

大規模ソーシャルネットワークにおける 時間的構造変化の解析手法の研究

渡部 優[†] 鈴木 豊太郎[‡]

[†] 東京工業大学情報理工学研究科, JST CREST 〒152-8550 東京都目黒区大岡山 2-12-1

[‡] IBM Research, University College Dublin, JST CREST

E-mail: [†] watanabe.m.ay@m.titech.ac.jp, [‡] toyo@jp.ibm.com

あらまし 近年, Twitter や Facebook などの SNS が急速に普及してきたが, このような多様化する SNS 上のネットワークの性質を深く理解するためには, 成長を続けるネットワークに対し時間的な変化を解析する必要がある. しかし, ネットワークの巨大化に伴うデータ収集や解析コストの増大が大きな課題となっている. そこで本研究では, Twitter 特有の構造変化のルールを抽出し, ユーザの行動パターンを定義することで, ネットワークの時間的な変化を解析するためのモデルを提案した. さらに, 2009 年の Twitter の実ネットワークとの比較実験を行うことで, このモデルが実際のネットワークに近い振る舞いを示すことを確認した.

キーワード Twitter, シミュレーション, 大規模グラフ, ソーシャルネットワーク

1. はじめに

1.1. 研究背景・目的

近年, Twitter, Facebook, LinkedIn といったソーシャルネットワークサービス (SNS) が急速に普及し, 私たちの生活に深く浸透してきた. これらのサービスは人と人をつなげるコミュニケーションツールとして大きな役割を果たすようになり, それと同時にそのコミュニティのあり方も多様化してきた. 例えば, Facebook では, 実名を使ったアカウント登録が行われ, 実生活における友人・知人とのネットワークが作られる. Twitter では, ハンドルネームを使ったアカウント登録が行われ, ユーザの興味や関心のある人とのネットワークが作られる.

こういったネットワークの時間的な変化を解析することは, 多様なコミュニティそれぞれの性質や構造の理解へとつながり, 研究として非常に興味深い領域である. しかしながら, 最近ではユーザの増加に伴うネットワークの巨大化により, 連続的なデータの取得や解析に大きなコストがかかり, 手軽に扱うことができなくなっている. そこで本研究では, SNS の 1 つとして有名な Twitter ネットワークを対象に, 以下の手順でネットワークの時間的な構造変化を解析する手法を提案する.

まず, スナップショットとして事前に収集した約 4.7 億ユーザからなる実ネットワークの予備解析を行うことで, Twitter におけるネットワーク特有のルールを抽出する. そのルールを考慮した上で, ネットワーク成長モデルの観点からユーザのフォローやフォロー解除のパターンを定義し, モデルとして提案する. そして, そのモデルをエージェントベースシミュレーション上

に実装することで, Twitter ネットワークの成長を再現する. このシミュレーションにより生成されたネットワークと実ネットワークとの比較を行うことで, 提案したモデルが実ネットワークに近い振る舞いを示すことを確認する.

1.2. 本論の構成

1 章では研究の背景・目的と, 本研究の概要を示した. 以降, 2 章では関連する研究を紹介し, 本研究との差分について言及する. 3 章では本研究で対象とする Twitter 特有の性質について紹介し, モデル化のパターンを提案する. 続く 4 章では, エージェントベースのシミュレーションにおけるネットワーク成長モデルの設計と最適化を含めた実装について紹介する. 5 章では, シミュレーションによる再現ネットワークの評価を行い提案手法について議論する. 最後に 6 章で本研究のまとめを行う.

2. 関連研究

SNS の大規模なネットワーク解析の研究として, 2009 年の Twitter ネットワークを対象にした研究[1]や 2011 年の Facebook ネットワークを対象にした研究[2]がある. [1]では, ユーザ数 4000 万のネットワークのスケールフリー性や隔たり (平均経路長) といった性質を明らかにしている. しかしながら当時のネットワークは, 2012 年現在と比較すると, 1/10 程度の規模でしかなく, ネットワークの巨大化が進む現在においては小さなものとなっている. また, [2]では, 2011 年の Facebook ネットワークを用い, スケールフリー性や隔たりの他, 地域性を考慮したサブグラフの解析を行い, Facebook 特有の性質を明らかにした. 規模としては,

7.2 億ユーザと 690 億エッジからなる巨大なネットワークを扱っている．しかしながら，時間的変化の観点では，1 年おきのスナップショットのネットワーク解析にとどまり，細かい粒度での時間的なネットワークの構造変化の解析には至っていない．

