

マイクロブログに現れる Dunbar 数とその起源

荒木直人¹, 日永田 泰啓², 只木 進一¹

¹ 佐賀大学工学系研究科

² 佐賀大学総合情報基盤センター

概要

Dunbar 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、その数は 150 人程度と推定されている。マイクロブログである Twitter ではフォロワーとリプライによって会話することができ、リアルタイムで親密な友人関係を反映していると考えられる。実際のデータから、このような相手の数は 200 人程度の限界があることが分かっている。Twitter にある Dunbar 数の起源を理解するために、Twitter 上の簡単なコミュニケーションモデルを提案し、メッセージを交わす関係の統計的性質を分析する。

Dunbar's number in a microblogging service and its origin

Naoto Araki¹, Yasuhiro Hieida², and Shin-ichi Tadaki¹

¹ Department of Information Science, Saga University, Saga 840-8502, Japan

² Computer and Network Center, Saga University, Saga 840-8502, Japan

Abstract

Dunbar's number represents the limit on the size of human group with close relationships, which is estimated about 150. In a microblogging service such as Twitter, a user can communicate with his/her friends through a reply method. Some observations of real data show that the number of reply-relationship friends seems to have the limit at about 200. For understanding the origin of Dunbar's number in Twitter, we construct a simple model of communication in Twitter and study statistical properties of the communication by simulations.

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。人間では親しい友達の数の限界が 150 人程度であるとする研究があり、その数は Dunbar 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 Dunbar 数が 150 人程度であることの例として、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いる [3]。

では SNS¹ ではどうであろうか。例えば Facebook² 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。友人の数が 150 人を超えるユーザが、Facebook 上に実際に存在する。一方マイクロブログ³ ではどうだろうか。マイクロブログの一つである Twitter⁴ ではリプライ⁵ によってコミュニケーションが可能である。リプライ関係は一般的な SNS のつながりに比べて親密であると考えられる。リブ

¹Social Networking Service

²<https://www.facebook.com>

³利用者は、主に自身の状況や雑記などを短い文章で、サービスを提供しているウェブサイトへ投稿する。ミニブログ内の利用者間でコミュニケーションを取れる様にもなっている。

⁴<https://twitter.com>

⁵ツイートに対する返信メッセージ

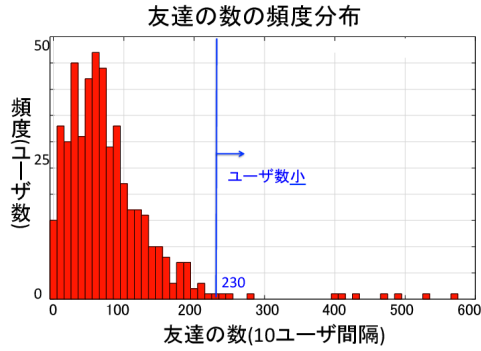


図1: 友達の数の分布。bin サイズは10としている。

ライを送った相手の数を友達と定義⁶し、その数を調査した結果の分布を図1に示している[4]。200人以上と友達であるユーザは極めて少ないことが明らかである。先行の研究において、友達(リプライ関係)1人当たりに対するリプライの数 ω を調査したところ、 ω は友達の数が100~150人程度でピークとなり、次第に減少することが報告されている[5]。このピーク位置はDunbar数と対応していると考えられ、人間には密な友人関係を維持できる人数に限りがあるというDunbar数が、マイクロブログにも表れていることを示している。

本研究では、Twitter利用者間のコミュニケーションを簡単にモデル化し、 ω が減少する要因を調査することでTwitterの中にDunbar数が表れる原因を分析する。

2 コミュニケーションモデル

Kwakらは、2009年の取得データをもとに行った分析から、Twitterネットワーク(フォロー関係)のスケールフリー性を示した[6]。そこで、本研究でもユーザの次数分布が $P(k) \propto k^{-\gamma}$ (k は友人数、 γ はべき指数)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、第3節で述べるCNNモデル(Connecting nearest-neighbor model)で作成した、頂点数 $N = 10^5$ 頂点のスケールフリーネットワークを用いる。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。

