

SNS に現れるダンバー数とその起源 -データ分析とモデル化-

荒木 直人^{1,a)} 日永田 泰啓² 只木 進一¹

受付日 2016年3月4日, 採録日 2016年8月1日

概要: ダンバー数とは親密な人間関係を維持できる大きさの上限を表しており, その数は 150 人程度と推定されている. 本稿で Twitter ユーザー間のリプライ関係を分析し, 親密な “友達” の数には 200 人程度の限界があることを示した. この数は, SNS におけるダンバー数に対応している. SNS にあるダンバーの数の起源を理解するために, SNS 上の簡単なメッセージングモデルを提案し, メッセージを交わす関係の統計的性質を研究する.

キーワード: ソーシャルネットワーク, 利用者・コミュニティ分析, ツイッター, ダンバー数

Dunbar number that appears in the SNS and its origin -Data analysis and modeling-

NAOTO ARAKI^{1,a)} YASUHIRO HIEIDA² SHIN-ICHI TADAKI¹

Received: March 4, 2016, Accepted: August 1, 2016

Abstract: Dunbar's number represents the limitation on the size of close human relation, which is estimated about 150. In this report, we analyze the reply relation between twitter users and find that the number of close “friends” is limited at about 200. This number seems to correspond to Dunbar's number in SNS. For understanding the origin of Dunbar's number in SNS, we propose a simple messaging model in SNS and study the statistical properties of the messaging relation.

Keywords: social network, user community analysis, twitter, Dunbar's number

1. はじめに

人はどれほどたくさんの友人を持つことができるのだろうか. 人間では友達の数の限界が 150 人程度であるとする研究があり, その数はダンバー数と呼ばれている. 霊長類の脳に占める大脳新皮質の割合と群れの個体数の相関から, 1993 年に Robin Dunbar が提唱したものである. 猿や類人猿の例から推定すると, ヒトの友達の数の限界は 150 人程度が限界であるとされる [1][2].

Dunbar はダンバー数が 150 人程度であることの根拠として, 企業や部族, ビジネスの世界, 軍隊, 学問の世界, 伝統的な社会の規模に触れており, 多くのコミュニティは 150 人程度を限界として構成されていると論じている [3].

一方 SNS^{*1} ではどうであろうか. 例えば Facebook^{*2} は遠くに住む友人, 昔の友人とのつながりを復活させてくれる. 実際 facebook の友人の数が 150 人を超えるユーザを多くみることができる [4]. さらに twitter^{*3} においてフォロワー・フォロワーの数は時と共に増えていくものである. 実際, 日本人ツイッター利用者の平均のフォロワー数は約 400 であるというデータもある [5]. このような大きな人数と Dunbar の指摘は矛盾はないのだろうか. これが, 本研

¹ 佐賀大学工学系研究科
Department of Information Science, Saga University, Saga
840-8502, Japan

² 佐賀大学総合情報基盤センター
Computer and Network Center, Saga University, Saga 840-
8502, Japan

^{a)} 15573001@edu.cc.saga-u.ac.jp

^{*1} Social Networking Service

^{*2} <https://www.facebook.com>

^{*3} <https://twitter.com>

究の出発点である。

本研究では、twitter 上の人間関係の規模を調べることで、ダンバー数との対応を検討する。また twitter 利用者間のコミュニケーションをモデル化し、ダンバー数の起源を検討する。

2. twitter 利用者の分析

twitter 利用者の友達の数を調査する前に twitter における友達の定義と分析の対象とするユーザの条件の定義を行う。その定義に従い twitter 利用者の友達の数の分布を調査する。また相互フォローの関係の中に、どの程度友達と呼べるつながりがあるか調査し、ダンバー数との関連を調べる。

2.1 本研究における友達の定義

twitter 以外の SNS として、facebook や mixi^{*4} がある。日本の SNS としては発足が早い mixi は、サービス当初は紹介制を採用しており、「友達」という概念が明確である。一方、facebook は利用規程において実名登録を求めている。「友達」という概念もある。両者とも、実世界の友人関係が反映する度合いが強い。しかし twitter には「友達」という概念がないため、本研究においてその定義を行う必要がある。

