事業者の凍結施策を考慮した拡張 SIR モデルを用いたデマ情報の拡散の分析

○関澤 佑（早稲田大学）

Analysis on Diffusion of Hoax Information Using Extended SIR Model with Freezing Measures by Business Operators

*Y. Sekizawa(Waseda University)

概要－ SNS 上で問題となっているデマ情報の拡散の問題に対し，従来の SIR モデルを用いたエージェントベースシミュレーションに，ネットワーク構造の違いや，悪質なユーザーを凍結させるという運営事業者の施策を含めて拡張し，デマ抑制に対しどのような効果があるかを確認する．

キーワード: CNN モデル，凍結施策

1. 研究背景と目的

1.1. 研究背景

2000 年代から，インターネット上に SNS が普及し，当初の gree や mixi などから，2010 年代以降 Instagram や微博，Twitter へと流行が移った．

2011 年の東日本大震災の際には，Twitter 上で被災地の状況や救援の情報や，テレビでは放送できなかった現場の生の情報が拡散され，情報伝達のツールとして大きな意義を見出した．報道各社も，ソーシャルリスニングとしてこういった生の情報から情報提供してもらい，実際のニュースに活用するなどの動きがみられる．

また，こういった各個人のクチコミは今現在の広告マーケティングなどに強い役割を果たし，デジタルマーケティングとして確立された．

その一方で，誤った情報や意図的なデマ情報（フェイクニュース）などは混乱を招き，社会問題になった．その原因として，正しい情報とデマ情報の分別が難しいという点がある．たとえば，東日本大震災で原子力発電が被害を受けた際，放射性物質を取り込まないためにうがい薬や昆布を摂取するとよいという情報や，関東地域において不足する電力は，他地域の節電によって補うことができるという情報が拡散されたが，これらの情報は後にすべてデマであると判明した．

その後，熊本地震や西日本豪雨の際にもデマ情報が流れた．また芸能人や，凶悪犯罪と名前が似ている人などへの根拠のないゴシップやそれに起因する誹謗中傷などは今もなお続いており，依然として問題が解決がなされないままとなっている．

アメリカでは，大統領選の際に対立候補のネガティブキャンペーンを意図的に行うフェイクニュースが拡散され，2019 年 9 月には，来年の大統領選の対立候補の不祥事を探すよう圧力をかけた，というデマに対し，トランプ大統領は釈明のため会見を開いた 1)

デマ拡散は，意図的に行われる場合と，周りを気遣って誤った災害情報などを流してしまう善意のパターンがあり，一概に発信者が悪いとも言えないのが現状である．

デマ拡散においては「凍結」の施策がとられることが多く，凍結されたアカウントは稼働することができなくなる．

しかし，実際のどのような場合に凍結するのが効果的かなどはわからないのが現状である．

1.1.1. デマ情報の実例①

2016 年 4 月に発生した熊本地震の際には，熊本県内の動物園からライオンが脱走したという内容のデマ情報が拡散された（Fig.1）2)

このツイートは 17,000 回以上リツイートされ，社会に大きな混乱をもたらした．熊本市動植物園にはこのツイートに関する問い合わせが 100 件以上寄せられた．

Fig.1: 熊本地震の際のデマ情報

のちにこのデマを発信した神奈川県内に住む会社員の男（20）は偽計業務妨害の罪で熊本県警に逮捕された．

1.1.2. デマ情報の実例②

2017 年に起きた東名高速のあおり運転の際には，犯人が，同じ苗字が含まれる会社の社長の家族とされるデマが拡散され（Fig.2），その会社にはそのデマを信じた多数のユーザーから誹謗中傷や抗議の電話が殺到した． 3)

