効力感）の3つの性格についても，近接中心性との間に強い負の相関があることがわかった．

Table 6.5: トピックネットワークごとのビッグファイブと近接中心性の関係

| 性格 | | ファッション | 映画 | イラスト | 他者依存 | 平均値 |
|-----------|--------------------------|--------------|--------------|--------------|--------------|--------------|
| A | Openness | -0.13 | 0.19 | -0.12 | -0.07 | -0.03 |
| B | Conscientiousness | -0.25 | -0.05 | -0.06 | -0.23 | -0.15 |
| B3 | Dutifulness | -0.10 | -0.12 | 0.04 | -0.23 | -0.10 |
| B5 | Self-discipline | -0.24 | 0.03 | -0.20 | -0.26 | -0.17 |
| B6 | Self-efficacy | -0.14 | 0.08 | -0.13 | -0.16 | -0.09 |
| C | Extraversion | 0.01 | -0.06 | -0.03 | -0.04 | -0.03 |
| D | Agreeableness | -0.03 | -0.23 | 0.07 | -0.03 | -0.06 |
| E | Emotional range | -0.00 | 0.17 | -0.07 | 0.08 | 0.04 |

6.2.3 媒介中心性

ユーザの性格の推定値と，メンションネットワーク上の媒介中心性の間の Pearson 相関係数を求め，Table 6.6 にまとめた．

ビッグファイブと媒介中心性の間には，下表の通り，相関は見られなかった．小分類含めた全ての性格の中でも，ある程度の相関を示したのは下表に示した Adventurousness のみであり，弱い正の相関を示している．

Table 6.6: トピックネットワークごとの性格指標と媒介中心性の関係

| 性格 | | ファッション | 映画 | イラスト | 他者依存 | 平均値 |
|-----------|------------------------|-------------|-------------|-------------|-------------|-------------|
| A | Openness | -0.07 | 0.08 | 0.02 | -0.01 | 0.00 |
| A1 | Adventurousness | 0.17 | 0.20 | 0.08 | 0.08 | 0.13 |
| B | Conscientiousness | 0.10 | -0.05 | -0.01 | -0.02 | 0.00 |
| C | Extraversion | -0.06 | 0.10 | 0.02 | -0.01 | 0.01 |
| D | Agreeableness | -0.04 | 0.03 | 0.02 | -0.06 | -0.01 |
| E | Emotional range | -0.05 | 0.09 | -0.08 | -0.02 | -0.02 |

6.3 リサーチクエスチョン3：Homophily な傾向を示す性格はあるか

本節では、性格の推定値が近いもの同士ほど会話が発生しやすく、離れていると会話が発生しにくいような性格の種類が存在するかを確認する。

次のページの Table 6.7 に各性格のネットワーク上の matching assortativity を 4 つネットワークの平均値が高い順に並べた。Table 6.7 において matching assortativity が高い Gregariousness と、matching assortativity の低い Self-consciousness について、実際にネットワーク内のリンクがどのような性格の値のノード間に分布しているか、映画ネットワークを例に Fig. 6.1 に示す。

