HOKKAIDO UNIVERSITY

Title	時間遅れを考慮した情報伝播モデルの提案
Author(s)	小松, 琢也; 鈴木, 育男; 山本, 雅人; 古川, 正志
Citation	第13回公益社団法人計測自動制御学会システムインテグレーション部門講演論文集, 190-195
Issue Date	2012-12-18
Doc URL	http://hdl.handle.net/2115/64593
Rights	©2012 SICE. 本文データは計測自動制御学会の許諾に基づきCD-ROMから複製したものである
Type	proceedings
Note	第13回システムインテグレーション部門講演会(SI2012). 2012年12月18日～20日. 福岡
File Information	1C1_3.pdf

Instructions for use

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

時間遅れを考慮した情報伝播モデルの提案

○小松 琢也（北海道大学大学院），鈴木 育男（北見工大）
山本 雅人（北海道大学大学院），古川 正志（北海道大学大学院）

Propose an information propagation model with the time delay

○Takuya KOMATSU (Graduate of Hokkaido Univ.), Ikuo SUZUKI (Kitami Institute of Technology),
Masahito YAMAMOTO (Graduate of Hokkaido Univ.), and Masashi FURUKAWA (Graduate of Hokkaido Univ.)

Abstract: Rapid spread of the Internet has changed the behavior of the information propagation process. In this paper, we propose an information propagation model with the time delay, because the timing for receiving the information is different among users. Therefore, the proposed model considers the lifestyle of the Internet users. An analysis of the information propagation process by using our proposed model is presented.

１．緒言

近年のインターネット，携帯端末の普及に伴う Facebook，Twitter 等のソーシャル・ネットワーク・サービス(SNS)の出現により，人と人との間を情報が伝わるという意味での情報伝播は大きく変化してきた．従来の情報伝播では，情報伝達手段が口コミなどに限定されたため，人と人とが直接会う必要があり，ある人からある人へと情報が伝わるのにその物理的距離に依存して時間がかかった．しかし，SNS はそれを大きく変えた．

SNS はインターネットという仮想空間上に人と人とのつながりによる社会ネットワークを現実と同様の形で形成している．ただし，人々は携帯端末やパソコンなどを利用しインターネットを通じて直接会うことなく，情報を知人に送ったり，受け取ったりできるようになった．これらの情報伝播は，インターネットを利用しているため，従来とは異なり，情報が伝わる際の時間が飛躍的に短くなった．また，SNS 自体もその特徴として，情報が高速かつ広範囲にわたることが知られている[1][2]．このことは，従来と比較して SNS 上での情報伝播においては，物理的な距離に依存してかかる時間よりも，その人がどのタイミングで情報を取得しに行くのかといった個人の振る舞いの方が，より大きく情報伝播に対して影響を与えるようになったと考えられる．このような中で，伝播にかかる時間が情報伝播の速度やその伝播した規模にどのような影響を与えているのかを考えることは重要な要素となると考えられる．

本研究では，伝播に必要とする時間を伝播の時間遅れとして捉え，インターネット上での情報伝播，特に SNS における情報伝播において伝播の時間遅れに注目した情報伝播モデルを提案する．またそのモデルにおいて，個人の振る舞いとして人の生活時間を考慮した場合の伝播への影響について議論する．

２．関連研究

情報伝播の数理モデル化に関する既存研究として，IC(Independent Cascade)モデル[3]，LT(Linear Threshold)モデル[3]，SIR モデル[4]などがある．

LT モデル，IC モデルは社会ネットワーク上における情報伝播の確率モデルである．両モデルともネットワークの各ノードが，"アクティブ"と"非アクティブ"という二つの状態をとるモデルであり，それぞれ，状態がアクティブになる条件が異なる．

LT モデルは受信側を中心とした決定論的な情報伝播モデルである．ネットワークの各ノードに閾値，エッジに重み(あるノードに接続されている重みの総和は 1 以下)が与えられており，隣接ノード中のアクティブなノードとのリンクの重みの総和が閾値を超えている場合，自身の状態がアクティブへと変化する．