人工的にネットワークを生成・再現する既存の研究として，BA モデル[5]や CNN モデル[6]が提案されている．BA モデルは一般的なネットワークがスケールフリー性を持つという観点から提案されたモデルで，ノードの度数に比例した確率（Preferential Attachment: 優先的選択）でリンクが張られる．また，CNN モデルは，2 ホップでたどり着ける頂点間を結びつける（間接リンクを直接化する）ことでネットワークの成長を再現するモデルである．しかしながら，これらのモデルでは，近年の人工的制限のある SNS ネットワークにはそのままでは適用できないという問題がある．

3. Twitter ネットワークの予備解析

Twitter ネットワークの成長モデルを提案するにあたり，まず Twitter 特有のネットワークの性質を明らかにする．本章では，まず 3.1 節で Twitter サービスの概要とそこで構築されるネットワークについて簡単に説明する．3.2 節では Twitter ネットワークの予備解析結果を示し，ネットワーク成長モデルに組み込むべきルールを抽出する．

3.1. Twitter サービス概要

Twitter とは，140 文字以内のショートメッセージ（“ツイート”）を投稿できるマイクロブログサービスである．このツイートは投稿者のタイムラインに保存され，ユーザはこの投稿者のタイムラインをいつでも閲覧することができる．また，タレントやアーティストなどの気に入った投稿者がいれば，その投稿者を“フォロー”することによって，自身のタイムラインにそのツイートを表示することができるようになる．Twitter ではこのフォロー機能によって，図 1 のように，たくさんのユーザをフォロー関係で結んだ巨大なネットワークが作られる．なお，本稿では便宜上，フォロー関係の定義として，A が B をフォローしているとき，A を B の“フォロワー”，B を A の“フレンド”と呼ぶこととする．

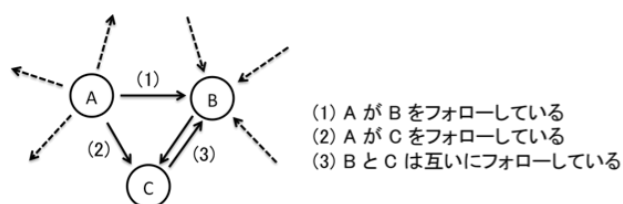


図 1. フォロー関係により構築される Twitter ネットワーク

3.2. Twitter ネットワークの特性

本研究では，Twitter 特有のネットワークの性質を明らかにするため，スナップショットとなる Twitter ネットワークを収集し，その予備解析を行った．ここで解析に使用したデータは，4.7 億ユーザからなる 2012 年 10 月時点の大規模スナップショットと，約 3,000 人のユーザを対象とした 2013 年 7 月から 9 月の期間の連続スナップショットである．前者の大規模スナップショットは，4.7 億人分のユーザプロフィールを含んでおり，ユーザのアカウント作成日や使用言語情報などを同時に調べることができる．なお，大規模スナップショットのネットワーク情報は GZIP 圧縮で 231GB あり，非常に大きなものとなっている．この解析では，東京工業大学のスパコン TSUBAME2.5 上で大規模データの分散処理基盤である Apache Hadoop[3]を用いた．

3.2.1. ネットワーク成長とユーザの増加

まず我々は，Twitter のネットワークの成長を再現するにあたり，ネットワークが時間的にどのように成長を成長してきたのかを観察を行った．図 2 はユーザのアカウント作成日をもとに，2006 年 3 月から 2012 年 10 月までの期間で，月単位での新規ユーザ数の変化を示したものである．この図では，2009 年から急激に新規ユーザ数が増加したことがわかる．しかしながらこの増加分は単調な増加とはなっておらず，前月から減少している時期もある．そこで，本研究では新規ユーザの増加をこのネットワークの成長速度として捉え，Twitter の成長モデルを考える上でのルールとして抽出する．これにより，実際の Twitter ネットワークの成長に近い振る舞いを，モデル上で再現する．

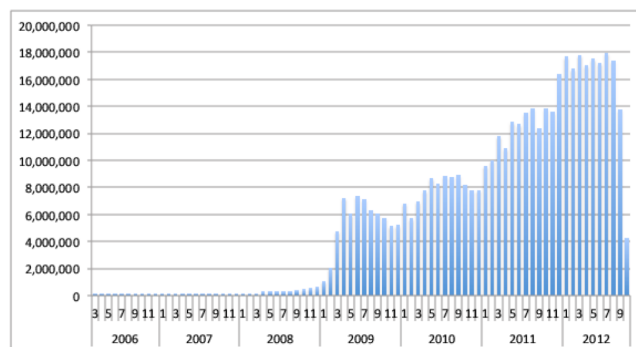


図 2. 月単位での新規ユーザ数の変化

3.2.2. スケールフリー性とフォロー制限

SNS における一般的なネットワークの性質の 1 つにスケールフリー性がある．スケールフリー性とは，グラフの度数の確率分布がべき乗則に従う（ネットワークの度数分布が両対数グラフで直線になる）性質である．直感的には，ごく一部のユーザが他のたくさんユーザにフォローされ大きな度数を持つ一方で，ほとん