先行の研究ではフォロー、フォロワー^{*5} の関係に焦点をあてている [6]。しかし Dunbar のいう友達の定義は「そこそこのお付き合い以上の関係」である。Dunbar は具体例をあげて「空港で会った時ににためらわずに話しかけられる関係」と述べている。

よって本研究ではフォロー・フォロワーの関係より強いつながりに焦点を当てる。A ユーザが B ユーザにリプライ^{*6}を送った時、A ユーザと B ユーザは友達であるとする。

2.2 分析対象ユーザについて

Twitter には様々なアカウントが存在する。この中には企業、ブランドアカウント、ニュース系アカウント等、実世界の普通の人ではないものが含まれている [7]。本研究ではそのようなユーザをできるだけ排除し、実世界の普通の人（一般ユーザ）を研究の対象とする。

ストリーミング API [8] で日本語が含まれているツイートをランダムに取得する。そのツイートを投稿したユーザのうち、以下の条件を満たすユーザを研究の対象とする。

^{*4} <https://mixi.jp>

^{*5} twitter では A ユーザが B ユーザをフォローすると A ユーザのタイムラインに B ユーザのツイートが表示される。その時 A は B ユーザのフォロワーと呼ぶ。相互のフォロワーの関係になると互いのツイートを互いに受け取る関係になる。

^{*6} 特定のユーザー名 (@...) から始まるツイートをリプライという。そのユーザー宛のツイートということになる。

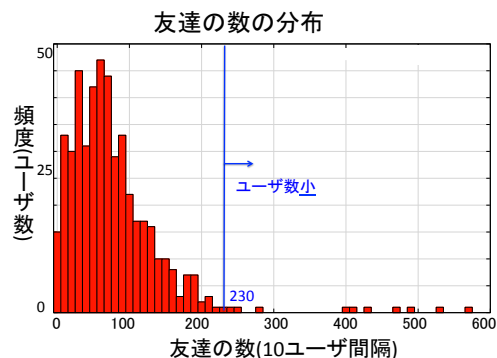


図 1 友達の数の分布. bin サイズは 10 としている。

Fig. 1 Distribution of the number of friends. The bin size is 10.

- (1) フォロワー数が 2000 以下であること^{*7}
- (2) ユーザ名、スクリーン名、自己紹介文に“bot”という文字列を含まないこと^{*8}
- (3) アカウントがプロテクトされていないこと^{*9}
- (4) リプライを 1 回以上していること^{*10}
- (5) 1000 件以上そのユーザ自身のツイートをしていること^{*11}

2.3 友達の数の調査

前述の方法で分析対象となる 483 ユーザを取得した。これらの各ユーザのツイートを最大 2000 件取得し、リプライを送ったユニークなユーザの数（友達の数）を調べた。図 1 はそれらユーザの友達の数の分布である。40～90 ユーザ程度の友達を持つユーザが多く、200 人以上の友達がいるユーザは数少ないことが明白である。

2.4 友達の数と相互フォロワーの数の関係

前節で収集した 483 ユーザに対して相互フォロワーの数を調べた。相互フォローの関係にあるユーザ同士は最低限お互いを認識しあっていると考えられる。その中でどの程度親密なつながりを持てるのかを分析する。

図 2 は相互フォロワーの数と友達の数の関係を表している。ダンバー数は 150 人程度とされるときに、その上限

^{*7} facenavi の調査によるとフォロワー数 2,000 人を超えるようなユーザは、一般的なユーザではなく、いわゆるインフルエンサー、企業、ブランドアカウント、情報・ニュース系アカウントがほとんどである。

^{*8} プログラムによる自動発言システム（“bot”と呼ばれる）が多数存在する。それらのアカウントにはユーザ名、スクリーン名、自己紹介文いずれかに“bot”という文字列が含まれていることが多い。

^{*9} プロテクトされたアカウントのツイートはフォローを承認されたユーザ以外見ることができない（ツイートのデータを取得できない）。

^{*10} リプライを全くしないユーザを調べたところ、企業、ブランドアカウント、bot 等がほとんどであると分かった。

^{*11} ツイート数が少ないユーザは twitter を始めて日が浅い場合や twitter 自体をあまり利用していない可能性がある。

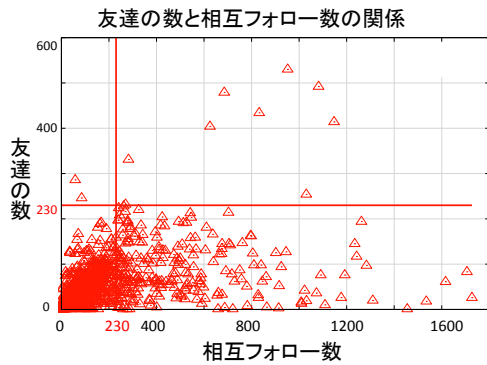


図 2 友達の数と相互フォロー数の関係. 二つの直線はダンバー数の上限 230 に対応している.