IC モデルは送信者を中心とした確率論的モデルである．ネットワークの各エッジに確率が定められており，アクティブ状態のノードが隣接ノードとのエッジに定められているその確率を用いて周囲のノード状態をアクティブ状態へと遷移させる．

また，SIR モデルは病気の伝播モデルとして発案されたものであるが，病気を情報ととらえることで情報伝播モデルとしても利用されている[2][6]．このモデルはネットワークの各ノードがそれぞれ状態 S(健康)，I(病気)，R(回復)という三つの状態をとり，感染確率 λ で状態 S から状態 I へ，確率 μ で状態 I から状態 R へと変化するモデルである．情報伝播においては，S は「情報を知らない状態」，I は「情報を受け取り周りに広めている状態」，R は「情報に対して興味がなくなった状態」に対応している．これらのモデルは一種類の情報が伝播する際のモデルであるが，複数種類の情報を取り扱

SY0012/12/0000 - 0190 © 2012 SICE

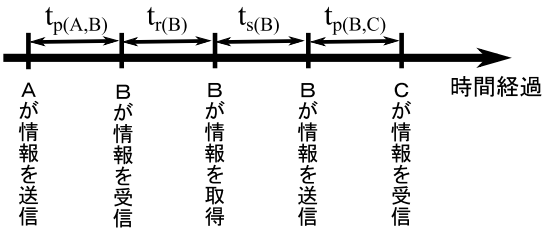

図 1　伝播の時間遅れ

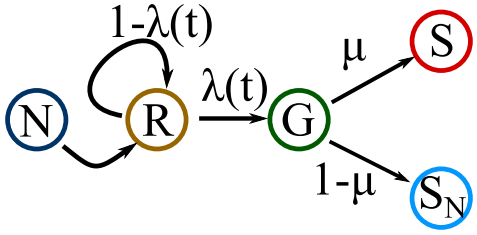

図 2　状態遷移図

うことのできるモデルも研究されている．藤田らは複数の商品の中から，どの商品を買うかを決定する状況において，機能や用途などの複数の情報が伝播する様子をモデル化している[7].

また，LT モデル，IC モデル，SIR モデルは，ネットワークにおけるノードが均一であると仮定したモデルであり，個性というノードごとの相違が表現されていない．そこで千葉らは，個人ごとの情報に対する興味の違いや話題の内容の違いを考慮したモデルを提案し，実データと比較することで提案したモデルの現実性を明らかにしている[8].

このように情報伝播モデルに対する研究は盛んに行われているが，人の生活時間による情報の伝播の遅れを考慮した伝播モデルは提案されていない．

３．情報伝播モデル

3.1 情報伝播と時間遅れ

図 1 はノード A からノードＢへ情報が伝播した後にノードＢからノードＣへと伝播した際の時間経過を表している．図中で″情報の受信″と″情報の取得″が分けられているように，本モデルでは，情報の受信と取得は同じ意味では用いない．情報の取得とは，我々がある情報を認識し，自分の知識として取り込む動作を指す．これに対して受信とは，情報を取得することができる可能性を持った状態になることを指す．つまり，情報を受信したあとでしか情報を取得することはできず，同時に情報を受信したからといって必ずしも取得できるとは限らない.図 1 にあるように本モデルでは，三つの伝播遅れを考慮する．

・物理的遅れ$t_p(i,j)$

送信者 i から受信者 j へと情報が伝播する際，物理的な影響による遅れで，送信者や受信者が原因とならないものである．例えば，SNS においては，書き込みを行ってからそれがサーバに送られ各ユーザーのタイムラインに情報が表示されるまでの時間に該当する．このような想定から$t_p(i,j)$はノード i とノード j のエッジに存在しているとし，エッジごとに設定されることが妥当である．

・受信遅れ$t_r(i)$

受信者 i が情報を受信してから，情報を取得するまでの時間遅れである．例としては，タイムラインに情報が表示されたあと，情報を受信者が確認して閲覧までの時間に該当する．