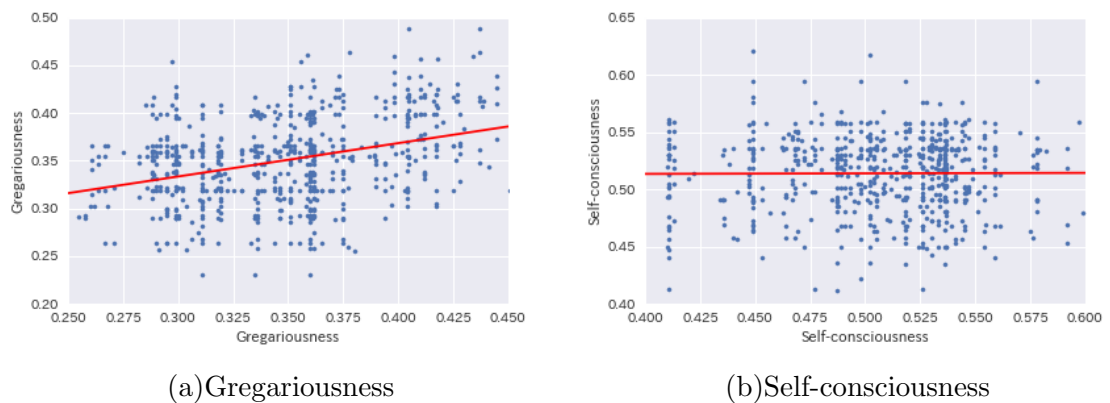


Fig. 6.1: 映画ネットワークのリンクの両端の性格値 (a)Gregariousness (b)Self-consciousness

Table 6.7: トピックネットワークごとの各性格についての matching assortativity

| | 性格 | ファッション | 映画 | イラスト | 他者依存 | 平均値 |
|-----------|---------------------------|--------------|--------------|--------------|--------------|--------------|
| C6 | Gregariousness | 0.323 | 0.348 | 0.260 | 0.109 | 0.260 |
| B | Conscientiousness | 0.207 | 0.201 | 0.324 | 0.168 | 0.225 |
| C3 | Cheerfulness | 0.295 | 0.222 | 0.146 | 0.115 | 0.195 |
| A2 | Artistic interests | 0.165 | 0.259 | 0.212 | 0.128 | 0.191 |
| E2 | Prone to worry | 0.311 | 0.052 | 0.201 | 0.119 | 0.171 |
| B5 | Self-discipline | 0.249 | 0.065 | 0.159 | 0.206 | 0.170 |
| A3 | Emotionality | 0.275 | 0.128 | 0.198 | 0.071 | 0.168 |
| A | Openness | 0.245 | 0.173 | 0.173 | 0.072 | 0.166 |
| A6 | Authority-challenging | 0.135 | 0.215 | 0.151 | 0.080 | 0.145 |
| B3 | Dutifulness | 0.217 | 0.034 | 0.191 | 0.137 | 0.145 |
| E6 | Susceptible to stress | 0.230 | 0.066 | 0.149 | 0.125 | 0.143 |
| D2 | Cooperation | 0.241 | 0.067 | 0.138 | 0.115 | 0.140 |
| A4 | Imagination | 0.261 | 0.011 | 0.113 | 0.130 | 0.129 |
| E1 | Fiery | 0.201 | 0.046 | 0.131 | 0.128 | 0.126 |
| B4 | Orderliness | 0.114 | 0.097 | 0.198 | 0.093 | 0.126 |
| A5 | Intellect | 0.132 | 0.177 | 0.126 | 0.044 | 0.120 |
| E3 | Melancholy | 0.125 | 0.059 | 0.145 | 0.129 | 0.114 |
| D | Agreeableness | 0.154 | 0.104 | 0.086 | 0.109 | 0.113 |
| D5 | Sympathy | 0.279 | 0.020 | 0.119 | 0.033 | 0.112 |
| C5 | Outgoing | 0.114 | 0.037 | 0.132 | 0.124 | 0.102 |
| D4 | Uncompromising | 0.185 | 0.080 | 0.102 | 0.038 | 0.101 |
| C4 | Excitement-seeking | 0.174 | 0.013 | 0.177 | 0.020 | 0.096 |
| B1 | Achievement striving | 0.044 | 0.043 | 0.168 | 0.127 | 0.095 |
| B2 | Cautiousness | 0.181 | 0.023 | 0.118 | 0.057 | 0.095 |
| C2 | Assertiveness | 0.098 | 0.013 | 0.180 | 0.067 | 0.090 |
| A1 | Adventurousness | 0.102 | 0.062 | 0.066 | 0.120 | 0.088 |
| E | Emotional range | 0.170 | 0.024 | 0.064 | 0.082 | 0.085 |
| B6 | Self-efficacy | 0.147 | 0.011 | 0.054 | 0.121 | 0.083 |
| D3 | Modesty | 0.237 | 0.001 | 0.035 | 0.037 | 0.077 |
| D1 | Altruism | 0.146 | 0.077 | 0.012 | 0.065 | 0.075 |
| E5 | Self-consciousness | 0.128 | -0.022 | 0.117 | 0.066 | 0.072 |
| C1 | Activity level | 0.065 | 0.050 | 0.131 | -0.000 | 0.062 |
| E4 | Immoderation | 0.125 | -0.017 | 0.085 | 0.042 | 0.059 |
| D6 | Trust | 0.077 | 0.050 | 0.040 | 0.052 | 0.055 |
| C | Extraversion | 0.056 | -0.001 | 0.014 | 0.024 | 0.023 |