Fig. 2 Relation between the number of mutual followers and the number of friends, in the area under 230. The line is drawn by the least square in this area.

が 230 人程度とされている [1][2]. 図中の直線は 230 人に対応している. 相互フォローの数が 230 人以上のユーザは多くいるが, 友達の数が 230 人以上のユーザは極めて少ない.

相互フォロー関係にあるユーザ間の親密さを調べるため, 友達の数と相互フォローの数の相関を調べた. 図 2 を全体的に見ると, 相関関係がないように見えるが, 相互フォローが 230 以下の辺りには, 正の相関があるように見える. そこで相互フォローが 230 以下のデータのみを使用して, 相互フォローの数と友達の数の関係を表した (図 3). また相互フォローの数が 230 以下のデータを用いて最小二乗法により直線を引き, 全体の傾向を表した. 図 4 は全体のデータをプロットしたものに, 図 3 で引いた直線を引いたものである (なお直線の効果を明瞭にするため y 軸の範囲を 0 から 250 にしている). 図 3, 図 4 を見ると, 相関があることがわかる.

以上から相互フォローが 230 人以下のユーザはそれらのユーザと一定の割合でコミュニケーションをとることができることが分かる. 230 人を超えて相互フォローをもつユーザはそれら全員とコミュニケーションを取ることはできず, そのほとんどが 230 以下のユーザとしかコミュニケーションをとることができない. つまり, Dunbar の指摘のように, 密にコミュニケーションをとることができる人数に上限があることが推定できる.

3. モデリング

前節において, twitter 利用者の調査から親密な友達の数には 230 人程度という限界があるという結果を得た. また先行の研究によると, 友人 1 人当たりに対するリプライの数を調査したところ, 友達の数が 100~150 人を超えると, 友達それぞれへのリプライの量が減ることが報告されている [9]. これらの結果は, 人間には密な友人関係を維持でき

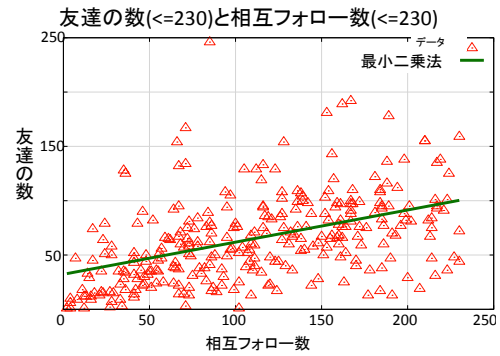


図 3 230 以下の領域における友達数と相互フォロー数の関係. 直線は, その領域に対する最小二乗法により求めている.

Fig. 3 Relation between the number of mutual followers and the number of friends, in the area under 230. The line is drawn by the least square in this area.

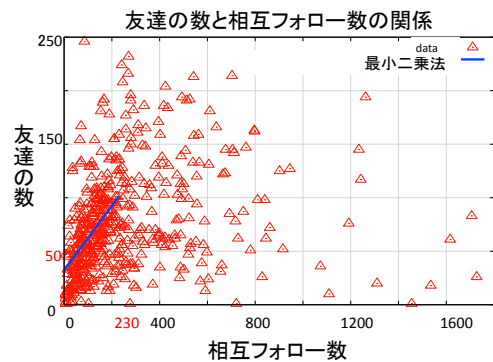


図 4 友達数と相互フォロー数の関係. 直線は, 230 人以下の領域で最小二乗法により求めている.

Fig. 4 Relation between the number of mutual followers and the number of friends. The line is drawn by the least square in this area under 230.