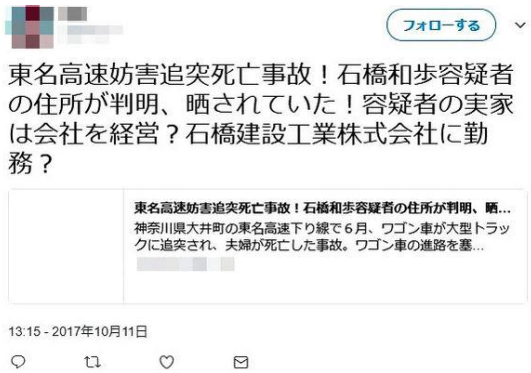

Fig.2: 熊本自身の際のデマ情報

会社はこのせいで休業を余儀なくされ，また社長は精神的苦痛を受けたとして書き込んだ男性らを相手に計 880 万の損害賠償を求める訴訟を起こした．

1.2. 研究目的

本研究では，デマ拡散の要因のうち，ネットワーク構造に着目し，その違いによって事業者の凍結施策の効果に違いがあるかどうか確認する．

2. ネットワークの諸モデル

2.1. モデルの概要

SNS 構造は，ノード（ユーザーを表す）とエッジ（ユーザー同士のつながりを表す）を用いてモデル化することができる．SNS を再現するモデルはいくつか提案されているが，本研究では次の 2 つのネットワークモデルを用いる．

2.1.1 ランダムグラフ

ランダムグラフは，各ノード間を一定の確率 u でつなぐことによって完成する．各ユーザーがみな同じような形態でつながるモデルである．

2.1.2 CNN モデル 4)

CNN モデルは，Connecting Nearest Neighbor モデルの略称であり，「友達の友達は潜在的には友達である」という仮定に基づき，SNS 構造を作り上げていく．

CNN モデルは以下のようなアルゴリズムで定義される．

① 確率パラメータ u を決める
② 以下のアルゴリズムを，ノード数が既定に達するまで行う
  ⅰ．確率 1-u で新規ノードを追加する
  ⅱ．追加したノードと，すでにネットワーク上に存在しているノードをランダムに一つ選び，エッジを繋ぐ
  ⅲ．追加したノードと，繋げたノードと繋がっているノードの間に潜在的エッジ（ポテンシャルエッジ）を張る
  ⅳ．確率 u で，ポテンシャルエッジをランダムに一つ選び，正式なエッジに変換する