どのユーザはフォロワーの数が少なく次数が小さいというものである。我々の先行研究[8]ではスケールフリー性について議論をしているが、ここではネットワーク成長のモデル化の観点から、Twitterにおけるフォロー制限について新たに言及する。

以下の図3と図4はTwitterにおける次数分布を示している。図3のフォロワー数の次数分布は、両対数グラフ上で直線的になっており、スケールフリー性を持っていることが確認できる。一方、図4のフレンド数の次数分布では、フレンド数が2,000のときに大きな値を示していることがわかる。これはユーザフォロワー数に対し一定の割合を超えて新規にフォローすることができないという、Twitterのフォロー制限が要因¹である。したがって、Twitterのネットワーク成長を考える上で、この次数分布の様子をTwitter特有のルールとして抽出する。

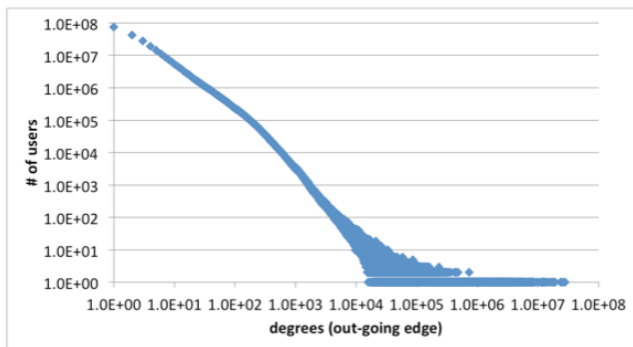


図3. フォロワー数の次数分布
(両対数グラフ)

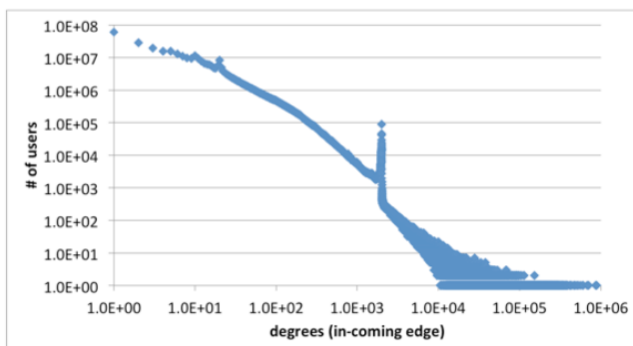


図4. フレンド数の次数分布
(両対数グラフ)

3.2.3. 相互フォロー率

Twitterにおけるユーザの行動の1つに”フォロー返し”というものがある。フォロー返しとは、あるユーザがTwitterで誰かにフォローされたときに、フォロー

を返すことで、相互にフォローしている状態になることである。我々はこの相互フォローの割合がTwitterでどのように分布しているのかを調べるため、Twitterのユーザをフォロワー数100~1,000人(グループA)、1,000~10,000人(グループB)、10,000人以上(グループC)の3つのグループに分け、そのグループ毎に相互フォロー率の分布を算出した。なお、ユーザ*i*の相互フォロー率 $P(i)$ は、 E_{fo} と E_{fr} をそれぞれフォロワーとフレンドの集合とし、以下の式で得る。

$$P(i) = \frac{|E_{fo} \cap E_{fr}|}{|E_{fo}|}$$

図5、図6、図7はA,B,Cそれぞれのグループにおける相互フォロー率の分布を示している。グループAの分布(図5)では、相互フォロー率の最頻値は、0%を除くと、80%となっており、ユーザ同士が相互にフォローしている割合が高いと言える。しかしながら、グループBとグループCでは、中間的な相互フォロー率を持つユーザ数が減り、相互フォロー率が極端に低いか高いかのいずれかに偏る傾向が見られる。この相互フォロー率の偏りは、Twitterネットワークの中にフォロー返しを積極的に行うユーザがたしかに存在していることを示している。したがって、本研究では、このフォロー返しの割合をTwitter特有のルールとして抽出する。

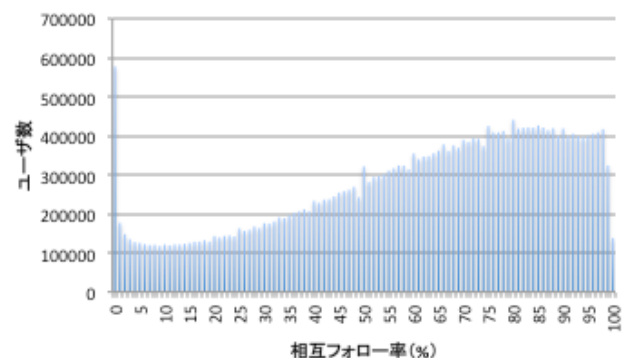


図5. 相互フォロー率分布
(グループA: フォロワー数100~1,000)