る人数に限りがあるというダンバー数が, SNS にも表れていることを示している. また興味深いことに, 友達の少ないユーザは友達それぞれに対するリプライの量が少ないことが分かっている [9].

そこで twitter におけるコミュニケーションを単純なモデルで表し, 隣接ユーザ 1 人当たりに対するリプライ回数 ω を調べることでダンバー数の起源を調べる. ユーザ i のリプライ総数 r_i を, 友人の数 k_i で除した値 (平均リプライ数) を次式で定義する.

$$\omega_i = r_i / k_i \quad (1)$$

Kwak らが 2009 年の取得データをもとに行った分析で twitter ネットワークのスケールフリー性を示している [10]. そこで, 本研究でもユーザの次数分布 $P(k) \propto k^{-\gamma}$ がべき乗則に従うネットワーク (スケールフリーネットワーク) であるとし, コンフィギュレーションモデル [11] を用いて,

べき指数 $\gamma = 2.4$, 頂点数 $N = 10^5$ ノードのスケールフリーネットワークを作成した. 各ユーザはこのネットワーク上で隣接しているユーザとリプライを交わすことによって他人とコミュニケーションをとる. なお, ここで構成したユーザのネットワークは, 単連結ではないことを注意しておく.

Dunbar は, ダンバー数の起源として記憶の有限性を想定している. そこで, 提案モデルでは, 一時的にメッセージを保持する記憶領域の有限性の効果を検討する. 各ユーザは, スタックとして記憶領域を持ち, その記憶領域が s_{size} という上限を持つと仮定する.

SNS におけるユーザーの活動度 (ツイートを行う頻度) は一様ではない. 例えばフォロワー・フォロワー数が多いユーザほど, ツイート数が多いことが示されている [10]. この効果をモデルに導入するために, 友達の多いユーザほど, 活動度が高いとする. 本モデルでは簡単のため, ユーザ i がツイートを行う確率 p_i はユーザーの次数 k_i に比例するとする. すなわち,

$$p_i \propto k_i \quad (2)$$

とする. また, 全ユーザのツイートを行う確率の平均をパラメタ P として表す. パラメタ P と個々のツイート確率 p_i の関係は次式とする.

$$P \approx \frac{\sum_i p_i}{N} \quad (3)$$

以降の節では, 普通のツイートとリプライツイートを区別するため, 普通のツイートをブロードキャストメッセージと呼び, 普通のツイートを行うことをブロードキャストすると呼ぶ.

3.1 各時刻における動作

本モデルでは各時刻にブロードキャストステップとリプライステップを行う. 各時刻の最初にブロードキャストを行い (以下の (1)), 次に特定のユーザに対するリプライを行う (以下の (2)). なお, 二つのステップとも, ユーザを選ぶ順序は, ランダムである.

- (1) 各時刻の最初に各ユーザ i はブロードキャスト確率 p_i に基づき隣接ユーザ全員へブロードキャストを行う. ブロードキャストメッセージは隣接ユーザの全員のスタックに入る. この時, スタックの上限まで既にツイートが入っている場合がある. その時はスタックの末尾のツイート (古いツイート) を押し出す.
- (2) 次にユーザそれぞれのスタックに対し次のような操作を実行する. あるユーザ i のスタックの中に入っているツイートの数と 0 の間の乱数を生成し, その数だけスタックの先頭からリプライを行う. ここでリプライ

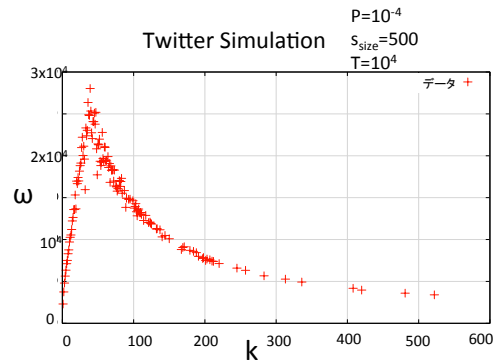


図 5 ユーザの次数 k と平均リプライ数 ω の関係. 各パラメータを $P = 10^{-4}$ と $s_{\text{size}} = 500$ と $T = 10^4$ とした. 同じ次数 k のユーザがいる場合その中央値を利用している.