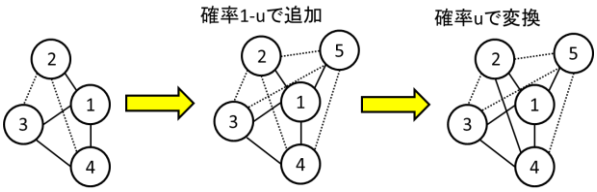

Fig.3: CNN モデルのアルゴリズム

以上の操作により結果，作られるモデルが CNN モデルであり，パラメータ u の値によって形状や特性の異なるネットワークグラフが得られる（Fig.4，5，6）．

この図においての青線は実際につながっているエッジ，緑線はポテンシャルエッジを表している．

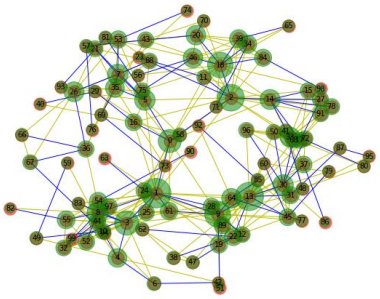

Fig.4: CNN モデル 確率 u = 0.1

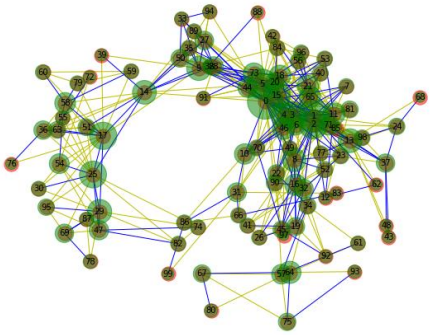

Fig.5: CNN モデル 確率 u = 0.5

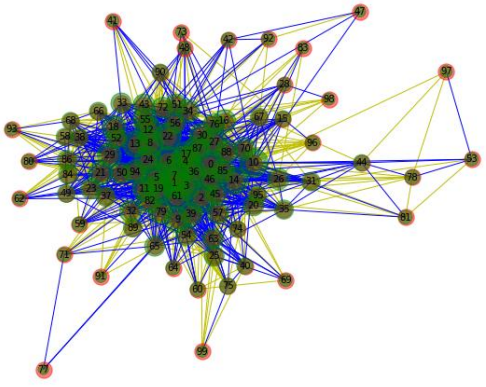

Fig.6: CNN モデル 確率 u = 0.9

上記の 3 つの図から, パラメータが低いと各ユーザーの密集度合いが低く, ハブとなるユーザーの規模が少ないことがわかる. パラメータ値が大きいと, 規模の大きいユーザーが生まれ, 情報拡散において強い役割を果たすと言える.

2.2. ネットワークモデルの評価指標

作られたネットワークグラフを評価するためにさまざまな指標が提案されている. 本研究では以下の代表的な評価指標を用いる.

2.2.1. 次数中心性

次数中心性とは, 各ユーザーと繋がっているユーザーの数の分布で定義され, そのネットワークのユーザーの規模の分布を知ることができる.

2.2.2. 近接中心性

近接中心性とは, あるユーザーからほかの全ユーザーへの最短経路の数の合計の逆数によって定義され, 近接中心性の高いユーザーは, 多くのユーザーから近くアクセスすることのできる「ハブ」になる可能性が高く, 影響力のあるユーザーであると言える.

2.2.3. 媒介中心性

媒介中心性とは, あるユーザーが任意の 2 ユーザーの最短経路の中にあるかどうかを表す尺度であり, 「そのユーザーが消失するとネットワーク自体が分断されてしまう」場合に高くなる.

3. 先行研究

白井らは, 伝染病が広まっていく様子を記述する SIR モデルを情報拡散に用い, SNS におけるデマ情報・訂正情報の拡散のモデル化を行った. 5)

SIR モデルは Susceptive（感染する可能性あり）, Infected（感染者）, Recovered（感染したのちに回復, 免疫を獲得）の 3 状態で感染症の患者数を表したモデルである.

池田らは, 白井らのモデルをベースとして, エージェントの多様性と情報価値の変化を考慮した「マルチエージェント型拡張 SIR モデル」を提案した. 6)

これにより, ユーザーの好みや感度をモデルに反映することができた. 先行研究や関連研究では, 実データの再現や要因分析が多く, SNS 構造そのものや事業者の役割がモデルに組み込まれいない. 本研究ではこの 2 点を表現した新たな拡張 SIR モデルを構築し, 特徴を分析する.

4. 本研究でのモデル

4.1. モデルの概要

本研究では, さまざまな形態のネットワークモデルにおいて, 運営事業者が悪質なユーザーを凍結させるといった施策を行った場合, デマ情報の拡散に違いがあるかどうかを分析し, どのようなネットワークにおいて有効なのか確認する.

従来研究のマルチエージェント型の SIR モデルに新たな状態"F（凍結）"を加えた SIRF モデルとすることで凍結施策を反映させている.

本モデルは, 初期状態作成・ユーザーの状態遷移・運営事業者による凍結・記録の 4 つのフェーズに分かれており, モデルの流れは以下のようになる.