¹フォロー制限については、Twitterサポートページにて確認ができる。
<https://support.twitter.com/articles/66885-i-can-t-follow-p-eople-follow-limits>

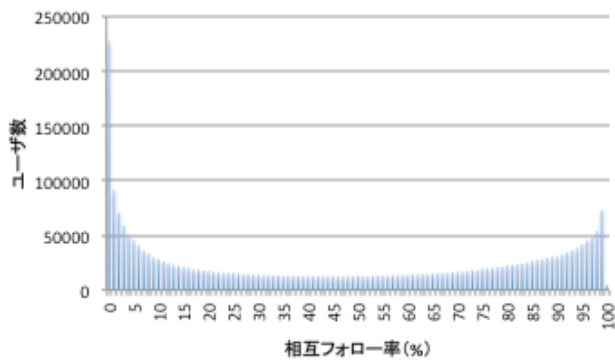


図 6. 相互フォロー率分布
(グループ B: フォロワー数 1,000~10,000)

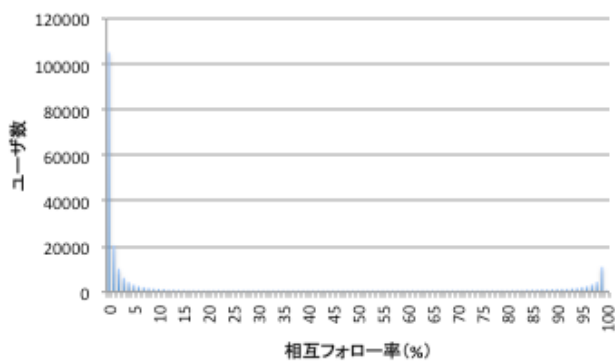


図 7. 相互フォロー率分布
(グループ C; フォロワー数 10,000~)

3.2.4. 言語間の隔たり

2013 年現在, Twitter は世界各国で使われるサービスとなっている. しかし, 異なる言語を使用するユーザ同士が直接たくさんのフォロー関係を持っているとは考えにくく, 異なる言語を使用するユーザの間には大きな隔たりがあると推察できる. そこで, Twitter のユーザを使用言語で分類し, 各言語へのフォローの割合を算出した.

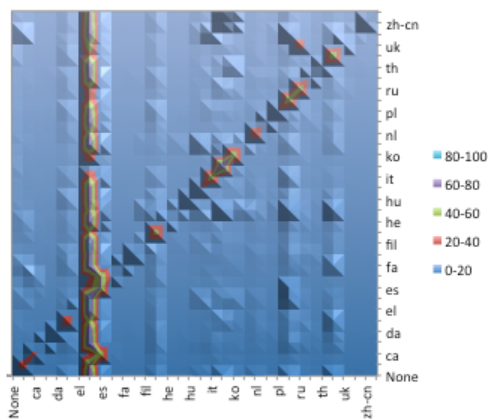


図 8. 言語間のフォロー率の分布

図 8 は言語間のフォロー率の分布を示しており, 縦軸にフォロー元の使用言語, 横軸にフォロー先の使用言語, 等高線にフォロー先言語へのフォローの割合を示している. 横軸の”en”における英語使用ユーザへのフォロー割合をみると, 英語使用ユーザの母数が多数であることもあり, 英語使用ユーザへのフォローは大半のフォロー元言語で高い割合 (70~90%) を示している. しかしながら, 対角線上にあたる, ある言語から同言語へのフォロー割合も高い割合を示しており, これはある言語におけるネットワークが他の言語に対して閉じていることを示している. 例えば, 中央付近の日本語ユーザから日本語ユーザへのフォローの割合は 90%にもなる一方, 日本語ユーザから英語ユーザへのフォローの割合は 8%しかなく, 日本語以外の言語に対して閉じていることが言える. 他言語に対して閉じている傾向の言語は日本語の他に, スペイン語の同言語へのフォロー率 64%や, ポルトガル語の 58%などが該当する. したがって, 本研究では, この言語間のフォローの割合を Twitter 特有のルールとして抽出する.

3.2.5. フォロー解除の頻度

本研究では, 小規模な連続スナップショットの時間的な解析を行うことで, フォローとフォロー解除の変動を調査した. 図 9 は, フォロワー数 10,000 以上のユーザをスナップショットデータから抽出し, 1 ヶ月間に何回の新規フォローとフォロー解除が行われたのかを示している. ユーザによって, 変動が大きいいため, 縦軸は対数とし, 横軸はアカウント作成日を考慮して, ユーザをソートしている. この解析により, フォロー解除が新規フォローに匹敵するほど多く行われていることが判明した. このことは, Twitter のネットワーク成長を特徴付ける重要な要素となりえるため, 本家研究では, このフォロー解除を考慮したネットワーク成長モデルを提案する.