・送信遅れ$t_s(i)$

送信遅れとは，受信者 i が情報を取得し，その情報を周囲に伝播させるまでの時間に該当する．この場合の伝播とは「実際に受信者伝わること」ではなく，送信者の手から情報が離れた状態をさす．この値は取得した情報を踏まえて情報を伝播するかどうか決定する時間，情報を周りに伝播するためのメッセージを作成する時間に該当する．本研究では，ノード毎に大きく変化することはなく，また，情報によっても同様に大きく変化しないものとして考えられる．

3.2 伝播モデル

伝播遅れを表現するために，本研究ではエージェントベースの *SIR* モデルと同様に状態遷移を用いてモデル化を行う．ただし，本研究で想定するのは，手紙や直接会ってのコミュニケーションではなく，*SNS* やインターネット上の掲示板，ブログなどの，一度の送信で送信者の近傍ノード全てに情報が伝播し，受信者がその情報を取得するかどうか選択するような状況を考える．そこで，本モデルでは *SIR* モデルにおける確率 λ による状態遷移を周囲のノードの状態と関係なく，一度情報を受け取った後は周りと独立して状態を変更することとする．また，同様に送信はある情報を受け取った場合，その情報に対して一回のみ送信されるものと考える．図 2 に本研究で扱うノードの状態遷移図を示した．図のようにネットワークの各ノードは５つの状態を持つ．ノード i が情報を受信し，ノード j へと情報を伝播させる状況における各状態と遷移条件は次のように定義する．

・*N: (not received information)*

情報を受信していない状態であり，周囲のノードから情報を受け取ったときに状態 *R* へと遷移する．

・*R: (received information)*

情報を受信した状態であり，毎ステップ，時間関数 $\lambda(t)$の確率で状態 Gへと遷移する．

・G：(get information)

情報を受信した後，情報を取得した状態であり， $t_s(i)$ ステップ後に確率 μ で状態 Sへ遷移し，確率 $1-\mu$ で状態 S_Nへ遷移する．

・S：(send information)

情報を取得し，情報を送信した状態であり，状態 S へと遷移したステップに一度だけ隣接ノードのうち，状態 Nのノードを状態 Rへと $t_p(i,j)$ステップ後に遷移させる．

・S_N：(not send information)

情報を取得はしたが送信しなかった状態であり，この状態は他のどの状態にも遷移しない．

次に図 2 中の時間関数 $\lambda(t)$と確率 μ について説明する．まず，μ であるが，これは送信確率に該当する．この値が高い場合は，周囲に情報を伝播しやすいことを表す．次に $\lambda(t)$であるが，これは情報を取得する頻度割合の離散的な時間変化関数である．人の 1 日の生活リズムにおいて，どの時刻にどの程度の頻度でメールを確認するか，ポストを確認しに行くかなどを割合かしたものである．ただし，すべてのステップ t に対して $0 \leqq \lambda(t) \leqq 1$ 満たすものとする．

3.3 伝播アルゴリズム

本論文ではインターネット上での情報伝播のみに限定して考えることとする．このため，物理時間遅れ $t_p(i,j)$，および，送信遅れ $t_s(i)$をそれぞれ 0 とすることとする．これは，背景でも述べたようにインターネット上の情報伝播では，物理的な制約がないため，送信してから受信するまでの時間はほとんどないに等しい．また，送信にかかる時間もメッセージを作成する時間だけであるため，それほど大きいものではないと考えられる．本論文で提案する情報伝播アルゴリズムを以下に示す．