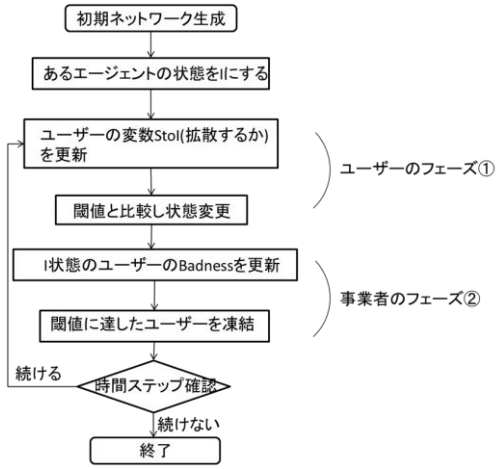

Fig.7: 本モデルのフローチャート

4.2. 各フェーズの挙動

ここでは, 既述の 4 フェーズを詳しく解説する.

4.2.1 初期状態作成

このフェーズでは, アルゴリズムに基づいてネットワークを作成し, その後の分析のための基本情報を定める. その具体的な方法などを以下に説明する.

4.2.1.(a) ネットワーク構築

このフェーズでは, その後のデマ拡散を行う舞台としていくつかの SNS モデルを構築する. まずスケールフリー性の有無による影響を調べるためにと, 確率パラメータ u を 0.1,0.5,0.9 の 3 通りの CNN モデルでネットワークを作成する（Fig.4, 5, 6）.

4.2.1.(b) 最初に感染させるユーザー

本モデルではデマ拡散の原因となるユーザーを最初に I 状態に変更してデマに感染させる. その数は総ユーザー数の 10%である. また, 最初にデマ情報を拡散させるユーザーは意図的にデマ情報を流しているものとし, その後の回復のフェーズでも回復しない.

4.2.1.(c) 時間ステップ

本モデルでは, 収束に十分かつ一般的にもデマ情報が収束, すなわち話題にならなくなる 3 か月に相当する, 90 ステップでモデル化する.

4.2.2. ユーザーの状態遷移のフェーズ

このフェーズでは, 拡張型 SIR モデルを用いて, ユーザーの S 状態から I 状態への推移, および I 状態から R 状態への推移を行う

4.2.2.(a) デマ情報への感染(S 状態から I 状態)

このフェーズでは, S 状態から I 状態への状態遷移をユーザーごとに行う.

その際, 自らの状態遷移の変数 StoI は以下の式のように定義する.

$$StoI_{user \cdot t} = StoI_{user \cdot t-1} + UserInterest_{user} UserSensitivity_{user} \sum_{n} UserInfluence_{n}$$

上記の式と, 初期値を与えた StoIthreshold を比較し,

S 状態から I 状態への状態遷移を行う．

また, UserInterest は興味度, UserSensitivity は感度, UserInfluence はフォローしているユーザーの影響度を表し，本モデルでは近接中心性を用いて定義する．

4.2.2.(b) 感染からの回復(I 状態から R 状態)

本モデルでは，感染から回復する確率を 0.205 とする．これは 7 日間で 8 割の患者が完治するように調整した値である．

4.2.3. 運営事業者による凍結のフェーズ

このフェーズでは, 悪質なユーザーを運営事業者が凍結するという施策を 2 つのパラメータを用いてモデル化する．その際，運営が凍結させる閾値を表す Operator BadnessThreshhold と，ユーザーの悪さを表す Badness という 2 つのパラメータを用いる．

OperatorBadnessThreshhold の値を変えることで凍結の強弱でどのようなデマ拡散の違いがあるかを確認する．

凍結のフェーズは以下のようなアルゴリズムで行う．

(a) .前の時間ステップで S 状態かつ，現在の時間ステップで I 状態に変わったユーザー，すなわちたった今デマに感染したユーザーを調べる．

(b) .そのユーザーがフォローしているユーザーを調べ，感染したユーザーの総フォロー数/感染したユーザーがフォローしている I 状態の人数の値を，そのユーザーがフォローしている I 状態のユーザーの Badness に均等に割り振り，蓄積させる．

(c) ．(a)，(b) をすべてのユーザーについて行う

(d) ．蓄積した Badness が運営事業者の凍結の閾値 OperatorBadnessThreshhold を超えたユーザーの状態を F 状態（凍結）に変更する．