Dunbarは、Dunbar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

⁶Dunbarは友達の定義を「空港で見かけた時に気軽に話しかけることができる関係」と表現している。

にメッセージを保持する記憶領域が有限である事の効果を検討する。各ユーザは、メッセージを保持する有限サイズ s_{size} のスタックを持つとする。

マイクロブログにおけるユーザの活動度(ツイートを行う頻度)は一樣ではない。例えば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[6]。この効果を本研究でのモデルにシンプルに導入する為に、ユーザ i がツイートを行う確率 p_i はユーザの次数 k_i に比例するとする。

$$p_i = Ak_i \quad (1)$$

ただし、 i に依存しない比例係数 A は、(以下の(2)式を介して) P の関数となる。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。

$$P = \frac{\sum_i p_i}{N} \quad (2)$$

以降の節では、リプライツイートとそれ以外の普通の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」と呼ぶ。

本モデルでは各時刻に各ユーザは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))。

(1)[ブロードキャスト・ステップ]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)を消去する。

(2)[リプライ・ステップ]次に全ユーザは、それぞれのスタックに対し次のような操作を実行する。ユーザ i のスタックに n_i 個のメッセージがあるとき、処理するメッセージ数 m_i ($0 \leq m_i \leq n_i$)をランダムに選び、スタックの先頭から m_i 個にリプライする。ここでリプライを送った数(m_i)だけユーザのリプライ回数 r_i を増やす。リプライはリプライ先のユーザのスタックに格納され

る。ブロードキャストステップ (1) 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消去され、新しいツイートが先頭から入れられる。

なお、二つのステップとも、各ユーザはランダムな順番で選ばれる。

このモデルは先行研究 [5] のモデルを参考にしている。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、ツイートに発信者の回数によって優先順位を設定し、ツイートを優先順位付き待ち行列に保存している。本研究では、回数によってツイートの優先順位は付けず、新しいツイートをスタックの中に入れる。つまり単純に新しいツイートほど先に返信される。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、本研究では回数に比例した確率でブロードキャストを行う。

3 CNN モデル

人間関係では共通の友人を持つ 2 人はランダムに選択された 2 人よりも友人関係になる可能性が高い。これを表現したのが CNN モデル (Connecting nearest-neighbor model)[7] である。このモデルは以下のルールを反復して構成される。反復の各回で以下のいずれかの操作が選択され、2 の方が選ばれる確率は u である。

1. 確率 $1 - u$ でネットワーク中に新しい頂点 v_j を追加し、すでにネットワーク中に存在する頂点 v_i をランダムに選び (v_j との間に) エッジを張る。さらに、頂点 v_i の隣接頂点の全てに頂点 v_j から潜在的なエッジ (ポテンシャルエッジ) を設定する。これらを (真の) エッジとは見なさない。
2. 確率 u でランダムにポテンシャルエッジを選び (真の) エッジに変換する。

図 2 で示すように CNN モデルはスケールフリー性を示し、またクラスタリング係数や平均経路長など、SNS の実データと類似している点が多い [8]。

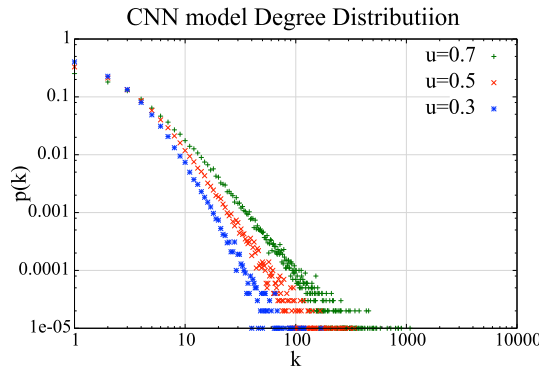


図 2: CNN モデルにおける回数分布。横軸は回数 k 、縦軸はその頻度 $p(k)$ である。

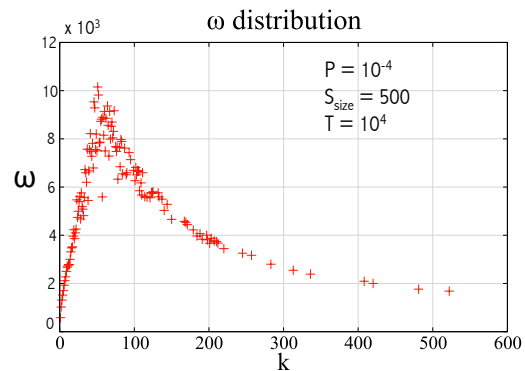


図 3: ユーザの回数 k と平均リプライ数 ω の関係。各パラメータを $P = 10^{-4}$ 、 $s_{\text{size}} = 500$ 、 $T = 10^4$ とした。同じ回数 k のユーザがいる場合その中央値を利用している。