6.4 リサーチクエスチョン4：コミュニティ内の性格の分布がコミュニティのふるまいに与える影響の分析

6.4.1 コミュニティの人数に対する寄与

コミュニティの人数と、コミュニティに所属するユーザの性格の推定値の平均、標準偏差の間の Pearson 相関係数を求めた。

Table 6.8 に4つのネットワークとの相関の平均の絶対値が0.2以上になる性格について、集計した。

Table 6.8: トピックネットワークごとのコミュニティの性格の平均や分散とコミュニティの人数の関係

| | 性格 | ファッション | 映画 | イラスト | 他者依存 | 平均値 |
|----|-------------------------|--------|-------|-------|-------|-------|
| 3 | Emotionality (平均) | -0.45 | -0.10 | -0.13 | -0.17 | -0.21 |
| 18 | Excitement-seeking (平均) | 0.36 | 0.04 | 0.28 | 0.20 | 0.22 |
| 23 | Cooperation (平均) | -0.50 | 0.15 | -0.32 | -0.34 | -0.25 |
| 30 | Prone to worry (平均) | 0.47 | 0.29 | 0.24 | 0.34 | 0.34 |
| 31 | Melancholy (平均) | 0.16 | 0.17 | 0.21 | 0.34 | 0.22 |
| 10 | Dutifulness (標準偏差) | 0.07 | 0.37 | 0.28 | 0.24 | 0.24 |
| 20 | Gregariousness (標準偏差) | 0.76 | -0.03 | 0.38 | 0.50 | 0.40 |
| 29 | Fiery (標準偏差) | 0.14 | -0.43 | -0.0 | -0.54 | -0.21 |

6.4.2 コミュニティの活性化に対する寄与

第4章の通り，コミュニティの活性度とは一回目のタイムライン検索から二回目のタイムライン検索にかけてのコミュニティ内の一人あたりのメンション付き投稿数の変化率である．

この指標と，コミュニティに所属するユーザの性格の推定値の平均，標準偏差の間の Pearson 相関係数を求めた．

Table 6.9 に4つのネットワークとの相関の平均の絶対値が0.2以上になる性格について，集計した．

Table 6.9: トピックネットワークごとのコミュニティの性格の平均や分散とコミュニティの活性化の関係

| | 性格 | ファッション | 映画 | イラスト | 他者依存 | 平均値 |
|----|---------------------------|--------|-------|-------|-------|-------|
| 8 | Achievement striving (平均) | -0.36 | -0.46 | 0.3 | -0.36 | -0.22 |
| 3 | Emotionality (標準偏差) | 0.06 | -0.14 | -0.23 | -0.64 | -0.24 |
| 10 | Dutifulness (標準偏差) | -0.23 | 0.27 | 0.63 | 0.18 | 0.21 |
| 13 | Self-efficacy (標準偏差) | 0.34 | -0.22 | 0.64 | 0.69 | 0.36 |
| 25 | Uncompromising (標準偏差) | 0.00 | -0.18 | -0.53 | -0.21 | -0.23 |
| 31 | Melancholy (標準偏差) | 0.10 | 0.26 | -0.03 | 0.68 | 0.25 |

第7章 考察

本章では、第1章で述べたリサーチクエスチョンごとに、分析結果に関する考察を行う。

7.1 外部モデルを使用した提案手法の妥当性の確認

本節では、ユーザの性格と行動の関係を分析した結果 6.1 についての考察を行う。

ユーザのツイートの頻度との関係

性格とツイートの頻度の関係については、Emotional range（感情起伏）が正の相関、Conscientiousness（誠実性）が負の相関を示すと思われていたが、得られた相関係数はそれぞれ 0.04, -0.01 であり、はっきりとした相関を確認することはできなかった。