凍結されたユーザーがアカウントを稼働できなくすることでネットワークから隔離され, フォロワーにデマ情報が伝わることがなくなる．

4.2.4. 記録

このフェーズでは, I 状態, F 状態の人数の記録を行う．

5. 結果・考察

本項ではネットワーク作成アルゴリズムによる違い・CNN モデルによるパラメータ u の確率による違い・凍結の閾値の違いの条件のもと, デマ拡散に対して凍結施策がどのような効果をもたらしているのかを分析する．

5.1. ネットワークモデルの性質

本モデルでは CNN モデル（u=0.1,0.5,0.9）とランダムグラフ（u=0.5）の 4 つのネットワークモデルを用いているが, 本項ではその性質について前述のネットワーク指標をベースに考察する．

5.1.1. 次数中心性

本モデルで用いた 4 つのネットワークモデルの次数中心性の分布は以下のようになる．

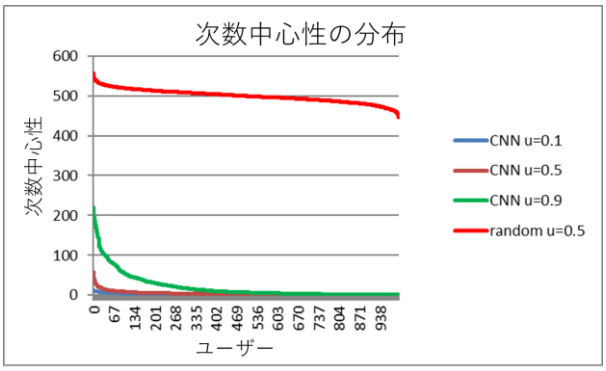

Fig.8: 4 つのネットワークモデルの次数中心性

この図から，CNN モデルでは u=0.9 の場合にほかの二つに比べ明らかに次数中心性の高い, すなわち多くのユーザーと繋がっているハブとなるユーザーが存在していることがわかる．

ランダムグラフにおいては，分布はほぼ u~0.5，すなわち 500 付近に集まっており，ハブとなるユーザーが存在しておらず, スケールフリー性のないモデルであることがわかる．

5.1.2. 近接中心性

本モデルで用いた 4 つのネットワークモデルの近接中心性の分布は以下のようになる．

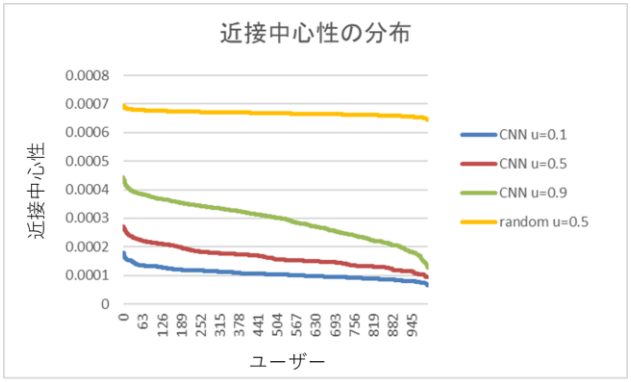

Fig.9: 4 つのネットワークモデルの近接中心性

この図から，近接中心性の分布は CNN モデルにおいては，u の値に従って分布することがわかる．u=0.9 の場合は近接中心性がほかの 2 つより倍近く高いユーザー, すなわちネットワークのすべての他ユーザーへ情報を伝えやすい, 影響度の高いユーザーが存在している．実際の SNS では芸能人やインフルエンサーと呼ばれるアカウントである．

ランダムグラフでは, ネットワークにスケールフリー性がなく, 確率に基づいてすべてのユーザーが同じようにつながるため, 近接中心性の分布はほぼ一定の範囲内に収まっている．