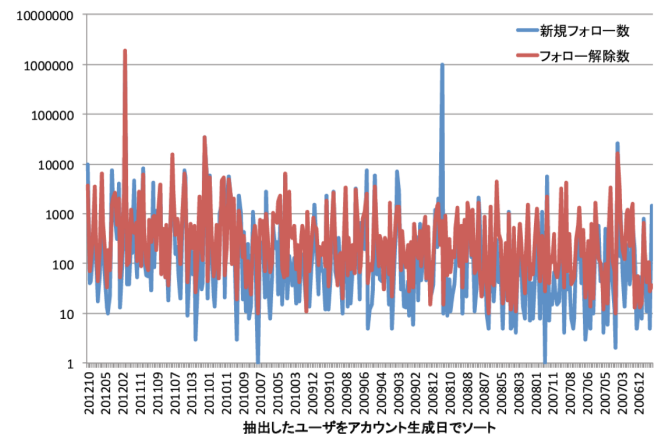


図 9. 新規フォローとフォロー解除数

4. ネットワーク成長のモデル化

3章ではTwitterネットワークの予備解析を通し、ネットワーク成長モデルに考慮すべきルールを抽出した。本章では、それらのルールを考慮した上で、ユーザの行動パターンを定義することで、ネットワーク成長モデルを提案する。まず、本研究で入れるネットワーク成長モデルの要件は以下の3つである。

1. 新規ノードの追加
2. 新規エッジの追加
3. 既存エッジの削除

4.1. 新規ノードの追加

Twitterにおいて、新規ノードの追加は新規ユーザのアカウント登録と等価である。したがって、実際のネットワークに近い振る舞いを再現するには、3.2.1節で示したように、Twitterの新規ユーザ数の変化を時間経過で再現できればよい。本研究では4.7億ユーザのアカウント作成日情報を用いることで、これを再現する。

4.2. 新規エッジの追加

4.2.1. フォロー頻度

新規エッジの追加はユーザが他のユーザを新しくフォローすることと等価である。ユーザの行動をベースにしたモデルを考える上で、フォローの頻度をユーザごとに与える必要がある。そこで、本研究では、2012年のフレンド数を利用し、単位時間あたりのフォロー数を算出することで、各ユーザの一定のフォロー頻度を与えた。

4.2.2. フォローのパターン

さらに、本研究ではフォローのパターンを5つのパターンに分類し、モデルとして定義した。

・ 知り合いをフォロー

このユーザの動作は、Twitter上にすでに友人や知人のアカウントが存在し、その友人や知人をフォローする行為に当てはまる。3.2.3節で示したように、多くの一般的なユーザは相互フォローの割合が高く、その相互フォローの関係にあるユーザ同士は、直接の知人である可能性も高い。そこで、本研究では、2012年のネットワークデータか一部の相互フォロー関係の一部を抽出し、入力データとして与えることで、これを再現する。

・ 有名人をフォロー

Twitterのサービス上には、一般ユーザ以外にも、TVやメディアで有名なアーティストの他、Newsサイトや企業などの幅広い種類のアカウントが存在している。そこで、このような著名なTwitterアカウントへのフォローを有名人へのフォローとして定義する。

この有名人をフォローする行動に相当する既存手法

に、適応度を考慮した優先的選択[cite]がある。この手法では、各ユーザに適応度（フォローされやすさ）のパラメータのパラメータを事前に付与することで、有名人へのリンクが高くなる事象を再現する。本研究のモデルでも、有名人をフォローするパターンにおいて、この手法を採用する。

・ リツイートの発信元をフォロー

Twitterでよく行われる行為に、他人のツイートを自分のフォロワーに共有するリツイートがある。リツイートを受け取ったユーザは、直接自身がフォローしていないユーザのツイートを受け取ることになるため、そのリツイートの発信元のユーザを新しくフォローする可能性がある。そこで、本研究では、この行動パターンを定義し、モデルに組み込む。

・ フォローに対するフォロー返し

3.2.3節で示したように、フォロワー数の少ない一般的なユーザは相互フォローの割合がかなり高いことを説明した。またフォロワー数のユーザについても、そのごく一部は相互フォロー率がほぼ100%に近い値となっており、フォロー返しはネットワークの性質を特徴づけるものとなっていることを示した。そこで、本研究ではフォロー返しをユーザの行動として定義し、ネットワーク成長モデルに組み込む。なお、相互フォロー率は、2012年の実データを元に算出して与えることで、実際のネットワークに近い振る舞いを再現する。