1 全ノードの状態を N にする．

2 任意に一つのノード j を選択し状態を S とし，その周囲の近傍ノードの状態を R とする．

3 状態 R のノード i は確率 $\lambda(t)$で状態 G へ遷移する．

4 上記で，状態 G に遷移した全てのノードについて同時に確率 μ で状態 S となり，確率 $1-\mu$ で状態 S_Nへと遷移する．

5 状態 S となったノードは $t_p=0$ であるので，状態 S になったノード j はその近傍ノードの状態を Rへと遷移させる．

6 状態 R のノードがなければ終了，状態 R のノードがあれば，ステップ(3)～(6)を実行する．

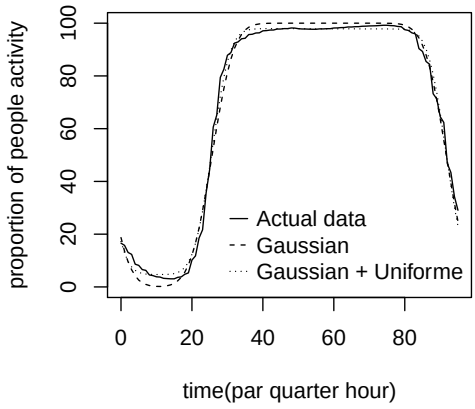

図 3　日本人の生活時間

3.4 人の生活時間

現実の情報伝播では，人は生活しながら情報伝播を行っている．このため，活動中の人は情報伝播可能であるが就寝中の人はすることができない．本論文では，この要素を表現するために，$\lambda(t)$を人の生活リズムに合わせた矩形波をと設定する．また，この矩形波を表現するためにパラメータとして，起床ステップ T^{wake}，一日のステップ数 T^{day}，活動ステップ数 T^{active}，矩形波の高さ λ を定める．また，このときに与えられる矩形波は

$$\lambda(t) = \begin{cases} \lambda(\text{活動中}) \\ 0(\text{就寝中}) \end{cases}$$

となる．ここで，現在のステップ t に対して，t'を

$$t' = \mathrm{mod}\, T^{day}$$

とし，T^{sleep} を

$$T^{sleep} = (T^{active} + T^{wake}) \,\mathrm{mod}\, T^{day}$$

と定義する．活動中となるステップ t の条件は 2 つあり，1 つは日をまたいで活動しない場合である．この場合は常に $T^{wake} \leqq T^{sleep}$ であるので

$$T^{wake} \leq t' < T^{sleep}$$

が成り立てば活動中になる．もう 1 つは，日をまたいで活動する場合である．この場合は $T^{wake} > T^{sleep}$ であるので

$$T^{wake} \leq t' \text{ もしくは } t' < T^{sleep}$$

が成り立てばよい．これ以外のすべての状態は就寝中に該当する．これらのパラメータを決定するために実データを利用する．利用するデータは日本政府が平成 18 年に調査した日本人の各時刻における活動内容に関して調査したデータである．このデータは全国 132520 人に対して，各時刻にどのような行動をしたか

を調査したデータであり，0 時から 24 時のあいだを 15 分ごと 96 に分割し，睡眠，食事，仕事などの全 20 項目の行動の内，各行動をした人の割合についてまとめてある．このデータから睡眠行動に関するデータのみを用い，解析を簡単にするため睡眠中ではない人は情報を受信できるものと仮定した．よって活動している人の割合 $A(t)$[%]は睡眠行動をとっている人の割合 $S(t)$[%]を用いて

$$A(t) = 100 - S(t)[\%]$$

とする．図 3 における実線が $A(t)$のグラフである．この $A(t)$[%]を用いて，パラメータ T^{wake}，T^{day}，T^{active} を定める．まず，問題を簡単化し，伝播遅れの影響を観測するために，ここでは各ノードの活動時間を全体で同一の値とする．$A(t)$は時刻 t における活動中の人の割合であるから，一日が T^{day} ステップで終わり，活動ステップ数が T^{active} である N 人の集合を想定し，それぞれの起床ステップ $T^{wake}{}_i$ を平均 a，分散，v の正規乱数で定め，各時刻における活動中の人数の割合 $F(t)$を求める．この $F(t)$と $A(t)$の二乗和誤差が最も小さくなるような T^{active} と平均 a，分散 v の組み合わせを求めた．ただし，探索時は N=10000，T^{day} =96 とする．