以降の分析ではこのようなネットワーク構造の違いが, アカウント凍結施策にどのような影響を及ぼすのか確認する．

5.1.3. 媒介中心性

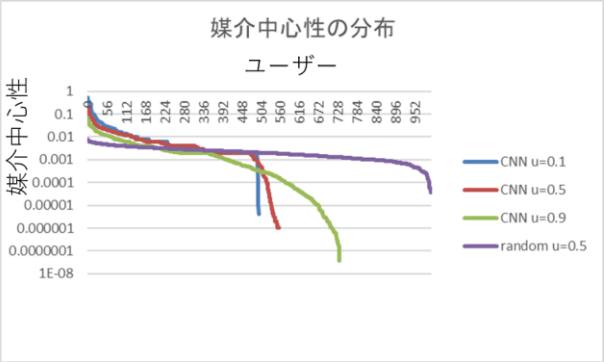

Fig.10: 4 つのネットワークモデルの近接中心性

この図から，媒介中心性に関しては CNN モデルおよびランダムグラフのいずれの場合によっても傾向が似ており，違いは現れなかった．

5.2. 試行結果

本モデルは全ての場合において 1000 試行行っている．その結果，ランダムグラフかつ凍結の閾値 0.1 の場合以外の全場合において同様の傾向が見られたため，以降の分析では平均値を典型例として扱う．

また都合上すべての図は載せられないため，Fig.11 に一部の場合を示す．ランダムグラフかつ凍結の閾値 0.1 の場合は 2 つのパターンが見られたため，別途考察する．

たとえば，Fig.11 の場合は I 状態の人数が最初右肩上がりに増え，15 ステップ前後でピークを迎え，その後回復および凍結のいずれかがなされることで一定の値まで下がり，収束しているといった傾向がすべての場合においてみられる．

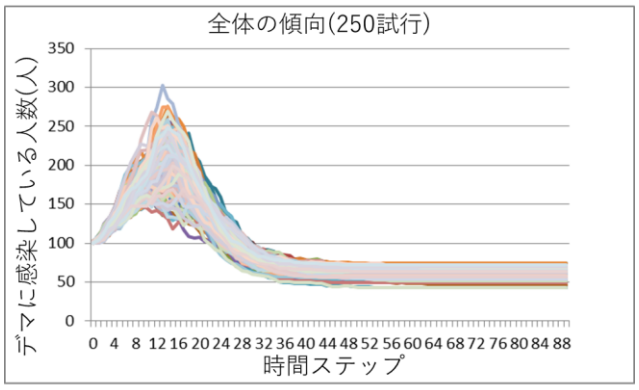

Fig.11: CNN モデル（u=0.9），凍結の強さ 0.1，250 施行の結果

5.3. ネットワーク作成アルゴリズムによる違い．

ネットワークの違いによって全体の I 状態の人数にどのような動きがみられるか確認する．

ランダムグラフでは確率に偏りが見られないようパラメータを 0.5 として出力を行っている．

凍結の強さは一番強い 0.1 として実行すると以下の Fig.12 のようになった．

この図から，最も強く凍結した際に CNN モデルで収束が行われ，45 ステップほどで全ユーザーが I 状態になることなくデマが収束したが，ランダムグラフでは，収束する場合（パターン 2）もあるが，ハブとなるユーザーが存在せず，多くのユーザーが均等につながっているため，あるきっかけで新たに感染する可能性がある（パターン 1）．

F 状態（凍結状態）の人数（Fig.13）も，パターン 1 では収束せず，2 回にわけて凍結がなされていることがわかる．

すなわち，悪質なユーザーを逐次凍結してもランダムグラフでは効果が表れず，デマが再度拡散してしまう可能性があり，凍結施策を行う場合はその SNS のネットワーク構造がスケールフリー性を持つ CNN モデル寄りなのかランダムグラフ寄りなのか着目し，実行するか決める必要がある．