・ ランダムフォロー

このパターンでは、前述までの4パターンに該当しない、ユーザ行動のランダム性を考慮する。例えば、趣味に関するワードで検索を行ったときに、検索結果に表示されたユーザをフォローする場合や、Twitter外部のブログやまとめサイトなどを通して新しいユーザを発見してフォローする場合が該当する。

4.3. 既存エッジの削除

3.2.5節で示したように、フォロワー数が多ければ多いほど、フォローを解除する割合が高いことから、本研究のモデルでは、ユーザ毎にフォロー頻度に比例したフォローの解除頻度を与える。また、フォローの解除の対象ユーザはフレンドの中からランダムで選出することで、これを行うこととする。

5. ネットワーク成長モデルの設計と実装

5.1. シミュレーション基盤 XAXIS

ここではネットワークの成長モデルを実装する基盤となるシミュレーション基盤 XAXIS について簡単に説明する。XAXISはPGAS言語X10[7]により実装されたエージェントベースのシミュレーション基盤で、単位時間（以下、ステップと呼ぶ）毎にエージェント間でメッセージ交換を行うことによりシミュレーショ

ンを実行する．また，XAXIS には Java のインターフェースが用意されているため，開発者は Java を用いることで，複数の計算機を用いた大規模な並列分散シミュレーションを行うことができる．基本的な実装は，エージェントの `run` メソッド内でユーザの行動を定義し，`receiveMessage` メソッド内でメッセージ送受信処理を実装するだけでよい．

XAXIS がサポートするエージェントには 2 種類あり，1 つは `Driver` エージェント，1 つは `Citizen` エージェントである．いずれのエージェントも自律的に動作し，メッセージを交換することが可能であるが，本研究のネットワーク成長モデルの実装にあたり，それぞれに以下のような役割を与えた．

・ `Driver` エージェント

Twitter ユーザ全体の属性情報の管理を行い，シミュレーションの入出力データを制御する．また，有名人を選出し，サジェストする機能をもつ．

・ `Citizen` エージェント

Twitter における 1 ユーザとして，自律的にフォローやフォロー解除を行い，他のユーザとのフォロー関係をデータ構造として持つ．

5.2. 基本設計

5.2.1. ネットワークの成長

ネットワークの成長は，基本的にノードの追加により再現される．これは Twitter ユーザの管理を行う `Driver` エージェントにユーザのアクティブ状態を示すビットマップを持たせる．シミュレーションのステップごとに，このビットマップを更新し，新規ユーザをアクティブ状態にすることで，ネットワークの成長を再現する．

5.2.2. フォロー関係

フォロー関係は，すべての `Citizen` エージェントにフォロワーとフレンドのリストを持たせることで表現する．新規のフォローや，フォロー解除を行う際は，このリストを更新する．

5.2.3. 新規フォローとフォロー解除

シミュレーションにおける新規フォローやフォロー解除の行動は，エージェント間でメッセージをやりとりすることにより行われる．メッセージの種類は，他のユーザをフォローする“フォローメッセージ”，フォロー解除を行う“アンフォローメッセージ”，他のユーザを推薦する“サジェストメッセージ”，その推薦を問い合わせる“リクエストメッセージ”がある．

5.3. ユーザ行動の流れ

基本的に，ステップ内でエージェントは図 10 のように，新規フォローやフォロー解除を行う．まず，ユーザが新規ユーザであれば，入力データから“知り合いをフォロー”し，初期エッジを追加する．続いて，単位時間あたりのフォロー数を将来のフォロー数から計算し，新規フォロー数を決定する．この新規フォローは確率的に，“有名人をフォロー”，“リツイートの発信元をフォロー”，“ランダムフォロー”のいずれかのパターンを選択される．最後にフォローの解除数を計算し，フレンドの中からランダムに“フォロー解除”を行う．なお，“フォローに対するフォロー返し”はフォローメッセージを受けとった時に，受動的に動作するため，このフロー図には書かれていない．

シミュレーション時に与えるパラメータは“有名人をフォロー”する確率 α ，“リツイートの発信元をフォロー”する確率 β ，“ランダムフォロー”の確率 $1 - \alpha - \beta$ とし，6 章の実験では，この確率を変化させることで，Twitter のネットワーク成長に適切なパラメータを調査する．