SNS への積極的な参加を示す Openness（知的好奇心）については、仮説通り、正の相関を確認することができた。

Imagination（想像力）、Authority-challenging（現状打破）は両者とも正の相関を示した。この二つの性格は Openness（知的好奇心）の中でも、現実の事実を超越しようとする姿勢と関わっている。従って、この性格傾向の強さは現実世界からソーシャルメディアの世界へ移ろうとする作用を促進する影響を持ち、ツイートの頻度の高さにつながるのではないかと推測する。

リプライ率

他人に向けた返信の傾向については、Agreeableness が仮説通り、正の相関を示すことがわかった。特に、Cooperation（共同性）、Modesty（謙虚さ）、Uncompromising（強硬さ）の3つが強い正の相関を示すことが分かった。他人との協調の姿勢を見せる人間は、人との関わりを好む傾向があることから、メンション付きのツイートの増加につながったと考える。

Adventurousness（大胆性）、Imagination（想像力）はともに Openness（知的好奇心）のカテゴリに属されるが、前者は強い正の相関を、後者は強い負の相関を示すことが分かった。これは、Adventurousness（大胆性）は外向きの知的好奇心であるのに対し、Imagination（想像力）は内向きの知的好奇心であることが関係している可能性がある。

リプライ傾向

リプライ傾向についてはほとんど性格との間で相関が見受けられなかった．相関が見受けられなかった要因としては二つ考えられる．

一つ目は，会話の持続させようとする傾向は，その時の話題や状況による影響を受けるものであり，個人の性格が寄与することはあまり多くはないという考えである．

しかし，共感性の高い人は，そうでない人と比べて会話をつなごうとする意識が強いと考えるのは飛躍した話ではない．相関が見られなかった二つ目の理由として，分析指標が不適切だった可能性があると考えられる．本研究で使用した分析指標は，会話が返信の送り合いであるという仮定のもと，他人から話しかけられた回数に対して，自分が余分にどれだけ話しかけたかを見た．しかし，本研究で用いた「繋がりたい」トピックネットワークにおいて，メンション付きの発言の多くが投稿に対する感想，という形で送られたもので，長い会話へ発展する頻度は多くなかった．したがって，長い会話のみに着目し，その上でその会話を持続させる傾向を測定する必要があると考える．

分析の妥当性の確認

以上より，Openness（知的好奇心）の高いユーザが SNS への積極的な参加傾向が高く，Agreeableness（協調性）が高いユーザは人とつながろうとする傾向が強いことがされた．これらは，Ryan ら [13] や Seidman[15] などが示した性格と SNS 使用の関係性と矛盾しない結果である．したがって，Personality Insights API を使用した本研究は有効なものであると考える．一方で，リプライ傾向との関係性の特定等の点で課題が残った．

7.2 集団の人間関係における個人の位置づけに性格が与える影響の分析

本節では，ユーザの性格とネットワーク指標の関係を分析した結果 6.2 についての考察を行う．

次数との比較

次数については，どの性格とも大きな相関は見受けられなかった．ユーザ数の少ない映画のトピックネットワークだけが他のトピックネットワークと比べて外れた値を示していることや，そもそも次数が示すのがユーザの会話相手の総数ではなく同じトピックネットワークに属する会話相手であることを鑑みると，次数は性格の影響よりも抽出したトピックネットワークのサイズやテーマの影響を強く受けるものと思われる．

近接中心性との比較

近接中心性は、トピックネットワークに対するユーザの依存の度合いを示す指標であると考えられる。これは、Conscientiousness（誠実性）とその小分類である Dutifulness（忠実さ）、Self-discipline（自制力）、Self-efficacy（自己効力感）が負の相関を示した。したがって、誠実な人や自製の聞く人はトピックネットワークの中心部に位置するまで SNS 上の関係には依存せず、自分にとって必要な場面でのみソーシャルネットワークの一部分を「つまみ食い」するようにソーシャルメディアと関わっているのだと推測される。