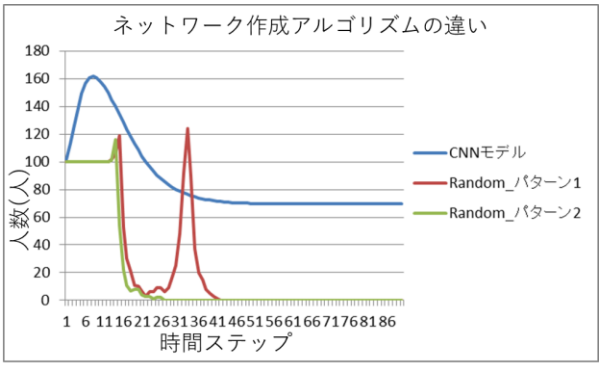

Fig.12:ネットワーク作成方法ごとの
I 状態の人数の推移

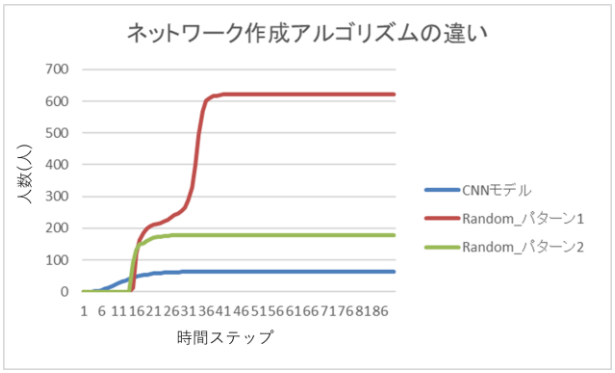

Fig.13:ネットワーク作成方法ごとの
F 状態の人数の推移

5.4. CNN モデルによるパラメータ u の確率による違い

ここでは違いがはっきり見られる CNN モデルの確率パラメータ u=0.1 0.5 0.9 の 3 つにおいて，どのような差がみられるか確認する．

凍結の閾値は，強く凍結する 0.1 の場合を用いた．結果が以下の図である

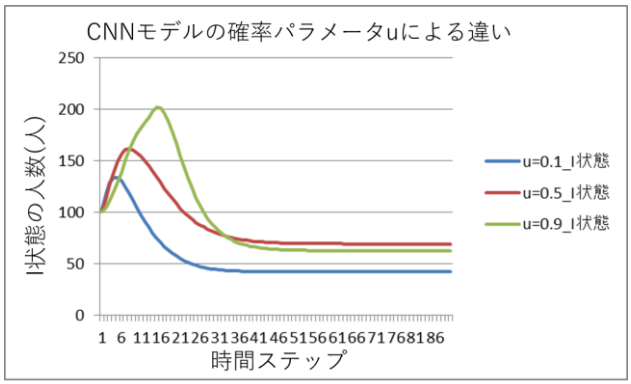

Fig.14: パラメータごとの I 状態の人数の推移

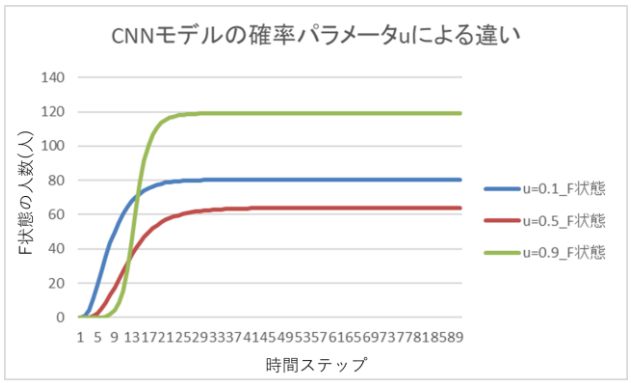

Fig.15: パラメータごとの F 状態の人数の推移

Fig.14 から，ハブの影響度が強く，多くのユーザーが互いにつながっている u=0.9 の場合は感染者の人数が一気に増えるため凍結が追い付かず，ピークが高くなる．