Fig. 5 The average number of replies ω as a function of user's degree k for $P = 10^{-4}$, $s_{\text{size}} = 500$, $T = 10^4$. The median values are used for users of the same degree k .

を送った数だけユーザのリプライ回数 r_i を 1 増やす. さらにユーザ j がユーザ i からのリプライを受け取った場合, ユーザ j のスタックにはユーザ i のリプライが入れられる. ブロードキャストステップと同様, この時すでにスタックの上限までツイートが溜まっている場合, 古いツイートは末尾から押し出され, 新しいツイートが先頭から入れられる.

このモデルは先行研究のモデルを参考にしている [9]. 先行モデルとの違いは, 以下の二点である. 第一は, リプライすべきツイートの選択である. 先行研究では, ツイートに優先順位を設定し, ツイートを優先待ち行列に保存している. 本研究では, ツイートの優先順位は付けない. 第二の相違点は, ブロードキャスト確率である. 先行研究では, 全ユーザが同確率でブロードキャストするのに対して, 本研究では次数に比例した確率でブロードキャストを行う. これらの相違点の効果は節 3.2, 節 4 で議論する.

3.2 結果と分析

全体のブロードキャスト確率を $P = 10^{-4}$, スタックサイズを $s_{\text{size}} = 500$, 総時間ステップを $T = 10^4$ としてシミュレーションを行った, その結果を, ユーザの次数 k と (1) 式の ω の関係として表したものが図 5 である. 次数 50 付近で ω の値がピークを迎え, その後次第に減少している様子が分かる. これは友達の数が増えるとその友達全てと有効にコミュニケーションを交わせないことを示している. また最小の次数からピークの位置まで単調にリプライ回数が多くなっている. これはピークの位置までは友達の数が増えるほどその友達それぞれと有効にコミュニケーションを取ることができることを示している. また極端に友達の数が少ないユーザはその数少ない友達とも有効にコミュニ

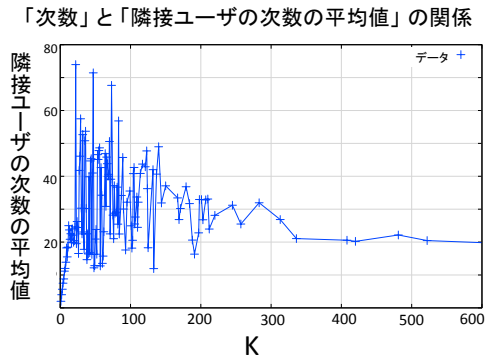


図 6 次数と隣接ユーザの次数の平均値の関係. 同じ次数のユーザが存在する場合その中央値を利用している.

Fig. 6 Relation between the degree k and the average degree of adjacent users. The median values are used for users of the same degree k .

ケーションを取ることができないことを示している.

$k \approx 50$ 以上において, ω が減少することは, スタックサイズが有限である効果である. 次数が大きくなり交わされるリプライの量が増えると ((2) 式), ユーザのスタックが飽和する. 飽和するとリプライできないツイートが増え, (1) 式の r の増加の傾きが小さくなる. ピークの位置より低い次数の地点の r の増加関数と, ピークの位置より高い次数の地点の r の増加関数の関数が違うため, このような結果となる.

次数が低いユーザの ω が低い原因は, 設定したブロードキャスト確率とネットワークの構造によるものである. ユーザのリプライ回数には隣接ユーザのブロードキャスト確率が影響する. ユーザの隣接ユーザのブロードキャスト確率の平均が高ければそのユーザはツイートを受け取る量が多くなるためリプライ回数が大きくなるはずである. 本ネットワークにおける隣接ユーザの次数の平均を計算すると次数が低いユーザが低くなっている (図 6). 次数が低ければブロードキャスト確率の値も低くなる ((2) 式) ため, 次数が低いユーザは周りのブロードキャスト確率も低くなる. よってスタックの中にツイートがたまらずリプライ回数が低くなっている. また自身のブロードキャスト確率が低いと自身から送るツイートの量が少なくなりリプライ回数が下がる原因となっている. 先行研究のモデルはキューの中にツイートが入る際, その送り主のユーザ次数の高さに優先順がつけられ, ユーザ次数の高い送り主のツイートが先にキューの中に入るといったものであった. その効果で次数が低いユーザはリプライをされず, 次数の低いユーザのリプライ回数が低くなる結果を示している.