媒介中心性との比較

媒介中心性について、相関が見られるような性格は Adventurousness のみであった。Adventurousness の高い人は、トピックネットワーク上の様々なコミュニティに積極的に関わりを求めることで、ネットワークのクラスタ間をつなぐような役割を果たしているものと推測される。

7.3 Homophily な傾向を示す性格はあるか

Gregariousness（社交性）、Conscientiousness（誠実性）、Cheerfulness（明朗性）、Artistic interests（芸術的関心度）の四つの性格は、傾向が似た者同士で集まりやすいことが結果から判明した。それぞれの性格が含まれるビッグファイブのカテゴリを集計すると、Openness（知的好奇心）が1つ、Extraversion（外向性）が2つ、Conscientiousness（誠実性）が1つであり、特別に正の相関が強いビッグファイブが存在するかは確認されなかった。

しかし、当初の仮説通り、Conscientiousness（誠実性）は全 35 性格の中で 2 番目に高い assortativity を示した。誠実な人はそうでない人の前ではイライラし、同時にそうでない人は誠実な人に対して面倒くささを感じてしまう。このように、Conscientiousness（誠実性）については、性格傾向の高低差によりお互いにストレスを感じてしまうことが原因であると考えられる。

Extraversion（外向性）の小分類の中では、Cheerfulness（明朗性）や Gregariousness（社交性）が高い Assortativity を示した。考察のため、Extraversion（外向性）の小分類について、それぞれの性格の性質を Table 7.1 に示す。表より、Cheerfulness（明朗性）、Gregariousness（社交性）は Extraversion（外向性）に含まれる性格の中でも他人に向けられた外向性であり、会話において表出しやすい性格である、また、同種の人との関わりが衝突や停滞を招く Assertiveness（自己主張）のような性格ではなく、むしろ会話のスムーズな進行に貢献するような性格であるから、同種の人の集まりが目立ったのだと思われる。

Table 7.1: Extraversion（外向性）の小分類の性質の考察

| 小分類 | 外向性の対象 | 説明 |
|---------------------------|-------------|------------------|
| Activity-level（活発度） | 経験 | 忙しさのレベル |
| Assertiveness（自己主張） | 人 | 他人の活動を指示したいという願望 |
| Cheerfulness（明朗性） | 人 | ポジティブな気持ちを表現する傾向 |
| Excitement-seeking（刺激希求性） | 経験 | 環境的刺激への欲求 |
| Outgoing（友好性） | 人付き合いの敷居の低さ | 他人に対する興味と友情 |
| Gregariousness（社交性） | 人 | 他の仲間に対する好意 |

7.4 コミュニティ内の性格の分布がコミュニティのふるまいに与える影響の分析

Excitement-seeking（刺激希求性）、Prone to worry（心配性）、Melancholy（悲観的）の高さがコミュニティの規模と正の相関を示すことがわかった。Excitement-seeking（刺激希求性）は、人との関わりによる刺激を、Prone to worry（心配性）とMelancholy（悲観的）は人との関わりによる心の安定を求めた結果の相関だと思われる。また、コミュニティの活性化には、Emotional Range（感情起伏）、Uncompromising（強硬さ）のふれはばの低さ、Dutifulness（忠実さ）、Self-efficacy（自己効力感）、Melancholy（悲観的）のふれはばの高さがコミュニティの活性化に関わることがわかった。前者は性格の違いが互いにかかるストレスが大きいため、コミュニティの成長を阻害すると推測する。後者は3つの性格は、自分のあり方、ふるまい方に影響を及ぼす性格である。このような特徴をもつ性格は、そのふれはばの高さが、コミュニティの多様性につながり、コミュニティ内の会話を活性化させるのだと予想する。