あまり各ユーザーがほかのユーザーとつながらない u=0.1 の場合は，ハブとなるユーザーがあまり存在しないため，そもそも感染があまり起こらず，起こってもすぐに凍結することでデマ拡散の抑制において効果があることがわかる．

また，Fig.15 から u=0.1,0.5 の場合は逐次凍結していくため，F 状態の人数は緩やかな曲線を描くが，一気にデマが拡散してしまう u=0.9 の場合は大量のデマ感染者が一気に凍結されるため，急な曲線を描いている．

ここから，凍結施策のみでは完全にデマを抑制することは出来ず，何らかの別の施策も併せて行うと望ましいことがわかる．

5.5. 凍結の強さによる違い

本モデルでは，一度凍結しなかった際の最も悪質なユーザーの Badness を基準として 9 段階に分け，凍結の閾値を感度分析している．凍結の閾値が低いほど，すぐに凍結される，すなわち強く凍結するということになる．

結果は，より違いが分かりやすくはっきりとみられる凍結の閾値 0.1,0.5,0.9 の場合を用いる．

ここでは，最もスケールフリー性を持つ SNS 構造が再現されている CNN モデル（u=0.9）の場合を用いて，凍結の閾値の感度分析の結果を示す．

その図が以下のようになる．

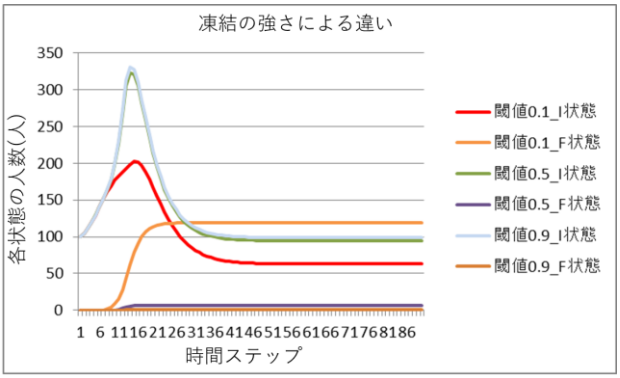

Fig.16: 凍結の閾値による I 状態と F 状態の人数の推移

Fig.16 から強い凍結を表す閾値 0.1 の場合はほかの 2 つの場合と比べ早く，6 ステップごろから凍結が始まり，結果感染した人数が明らかに少なくなっていることがわかる．

しかし，凍結の閾値が 0.5（中），0.9（弱）の場合はほとんど違いがみられないことがわかる．

ここから，凍結を実質する場合は中途半端に行っても意味がなく，強く凍結する必要があることがわかる．

6. 結論

本研究では SNS 上でのデマ拡散の問題を，SIR モデルに凍結施策を反映させたスケールフリー性を持つ SNS のネットワークモデルを構築した．

また，CNN モデルのパラメータや，凍結の強さによる感度分析を行って，運営事業者の凍結という施策の効果を確認した．

その結果，影響力の強いハブのうちユーザーが集中しているスケールフリーネットワークにおいて，強い凍結施策を行うとデマ状態が抑制されることが分かった．

参考文献

1)http://www.news24.jp/articles/2019/09/26/10505826.html
2)https://www.huffingtonpost.jp/2016/07/20/lion-escape_n_11081056.html
3)https://www.asahi.com/articles/ASM3743XKM37TIPE019.html
4) Alexei Vazquez ：Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, Physical Review E 67,1/15 (2003)
5) 池田圭佑, 岡田佳之, 榊剛史, 鳥海不二夫, 篠田孝祐, 風間張一洋, 野田五十樹, 諏訪博彦, 栗原聡 ：マルチエージェント型拡張 SIR モデルを用いた情報拡散シミュレーションの評価, 情報処理学会, ICS-173 No.7, 1/7 (2014)
6) 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, 人工知能学会, SIG-DOCMAS-B102,pp.1/9 (2012)