3.2.1 時間の上限 T の効果

時間の上限 T はネットワーク内で循環するメッセージの数を調整する. 図 7 で T の値を変えたときの ω の挙動を示す. T を 3000~18000 まで 3000 刻みで値を変えてシ

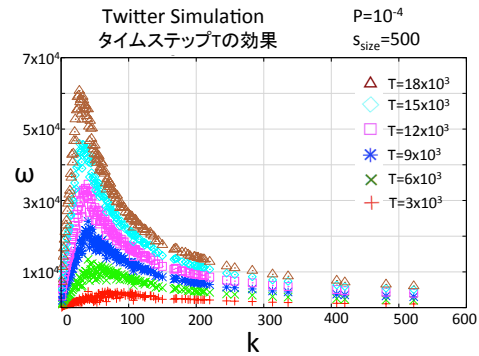


図 7 ユーザの次数 k と平均リプライ数 ω の関係に対する T の効果.

Fig. 7 Effect of T on the average number of replies ω as a function of user's degree k for $P = 10^{-4}$, $s_{\text{size}} = 500$.

ミュレーションを行った. T の影響で分布が縦軸方向に高くなっている. 時間が長いほど, ピークの構造が明確となる. つまりダンパー数が明瞭に現れる.

3.2.2 ブロードキャスト確率 P の効果

ブロードキャスト確率 P が次数の低いユーザに大きな影響を与えることは節 3.2.1 節直前の段落で明らかである. 図 8 で P の値を変えたときの ω の挙動を示す. 全体のブロードキャストの確率 P を 10^{-5} , 5×10^{-5} , 10^{-4} , 5×10^{-4} , 10^{-3} , 5×10^{-3} と値を変えて, シミュレーションを行った. ピークの位置が左上に上がっている. 次数が低いユーザのブロードキャスト確率が上がることで, 次数が低いユーザまでキューが飽和してしまうことが原因である. またこのパラメータは T 同様にシステム内を循環するメッセージの量を調整する. ブロードキャスト確率 P が上がることで全体のリプライの量が増える.

3.2.3 スタックサイズ s_{size} の効果

スタックサイズ s_{size} はピークの位置を調整する. 図 9 で s_{size} の値を変えたときの ω の挙動を示す. スタックのサイズを 500, 1000, 1500, 2000, 2500, 3000 と変えてシミュレーションを行った. ピークの位置が右上にずれていることは明らかである. まずピークの位置が右に変化しているのは s_{size} を大きくすることで次数が大きいユーザのスタックの中にツイートが溢れなくなることが原因であると考えられる. また上に上がっているのは棄却するツイートの量が減り, リプライする回数が増えることが原因である.

4. まとめと今後の課題

Twitter のユーザに関する分析はいくつか行われており, その多くはツイッターに存在する全てのアカウントを対象としている. しかしツイッターに存在するアカウントには一般ユーザとは言えないものが数多く存在する. 本研究では一般ユーザであるユーザを取得し, 友達の数分析を行った.

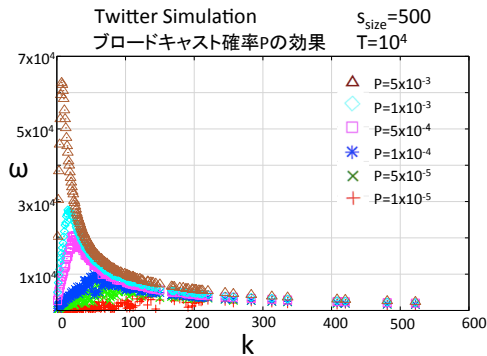


図 8 ユーザの次数 k と平均リプライ数 ω の関係に対するブロードキャスト確率 P の効果.

Fig. 8 Effect of broadcast probability P on the average number of replies ω as a function of user's degree k for $T = 10^4, s_{\text{size}} = 500$.