4 結果と分析

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。

$$\omega_i = r_i / k_i \quad (3)$$

式 (2) のパラメータを $P = 10^{-4}$ 、スタックサイズ (第 2 節) を $s_{\text{size}} = 500$ 、第 2 節の (1)、(2) を繰り返す数を $T = 10^4$ 、CNN モデルの u (第 3 節) を $u = 0.5$ としてシミュレーションを行った。その結果を、ユーザの回数 k と隣接ユーザ当たりのリプライ数 ω の関係として表したものが図 3 である。回数 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見える。これは友達の数が増えると友人 1 人当たりのリプライ数が減ることを示している。

ω がある k を越えると減少することは、スタックサイズが有限である効果である。回数が大きくなり交わされるリプライの量が増えると、ユーザのスタック

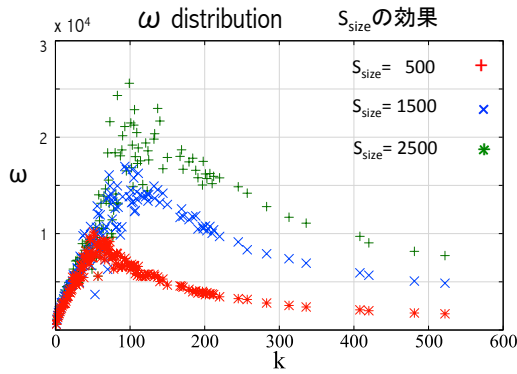


図 4: ユーザの次数 k と平均リプライ数 ω の関係に対するスタックサイズ s_{size} の効果。

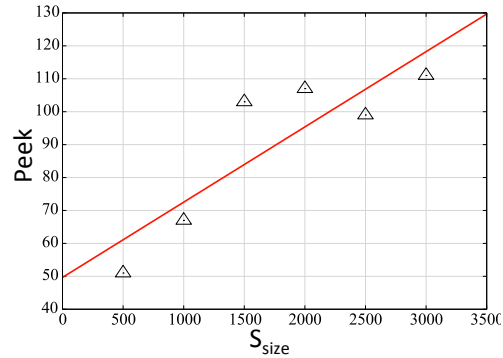


図 5: ω のピーク位置と s_{size} の関係。ピーク位置は ω が最大値をとる k 。

クが飽和する。飽和すると s_{size} を超えてメッセージを送ることができず、 k に対して r はほとんど変化しなくなる。よって式 (3) で r を k で除した時、 ω の分布にピークができる。また、ある程度大きな N では、結果が N に依存しない。

ここで、特にスタックサイズ s_{size} の効果を検討する。図 4 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、1500、2500 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。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。図 5 にピーク位置と s_{size} の分布を示す。ピーク位置のデータを直線で fit し、赤線で示している。 s_{size} によってピーク位置が線形に変化しており、ピーク位置を決める重要なパラメータであることは明らかである。また ω のピーク位置 (図 4) が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。

人間の記憶能力の制限をスタックサイズ s_{size} を使って表現したが、能力の制限が Dunbar 数の原因になっていることが予測できる。

5 まとめ

モデルを用いて、Twitter に現れる Dunbar 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、ある位置 k でピークをもち、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 [5] 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。

k に対する ω の関数のピーク位置は CNN モデルの u の値に依存している。 u の変化によるネットワーク特性の変化とピーク位置の関係の理解は今後の課題である。

参考文献

- [1] Dunbar, R. L. M., *J. Human Evo. A*, **469** (1992) 469.
- [2] Dunbar, R. L. M., *Evo. Anthro.*, **6** (1992) 178.
- [3] Dunbar, R. L. M. :*How Many Friends Does One Person Need?*, (Faber and Faber, 2010).
- [4] 荒木直人, 日永田泰啓, 只木進一, 情報処理学会火の国情報シンポジウム論文集⁷2016.
- [5] Gonçalves, B., Perra, N., and Vepignani, A., *PLoS ONE*, **6** (2011) e22656.
- [6] Kwak, L., Lee, C., Park, H., and Moon, S., *Proc. 19th Int. Conf. on WWW*, (2010) 591.
- [7] Vázquez, A., *Phys. Rev. E*, **67** (2003) 056104.
- [8] 内田誠, 白山晋, 情報処理学会論文誌, **47** (2006) 2840.

⁷<https://www.ipsj-kyushu.jp/page/ronbun/hinokuni/1005/2B/2B-3.pdf>