この結果 a=25.5，v=4.94，T^{active}=66.34 となった．その結果が図 3 の長い点の点線である．このときの二乗和誤差は 995.9338 である．ただし，図からもわかるように実データでは，各時刻において起きている人の割合は 100%になることはなく，また，0%になることもないが今回求めた $F(t)$では発生している．これは，正規乱数を使用しているためである．そこで，起床ステップを決める際に r の割合で一様乱数を発生させることで，全員が活動中であったり，就寝中であったりすることがないようにした．この結果，a=25.8，v=4.14，T^{active}=65.94，r=0.06 で最も二乗和誤差が低くなる．これが図 3 中の短い点の点線である．このときの二乗和誤差が 580.5038 である．

正規乱数のみを用いる場合よりも計算誤差が低くなっていることが確認できる．よって，今後の実験では，人の生活時間を再現する際には，この値の組み合わせ矩形波を作成する．ただし，この結果は一日のステップ数が 96 として考えているため，これ以外の場合，a=25.5，v=4.94，T^{active}=66.34 をそれぞれ T^{day}/96 倍する．

４．数値計算実験

4.1 実験条件

本論文で行う全ての実験では，一日のステップ数を 1440 ステップとする．これは一ステップが実際の時間における 1 分に対応する．このため，生活時間を考慮する場合は，人の起床ステップの分布を得るためのパラメータをそれぞれ 15 倍にし，a=382.5，v=74.1，T^{active} =995.1 を用いる．また，一様乱数を含ませる割合 r=0.05 とし，T^{wake} を設定する．各ノードの λ の値は 0.1 とする．生活時間を考慮しない場合の $\lambda(t)$は全ての t に対して

$$\lambda(t) = \alpha$$

とする．また，ここでは，伝播遅れの影響を観測するため情報を受け取った場合は必ず情報伝播が起こるものとして μ=1 としている．このため実験の終了条件を全ノードの状態が S となることとする．実験で使用するネットワークは CNN モデル[5](リンク率 0.8)により作成したネットワークを用いることとし，ノード数 N は N=10000 とした．CNN モデルは「友達の友達は自身の友達である」という実際の社会現象を再現したもので，次数分布が両対数グラフにて直線にのるためスケールフリー性を満たすほか，平均最短パス長も低くスモールワールド性を満たし，かつパラメータを調節することでクラスタ性を操作することのできる現実に近いモデルである．利用したネットワークの特徴量は次数平均 k=9.833，平均最短パス長 L=5.849,クラスタリング係数 C=0.166 情報伝播のスタートノードはネットワークの平均最短パス長 L=5.849 に最も近い平均最短パス長を持つノードを用いることとする．これは，スタートノードとして，ネットワークのどのノードとの距離も平均的なノードを選択したいためである．

生活時間を考慮した伝播の場合スタートステップが重要と予測される．そこで以下のステップにて，指定したノードの状態を S とし情報伝播を開始するものとする．

・ほぼ全ての人が就寝している(t=0)

現実での 0:00 に該当．殆どの人が寝ている時間．

・人が活動し始める(t=308)

現実世界での 5:08 に該当．正規乱数の平均 a=382.5 から分散 v=74.1 を引いた値とする

・ほぼ全ての人が活動している(t=720)

現実での 12:00 に該当．殆どの人が起床し活動している時刻．全ステップが 1440 であることからその半分の t=720 とする．

・人が就寝し始める(t=1303)

現実世界での 21:43 に該当する．正規乱数の平均 a=382.5 から分散 v=74.1 を減算し，活動時間 T^{active} を加算した値とする．

実験では，スタートノード，ネットワークトポロジー，起床時間の分布は同一のものを利用する．また，情報伝播を送信者が伝播させた情報を受信者が受け取り，認識するまでの過程であると考えるため，ネットワークにおける状態 G のノード数に注目する．