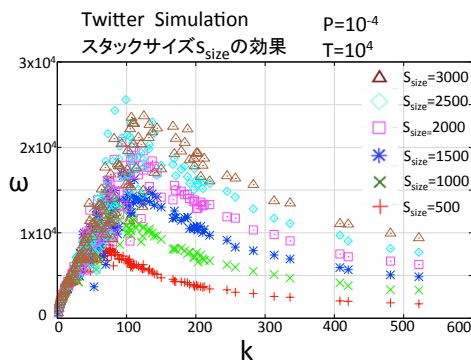


図 9 ユーザの次数 k と平均リプライ数 ω の関係に対するスタックサイズ s_{size} の効果.

Fig. 9 Effect of stack size s_{size} on the average number of replies ω as a function of user's degree k for $P = 10^{-4}, T = 10^4$.

フォロー・フォロワー関係を見ると多数と関係を持っているユーザが存在する。しかし“親密なつながり”に注目し、リプライを交わした相手が親密な友達関係を表すと考え、230人以上と親密であるユーザは稀であることが明らかになった。つまり、ダンバー数に相当する人数の上限が見えた。選択肢が増えれば増えるほど、その中で選んでいくことが難しくなるという指摘もあるように [12], SNS で沢山のひとと触れ合う機会が増えたとしても個人が持つ情報処理能力はそのような社会の変化に合わせて拡張しているわけではないように見える。

次にモデルを用いて、SNS に現れるダンバー数の起源を調べた。メッセージを一時保存するスタックのサイズが有限であることが、平均リプライ数 ω が $k \sim 50$ で頂点を持ち、それ以上の次数で減少する理由であることがわかった。先行研究では、メッセージに優先度をつけたが、優先度とは関係なく、次数の大きなユーザで平均リプライ数 ω の減少が起こることを示した。次数の小さなユーザで、平均リ

プライ数 ω が次数とともに増加することは、活動度が次数に比例して増加することの結果である。この点も先行研究にない効果を示している。

本研究のモデルでは、あるユーザのリプライ回数の総数を見ており、隣接ユーザそれぞれに対してリプライ回数の重み付けを行っていない。リプライを受けやすいユーザや多数のリプライを行うユーザの特徴を探り、SNS ネットワークの特性を探ることが課題の一つである。また、本研究でリプライ回数にはネットワークの構造による効果があることが明らかとなった。コミュニケーションにおけるネットワーク構造の効果について今後詳しく調査することも課題である。

参考文献

- [1] Dunbar, R. L. M. : *Neocortex Size As A Constraint On Group Size In Primates*, J. Human Evo., Vol. 22, No. 6, pp. 469-493(1992).
- [2] Dunbar, R. L. M. : *The Social Brain Hypothesis*, Evo. Anthro., Vol. 6, pp. 178-190(1992).
- [3] Dunbar, R. L. M. : *How Many Friends Does One Person Need?*, Faber and Faber(2010).
- [4] facenavi : facebook 日本人ユーザー 平均友達数, ページ数, 外国人友達率, 入手先 (<http://facebook.boj.jp/facebook日本人ユーザー-平均友達数, ページ数, 外国人>) (参照 2016-01-31).
- [5] Ishii, K.: Understanding Japanese Users on Microblog Twitter, Department of Social Systems and Management Discussion Paper Series, pp. 1-7(2011).
- [6] Akioka, S., Kato, N., Muraoka, Y., and Yamana, H. *Cross-media impact on twitter in japan*, Proc. 2nd Int. Workshop on SMUC., pp. 111-118, (2010).
- [7] facenavi : Twitter 日本人ユーザー・データ調査 フォロワー数, ツイート数など, 入手先 (<http://facebook.boj.jp/twitter-user-survey-2012>) (参照 2016-01-31).
- [8] 山本 裕介: Twitter API ポケットリファレンス, 技術評論社 (2010).
- [9] Goncalves, B., Perra, N., and Vepignani, A.: *Modeling Users' Activity on Twitter Networks: Validation of Dunbar's Number*, PLoS ONE, Vol. 6, No. 8, e22656, DOI: 10.1371/journal.pone.0022656(2011).
- [10] Kwak, L., Lee, C., Park, H., and Moon, S.: *What is Twitter, social network or a new media?*, Proc. 19th Int. Conf. on WWW, pp. 591-600 (2010).
- [11] 増田 直紀, 今野 紀雄: 複雑ネットワーク基礎から応用まで, 近代科学社 (2010).
- [12] Iyengar, S. : *The art of choosing*, Twelve, New York (2010).