

## ツイッターの確率モデル

川本達郎

東京大学大学院理学系研究科 物理学専攻

ツイッターはマイクロブログとして世界的に人気の高いウェブサービスですが、ユーザーとしての興味はやはり「どれだけの人自分の文章(ツイート)を読んでいるか」だと思います。ツイッターでは、自分のツイートを直接受け取るフォロワーと呼ばれるユーザーたちが読むだけでなく、リツイートという機能によって、さらにたくさんの人に伝播していく可能性があります。ではそのツイートによる拡散過程には何か統計則があるのでしょうか。またあるとすればどのようなダイナミクスとして記述されるのでしょうか。ネットワーク上の拡散の問題としては、もっとも単純には各ノードが一定の確率で他とつながるパーコレーションが考えられます。また、病気の伝染モデルのように、各ノードに幾つかの相互作用のルールが与えられ、その運動を解くという問題もよく考えられます。ツイッター上の情報拡散は、素朴なパーコレーションほど単純ではなく、伝染モデルのような詳細な依存性を必要とするほど複雑でもないように思えます。今回、種ユーザーからどれだけ距離にいるかによってフォロワーを“世代”に分類し、各世代を一つの単位とした確率過程としてツイートの拡散ダイナミクスをうまく記述できるのではないかと考えました<sup>1</sup>(図 1 参照)。そのようなダイナミクスを考えると、 $(m-1)$  番目の世代のフォロワーがリツイートする回数  $n_m$  は、以下のようなランダム乗算過程として表されます：

$$n_m = \prod_{g=1}^m J_g N_0, \quad (1)$$

ここで  $J_g$  は確率変数、 $N_0$  は種ユーザーのフォロワー数です。このモデルの正当性は、幾つかのデータ解析で、大抵の  $J_g$  がおよそ対数正規分布に乗っていることによって示します。また、リツイート回数がゼロ回以上のツイートをランダムにサンプルした場合、リツイート回数の頻度分布はほぼベキ則に従うのが見えますが、他のウェブコンテンツにおいてリツイート回数のような、人気の指標となる量の分布 (Wikipedia の編集回数の分布など) は、多くの場合対数正規分布を示します [1,2]。ここでのツイッターのモデルは、このベキ則が生じることも説明します。

非常にシンプルなモデルですが、大まかにはこの程度でツイッターのダイナミクスは記述できてしまうこと、どれくらいリツイートによって拡散するかなどが定量的に解くことは、興味深いことではないでしょうか。

[1] Wu F, Huberman B A (2007) Novelty and collective attention. Proc Natl Acad Sci USA 104(45):17599-17601.

[2] Wilkinson D M (2008) Strong regularities in online peer production. Proc. of the 2008 ACM Conference on E-Commerce 302-309.

<sup>1</sup>ここで扱う拡散は、フォロワー間に限り、かつツイッター上のリツイートボタンを押したことによる伝播(公式リツイート)のみを考えます。

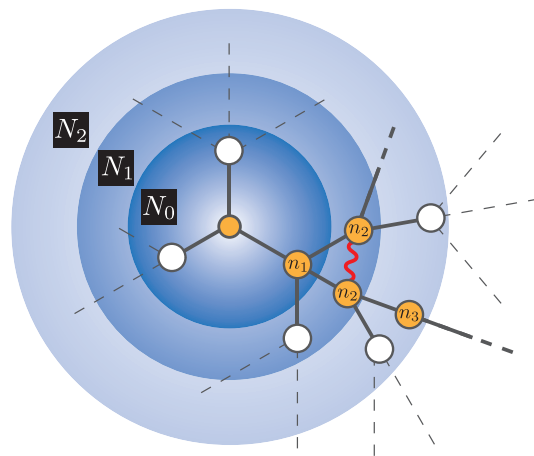


図 1: ツイッター上のフォロワーの世代ごとの分類と拡散の様子。中心のノードが種ユーザーを表し、リンクでつながったノードがフォロワーで、太線がリツイートによってツイートが伝播したことを示す。赤波線のようなループ構造は無視する。