第8章 結論

8.1 本研究の結論

ソーシャルメディアでの人々の関わり合いは、今後もますます発展し多様化するであろう。そんな時流で、ソーシャルメディア上のチームワークにおけるパフォーマンスの最大化や、ソーシャルメディアの使用によりユーザが得られる効用の最大化のため、より詳しいソーシャルメディア上のコミュニティ形成のメカニズムの把握が重要である。本研究では、コミュニティ形成に影響を与える因子を「環境要因」、「その時点での個人の感情や思想」、「個人の性格」という3点に絞った。その中で解明されていない点の多い「個人の性格」がコミュニティの形成に与える影響を解明することが、本研究の目的であった。

関連研究では、Homophily の作用が最も強まるのが民族や人種といった「環境要因」であることや、「政治に対する感情や思想」から SNS のプロゴスフィアで Echo Chamber 現象が見られることなど、「個人の性格」以外の2要素がコミュニティ形成に与える影響は小さいことが示されている。従って、研究目的の達成には「個人の性格」によるネットワークのダイナミクスを他の要因から切り離して捉える必要があった。そこで、本研究では話題やユーザの興味を統一した共通の「トピックネットワーク」に所属するユーザを抽出した上で、性格の計算、分析指標との比較を行うという手法を提案した。そして日本語の Twitter のデータと IBM Watson の性格推定モデルを使用し、以下の分析を行った。

- ユーザの性格とその Twitter 上の行動の関係の分析
- ユーザの性格とメンションネットワーク上の位置づけの関係の分析
- Homophily な作用が起こる性格の存在の確認
- コミュニティ内の性格の分布がコミュニティのふるまいに与える影響の分析

性格と SNS 上の行動との関係性については、Openness（知的好奇心）が高いユーザはツイート頻度が高いこと、Agreeableness（協調性）が高いユーザは他人と会話する傾向が強いことを示した。これは、SNS 上のユーザの行動に関する関連研究で確認されている現象であり、これにより、日本語の Twitter データと IBM Watson の性格推定モデルを使用した提案手法が一定程度、妥当性を持つことが示される。

性格とネットワーク上の位置づけの関係に関しては、Conscientiousness（誠実性）、Self-discipline（自制力）の高い人ほどネットワークの中心からは遠くに位置する傾向にあることを示した。これは、そのような性格傾向の高い人ほど、ソーシャルメディア上の一つのコミュニティに依存する事態に陥りにくいからであると推測される。

それぞれの性格に Homophily な作用があるかどうかは、matching assortativity と呼ばれる指標を用いて分析した。その結果、Conscientiousness（誠実性）には Homophily な作用があることが示された。既存研究によるとチームビルディングにおいてメンバーの Conscientiousness（誠実性）はチームの生産性を向上させるが、チーム内の Conscientiousness（誠実性）のふれはばが大きいと、生産性は返って低下してしまうことが分かっている。ソーシャルネットワークにおいても同様に、Conscientiousness（誠実性）の違いによって生まれるコンフリクトが存在し、Homophily な作用を生み出しているという可能性を示した。また、Extraversion（外向性）の小分類は、Cheerfulness（明朗性）や Gregariousness（社交性）などの「人に向けられた外向性」と、Activity-level（活発度）や Excitement-seeking（刺激希求性）などの「経験に向けられた外向性」に分かれるが、そのうち「人に向けられた外向性」のみに強い Homophily な傾向が見られることが分かった。

最後に、コミュニティ内の性格分布とコミュニティのふるまいの関係に関しては、コミュニティの規模の大きさに関わる要素と、会話の活性化に関わる要素の二点について分析した。コミュニティの規模は Prone to worry（心配性）や Melancholy（悲観的）のように人との関わりによって心の安定を図る性格と正の相関を示すことが分かった。コミュニティの活性化に関しては、Emotionality（情動性）や Uncompromising（強硬さ）など、他の人との接し方に関わるような性格は、似た者同士である方が活性化が進み、Dutifulness（忠実さ）、Self-efficacy（自己効力感）、Melancholy（悲観的）、のような自らのあり方に関わる性格は、多様性に富んでいるコミュニティの方が活性化に寄与することが分かった。