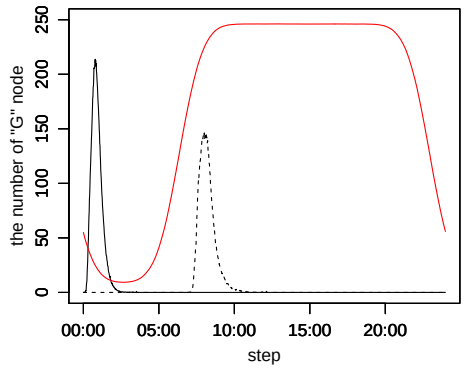

図 4　深夜に伝播させた場合

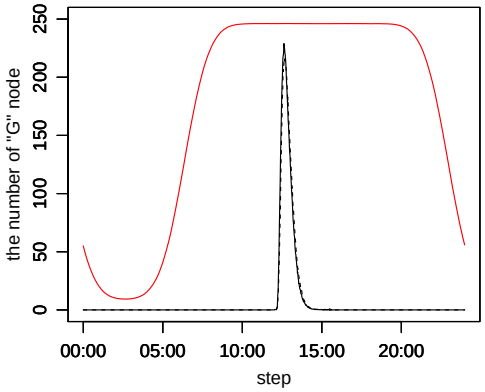

図 6　昼間に伝播させた場合

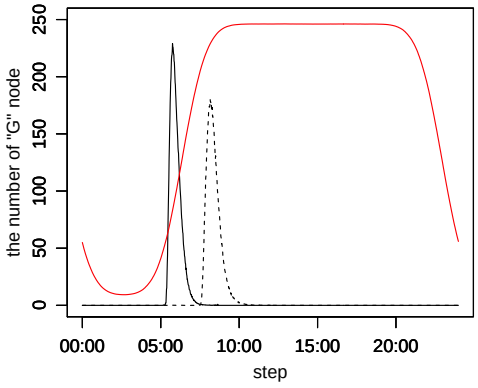

図 5　起床し始めに伝播させた場合

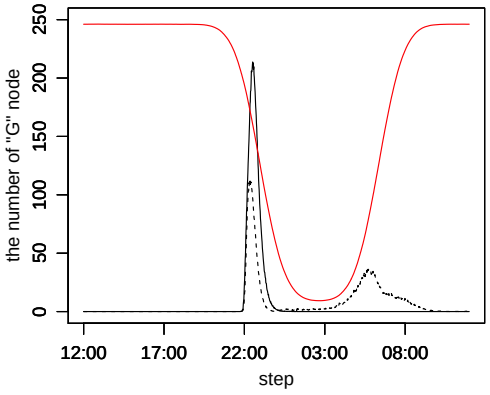

図 7　就寝し始めに伝播させた場合

4.2 実験結果

図 4～図 7 は 50 回の情報伝播を行い横軸にステップ数，縦軸に各ステップでの状態 G のノード数の平均をプロットしたものである．図 6 において，生活時間を考慮した場合としていない場合で，グラフの形はほぼ一致している．これは，伝播開始時刻を 12:00 と設定したため，ほとんどの人が起きている時刻から伝播される．よって，ほとんどのノードにおいて $\lambda(t)=\alpha$ となるため，生活時間を考慮しない場合と $\lambda(t)$が一致する．このため，ほぼ同様の結果となったと考えられる．次に図 4 を見ると，生活時間を考慮しない場合では，伝播が始まりすぐに全体に伝播が起きることが確認できる．これに対して，生活時間を考慮した場合では 450 ステップあたりから情報伝播が起きることが確認できる．生活時間を考慮すること，現実の 0:00 付近では，多くの人が寝ている．このため伝播が起こらず，朝になり大部分の人が起床し始めるステップ数まで，伝播遅れが発生したといえる．また，図 5 でも同様の現象が起きている．このことから，大多数の人が寝ているときに伝播をはじめると，ほとんど伝播せず，朝になり多くのノードが活動を始めると同時に伝播し始めることがわかる．図 7 において，生活時間を考慮した場合もしていない場合もシミュレーション開始後すぐに伝播し始めている．しかし，生活時間を考慮した場合では，すぐに伝播が収まり，その後，約 1700 ステップ付近でもう一度伝播が始まる．