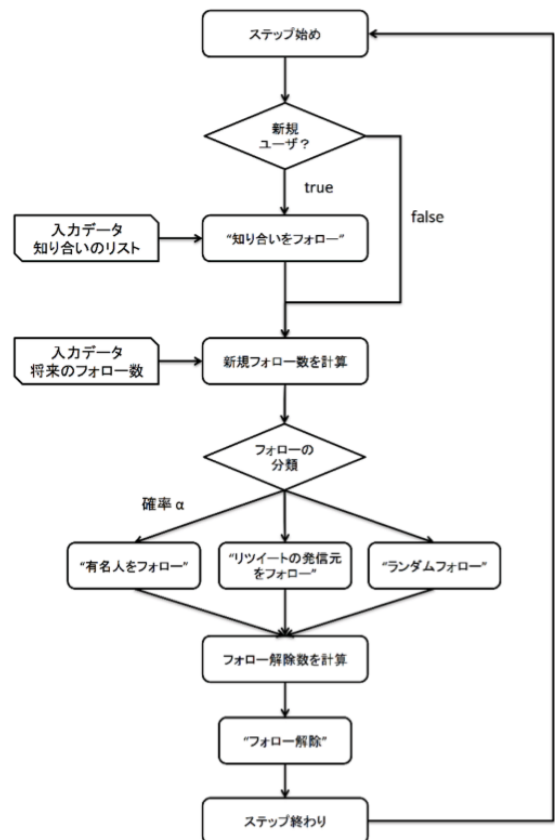


図 10. `Citizen` エージェントの動作の流れ

6. ネットワーク成長シミュレーションの実験と評価

6.1. 入力データ

本シミュレーションにおける入力データと使用用途は以下の通りである。

- アカウント作成日情報

アカウント作成日順にソートされたユーザ ID のリスト。各ステップで新規ユーザを作成し、シミュレーションステップに反映させるために用いる。

- 初期ネットワーク

2012 年のネットワークから、2006 年 4 月に存在するアカウントのみを抽出し、初期ネットワークとして与える。ノード数は 80、エッジ数は 526 になる。

- ユーザ属性情報

各ユーザの言語や相互フォロー率を含んだリスト。相互フォロー率は実ネットワークから事前に計算しておく。

- 言語間のフォロー率情報

言語間のフォロー率を記述したリスト。各ユーザは 3.2.4 項で示した言語間の隔たりを考慮し、多言語へのフォローを行う。

6.2. 実験環境

我々は XAXIS 上のネットワーク成長シミュレーションを東京工業大学のスパコン Tsubame2.5 を用いて実験を行った。使用した計算機は最大 20 ノードで、各ノード 54GB メモリ、24 コアで構成されている。

また、ソフトウェアは IBM Java 1.6.0 と X10 2.2.1 を使用し、実験を行った。

6.3. 小規模実験

実験方法

まず、シミュレーションに与えるパラメータの影響と、適切なパラメータを選択する必要がある。そこでステップの単位時間を 1 日とし、800 ステップまで小規模なシミュレーションを行った。このシミュレーションに与えるパターンは表 1 の通りである。ランダムフォローはネットワークに与える影響が大きいため、小さい値を割り当てている。

試行	有名人	リツイート	ランダム
1	0.5	0.5	0
2	0.45	0.45	0.1
3	0.495	0.495	0.01

表 1. パラメータの設定

実験結果

このシミュレーションの結果は表 2 の通りである。

この結果では、ノード数、エッジ数、直径や平均経路長といった、基本的なネットワークの性質に大きな違いは見られなかった。しかしながら、次数がゼロとなるユーザ数を比較してみると、試行 2 の結果が他の試行よりも小さな値となっている。これは少し高い確率に設定されたランダムフォローが、ネットワークに大きく影響を与えたことを示している。そこで、本研究では、ユーザのランダム性を考慮しながらも、ネットワークに与える影響が小さい試行 3 のパラメータを採用し、2009 年のネットワークとの比較実験を行う。

	試行 1	試行 2	試行 3
ノード数	2,098,094	2,098,094	2,098,094
エッジ数	42,225,070	43,239,520	42,316,517
平均次数	20.1	20.6	20.1
直径	12	11	11
平均経路長	3.67	3.65	3.66
次数 0 のユーザ数	987,769	300,072	823,442

表 2. 小規模シミュレーション結果

6.4. 大規模実験

実験方法

大規模実験では、2009 年と同等の規模となる 1200 ステップまでネットワーク成長シミュレーションを行い、実ネットワークとの比較を行うことで、本研究で提案したモデルの評価を行う。なお、この 2009 年の実ネットワークは、[1]の研究で収集された 4100 万ユーザからなる実際の Twitter ネットワークである。実験で与えるパラメータは、小規模実験で検証した表 3 のパラメータである

有名人	リツイート	ランダム
0.5	0.5	0

表 3. パラメータの設定

実験結果

図 11 は 1200 ステップのシミュレーションにより生成されたネットワークの次数分布を示している。この結果から、フォロワーの次数分布は両対数軸上で直線的になっており、スケールフリー性を満たしていることが確認できる。また同時に、フレンド数 2000 のユーザ数は大きな値をみだしており、Twitter の人工的なフォロー制限も再現できていることが確認できる。