以上のように、提案手法を用いることで、既存の心理学の発見と矛盾しない、性格とユーザの行動やネットワークダイナミクスとの関係が確認された。さらに、これまで知られなかったようなコミュニティのふるまいと性格の分布に関する相関関係や因果関係を明らかにできる可能性が提示された。

8.2 課題と今後の展望

本研究では、性格によってコミュニティのふるまいを予測する手法の提案に成功した。しかし、正確な相関関係や因果関係の抽出、実用的なサービスに結びつく知見を得るための、より一層の研究の方向性として、今後の展望を述べる。

8.2.1 より多様なトピックネットワークによる分析

本研究では、4つのネットワークについて、それぞれ10から20程度のコミュニティを取り出し、性格分布とふるまいの相関関係を調べた。より多種多様なトピックネットワークの抽出と分析により、信頼性の高い知見の獲得が可能となる。

8.2.2 スパースでないトピックネットワークより多く抽出する方法の考案

上で述べたような多様なネットワークを用いた分析を行うためには、より多くのスパースでないトピックネットワークを抽出する手法の考案が必要と思われる。本研究ではハッシュタグ検索により同一の興味を持つユーザを抽出したのちに、ネットワークのスパース性を確認したが、ソーシャルメディア上の密なネットワークを取り出してから、共通の興味をもつ部分の検知と抽出を行うことも、必要な手法と思われる。

8.2.3 異なる性格の組み合わせの評価

本研究では、一つの性格指標について、コミュニティ内の平均値と標準偏差のみを扱った分析を行った。しかし、現実のネットワークにおいて、人の関わりあいはいずれも複雑であり、たとえば激情的で自己主張の強い人物と共感度と忠実性の高い人物が接触すると、コミュニティの人間関係がどのように変化するかは、本提案手法では知ることはできない。説明変数を増やした分析手法の考案が求められる。

8.2.4 性格分析モデルの精度の向上

最新研究では、SNS データによる性格分析は、投稿テキストのみではなく、投稿画像や「いいね」の情報などあらゆる種類の特徴量を用いることでさらなる精度を実現している。本研究では、時間的コストやリソース的コストの制約により、外部 API を用いた分析に終始したが、性格分析の精度をあげることは、今後より複雑な相関関係を明らかにする上で不可欠だろう。

謝辞

本研究を行うにあたり、非常に多くの方々にお世話になりました。この場を借りて感謝の意を述べさせていただきます。

指導教員の坂田一郎教授は、お忙しい立場の中、貴重な時間を割いていただき、自分のやりたいことを実現するために目指すべき研究の方向性を指し示していただきました。心よりお礼申し上げます。

共同研究室の森純一郎教授には、本研究を進める上で不勉強な点が多かった私に足りない知識や手法が何か、優しく指摘いただきました。ありがとうございました。

株式会社ホットリンク兼本学研究員の榊剛史氏、本学研究員の大知正直氏、浅谷公威氏には、日頃から研究の技術的なアドバイスや、使用できる手法やツール、最新の関連研究に関する情報などを共有していただきました。特に大知さんには、手法に関するアドバイスから、論文の構成、そして執筆の最後の最後まで助言や叱咤激励をいただきました。深く感謝いたします。

研究室の先輩や研究室秘書の方々には、研究を進める上での様々なサポートをしていただきました。特に、同じ部屋で作業をしていた修士2年生のみなさんは、ご自身の修士論文で忙しい中、声をかけてくださったり、研究面でも、生活面でも、さまざまなアドバイスをくださったり、大変お世話になりました。

一緒に研究を進めてきた研究室の同級生である中元雪絵さん、山本瑛伝さんとは、時には息抜きもしながら、切磋琢磨しあいました。半年間の研究を頑張りきれたのは、同級生の仲間のおかげです、本当にありがとうございました。