これは，深夜に向かうにつれ就寝するノードが増加したため伝播がとまり，翌日の朝になり各ノードが活動し始めるまで伝播が遅れたといえる．

ここで，伝播に必要としたステップ数を状態 G のノード数が初めて 1 を超えてから再度 1 を下回るまでのステップ数と定義する．各スタートステップにおける

	生活時間を考慮す	生活時間を考慮しない
深夜	204	137
起床時	134	131
昼間	182	131
就寝時	712	137

表 1　伝播に必要とするステップ数

伝播に必要としたステップ数を表 1 にまとめた．図 4, 図 5 の両グラフにおいて，生活時間を考慮しない場合に対して，考慮した場合の伝播にかかるステップ数が多くなっている．また，生活時間を考慮した場合の情報伝播において，伝播にかかるステップ数が近い値を示している他，伝播を開始するタイミングがほとんど同じタイミングである．このことは，生活時間を考慮した伝播の場合，全体が起きる前にある程度の人が起きた時点から伝播が開始されることを示しており，このため，伝播が早い段階から始まるが，全体が起ききるまでは伝播が行われないため，伝播に遅れが生じたものと考えられる．今回の結果では，伝播開始が 450 ステップ付近であるが．これは，起床時間の分布を固定しているためであり，本来はもう少し早いタイミングになると考えられる．実際，起床時間の分布を変更して何度か実験を再度行うと平均して 400 付近の値に落ち着いた．以上の結果から，生活時間を考慮することで，伝播遅れが発生することを確認した．

５．結言

本研究では，情報伝播の時間遅れを考慮した情報伝播モデルを提案し，人の生活時間を考慮することで伝播遅れのモデル化を行った．人の生活時間を考慮した場合，ほぼすべての人が起きているときには，考慮しなかった場合と同様の伝播が起こるが，ほとんどの人が就寝している間に伝播した場合や，人が就寝し始めるタイミングで情報伝播を行った場合は周囲のノードが起床を始めるまで伝播が遅れることを確認した．また，朝，全体が起きていない時間に伝播を開始すると，昼に伝播を行うよりも伝播に時間がかかることを確認した．今後の目標は，活動時間を人により別々の値を設定し，それにより伝播遅れにどのような違いが発生するのかを調査したい．また，$\lambda(t)$に睡眠以外の要素を考慮した関数を与えることで，どのような違いが伝播に発生するのかを調査したい．

参考文献

[1]風間 一洋,"Twitter における情報伝播(<特集>Twitter とソーシャルメディア)",人工知能学会誌, 09128085, 27, 1, pp35-42, (2012)

[2]白井,嵩士，榊,剛史，鳥海,不二夫，篠田,孝祐，風間,一洋，野田,五十樹，沼尾,正行，栗原,聡,"Twitter ネットワークにおけるデマ拡散とデマ拡散防止モデルの推定",SIG-DOCMAS, B102 (2012)

[3]Kempe and J. Kleinberg　and E. Tardos,"Maximizing the spread of influence through a social network",Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137.146 (2003)

[4]Kermack.W.O, McKendrick,A.G.,Proceedings of the Royal Society of London A,115, pp.700-721,(1927)

[5]Vazquez, Alexei,"Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations."Phys. Rev. E pp.056.204 No.5 5 May(2003)

[6]中村 彰，片平昌幸，"感染症伝播の疫学と情報の拡散伝播"，第 5 回ネットワーク生態学シンポジウム予稿集, pp71-80 (2006)

[7]藤田 幸久,鷲田 祐一,鳥海 不二夫,植田 一博,石井健一郎,"情報の多様化を考慮した情報伝播のモデル化とシミュレーション",情報処理学会研究報告．MPS，数理モデル化と問題解決研究報告 09196072，2009，6, pp1-6 (2009)

[8]千葉 隆雄，鈴木 育男，山本 雅人，古川 正志，個人の特性と話題を考慮した情報伝播モデルの研究"，全国大会講演論文集, 71, 1, pp "1-247"-"1-248" (2009)

[9]e-Stat 政府統計の総合窓口　平成 18 年社会生活基本調査　調査票 A に基づく結果　生活時間に関する結果　時間帯編 全国 http://www.e-stat.go.jp/SG1/estat/Xlsdl.do?sinfid=000001095506

[4]神嶌 敏弘,"強調フィルタリングの課題：プライバシー，桜攻撃，評価値のゆらぎ",情報処理, 48 ,9 (2007)