また、表 4 に、2009 年の実ネットワークとの性質の比較を示した。シミュレーション結果の平均経路長は、実ネットワークのそれと近い値を示しており、ネットワークのsmall-world性を満たしていることが確認できる。しかしながら、エッジの総数（平均次数）の点では、シミュレーション結果は小さい値となっており、実ネットワークに比べ疎なネットワークとなっ

ていることを確認した．これは，フォロー先を選択する際に，すでにフォロー済みのユーザを選択してしまった場合や，言語間のフォロー率からフォローを決定しなかった場合に，新規フォローがされず，結果としてリンク数が減ってしまったことが要因と考えられる．

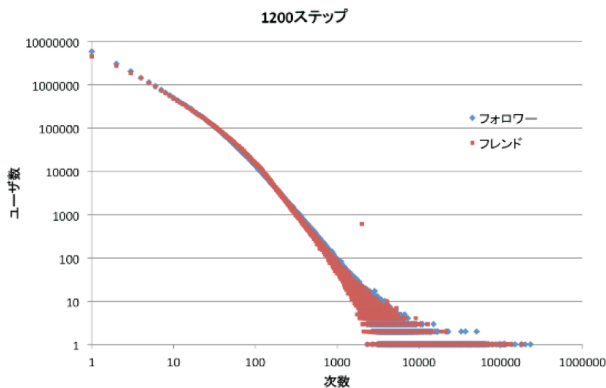


図 11. 1200 ステップ時の次数分布

	1200 ステップ シミュレーション	2009 年 実ネットワーク
ノード数	47,00,274	41,652,230
エッジ数	698,933,251	1,468,365,181
平均次数	13.94	32.25
直径	13	28
平均経路長	4.23	4.49

表 4. シミュレーション結果と
実ネットワークの比較

7. まとめと今後の展望

7.1. まとめ

本研究では，ネットワークの時間的構造変化をシミュレーションで再現するため，まずスナップショットとして事前に収集した約 4.7 億ユーザからなる実ネットワークの解析を行い，Twitter におけるネットワーク特有の性質を検証した．そこで得られた検証結果を Twitter における構造変化のルールとして抽出し，ネットワーク成長の観点からユーザのフォローやフォロー解除のパターンのモデル化を行った．そして，モデル化したパターンをエージェントベースシミュレーションに落としこみ，ネットワーク構造の時間的な変化を再現した．これを 2009 年の実ネットワークとの比較を行うことで，提案したモデルが実ネットワークに近い振る舞いを示すことを確認した．

7.2. 今後の展望

本研究のモデルをさらに精度の高いものにするために，Twitter ネットワークのモデルに検討すべき項目を示す．現状では各ユーザに対し，固定のフォロー頻度（単位時間あたりのフォロー数）を割り当ててい

るが，ユーザの成熟度などを考慮した動的なフォロー頻度を導入することで，より高度なモデル化を行う必要がある．また，ユーザの属性情報として，言語情報の他にも，BOT のような機械的ユーザや，News アカウントといった特殊なユーザの存在を考慮し，精度を高めることが求められる．

実験においては，30 ノードを用いた分散環境において，パフォーマンスの問題により規模として 2009 年までのシミュレーションにとどまった．今後は実装の最適化などを行い，より大規模なシミュレーションを行える環境を整えていく必要がある．

謝辞

本研究の一部は JST CREST「ポストペタスケールシステムにおける超大規模グラフ最適化基盤」の援助による．また，本研究を進めるにあたり，ご指導を頂いた東京工業大学計算工学専攻徳田雄洋教授と実験用データを収集・提供して下さいた東京工業大学上野晃司氏に感謝いたします．

参 考 文 献

- [1] Haewoon Kwak, et al. "What is Twitter, a social network or a news media?", *Proceedings of the 19th international conference on World*
- [2] Lars Backstrom, et al. Four degrees of separation. In *ACM Web Science 2012: Conference Proceedings*, pages 45–54. ACM Press, 2012.
- [3] Apache Hadoop : <http://hadoop.apache.org/>
- [4] Leskovec, et al. "Graphs over time: densification laws, shrinking diameters and possible explanations." *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. ACM, 2005.
- [5] Albert, Barabási. "Statistical mechanics of complex networks." *Reviews of modern physics* 74.1 (2002): 47.
- [6] Vázquez, Alexei. "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations." *Physical Review E* 67.5 (2003): 056104.
- [7] Saraswat, Vijay A., et al. "X10: concurrent programming for modern architectures". In *Proceedings of the 12th ACM SIGPLAN symposium on Principles and practice of parallel programming (PPoPP '07)*
- [8] Masaru Watanabe, et al. "How Social Network is Evolving ? A Preliminary Study on Billion-scale Twitter Network" *The 2nd International Workshop on Large Scale Network Analysis (LSNA 2013)*