最後に、楽しい時も、大変な時も変わらず私に接して、荷を軽くしてくれる家族、友人にも感謝いたします。

改めまして、お世話になったみなさまへ感謝申し上げます。そして、研究室のさらなる発展を祈って謝辞とさせていただきます。ありがとうございました。

参考文献

- [1] 北村英哉, 那須川哲哉, and 上條浩一. 特別鼎談 文章を解析し, 書いた人の性格を推定: 「ibm watson personality insights」の可能性 (特集イノベーションを創出するソーシャル・コラボレーションとテクノロジー). *Provision*, (89):26–31, 2016.
- [2] Eric Gilbert, Tony Bergstrom, and Karrie Karahalios. Blogs are echo chambers: Blogs are echo chambers. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*, pages 1–10. IEEE, 2009.
- [3] Lewis R Goldberg. An alternative” description of personality”: the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216, 1990.
- [4] Paul T Costa Jr, Robert R McCrae, and David A Dye. Facet scales for agreeableness and conscientiousness: A revision of the neo personality inventory. *Personality and individual Differences*, 12(9):887–898, 1991.
- [5] Oliver P John and Sanjay Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2 (1999):102–138, 1999.
- [6] Robert R McCrae and Paul T Costa Jr. Personality trait structure as a human universal. *American psychologist*, 52(5):509, 1997.
- [7] David P Schmitt, Jüri Allik, Robert R McCrae, and Verónica Benet-Martínez. The geographic distribution of big five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of cross-cultural psychology*, 38(2):173–212, 2007.
- [8] Paul T Costa Jr and Thomas A Widiger. *Personality disorders and the five-factor model of personality*. Washington, DC, US: American Psychological Association, 1994.
- [9] Murray R Barrick and Michael K Mount. The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26, 1991.
- [10] Murray R Barrick, Greg L Stewart, Mitchell J Neubert, and Michael K Mount. Relating member ability and personality to work-team processes and team effectiveness. *Journal of applied psychology*, 83(3):377, 1998.

- [11] James E Driskell, Gerald F Goodwin, Eduardo Salas, and Patrick Gavan O'shea. What makes a good team player? personality and team effectiveness. *Group Dynamics: Theory, Research, and Practice*, 10(4):249, 2006.
- [12] Patrizia Steca, Dario Baretta, Andrea Greco, Marco D'Addario, and Dario Monzani. Associations between personality, sports participation and athletic success. a comparison of big five in sporting and non-sporting adults. *Personality and Individual Differences*, 121:176–183, 2018.
- [13] Tracii Ryan and Sophia Xenos. Who uses facebook? an investigation into the relationship between the big five, shyness, narcissism, loneliness, and facebook usage. *Computers in human behavior*, 27(5):1658–1664, 2011.
- [14] Jason L Skues, Ben Williams, and Lisa Wise. The effects of personality traits, self-esteem, loneliness, and narcissism on facebook use among university students. *Computers in Human Behavior*, 28(6):2414–2419, 2012.
- [15] Gwendolyn Seidman. Self-presentation and belonging on facebook: How personality influences social media use and motivations. *Personality and Individual Differences*, 54(3):402–407, 2013.
- [16] Teresa Correa, Amber Willard Hinsley, and Homero Gil De Zuniga. Who interacts on the web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, 26(2):247–253, 2010.
- [17] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.
- [18] Lisa A Fast and David C Funder. Personality as manifest in word use: correlations with self-report, acquaintance report, and behavior. *Journal of personality and social psychology*, 94(2):334, 2008.
- [19] Jacob B Hirsh and Jordan B Peterson. Personality and language use in self-narratives. *Journal of research in personality*, 43(3):524–527, 2009.
- [20] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. Predicting personality from twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 149–156. IEEE, 2011.
- [21] P.-H. Arnoux, A. Xu, N. Boyette, J. Mahmud, R. Akkiraju, and V. Sinha. 25 Tweets to Know You: A New Model to Predict Personality with Social Media. *ArXiv e-prints*, April 2017.

- [22] Danny Azucar, Davide Marengo, and Michele Settanni. Predicting the big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, 124:150–159, 2018.
- [23] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.
- [24] 那須川哲哉, 上條浩一, 山本眞大, and 北村英哉. 日本語における筆者の性格推定のための言語的特徴の調査. 言語処理学会第 22 回年次大会予稿集, pages 1181–1184, 2016.
- [25] Mark EJ Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, 2003.