

パーコレーション理論を応用したコミュニティの情報伝播モデルの作成

Community Information Propagation Model using Percolation Theory

安藤 祐貴^{*1}
Yuki Ando

平田 均^{*1}
Hitoshi Hirata

矢入 郁子^{*1}
Ikuko Yairi

^{*1} 上智大学大学院 理工学研究科理工学専攻情報学領域
Graduate School of Science and Engineering, Sophia University

Abstract: In recent years, with the spread of the Internet and smart phones, the method of information propagation has diversified. This is probably because not only the method of acquiring information increases but also people who did not have a method of transmitting information so far can easily transmit information. There are few models that can mimic this diversified information propagation, and examples in which the prediction is out of the way with the conventional method are starting to be seen. In this paper, we propose a new model applying percolation theory as an information propagation model applicable to modern complex society. In addition, we show the significance by extending the model and measuring its effect. This model It is possible to show information propagation in a more realistic environment by adding parameters.

1. はじめに

近年のインターネットや SNS の普及に伴い情報の流通方法が多様化するなか、人々が情報を取捨選択して意思決定するまでのプロセスも複雑化してきている。2016 年の米国大統領選で事前予想が大きく外れた例は記憶に新しい。この事例ではあらゆるメディアを用いた虚々実々の情報合戦によって人々を惑わし多くの浮動票を生んだこと、SNS による情報操作があったことなどが指摘されている。本稿では、人々がある情報を正しいと判断して、他の人に伝える行動を情報伝播とよび、コミュニティ内での情報流通の仕組みを数理的に解明することを目的とする。そしてこの情報伝播を、化学における浸透現象と類似のものと捉えてパーコレーション理論を適用し、コミュニティ内での情報伝播をコンピュータシミュレーションを通して分析する。情報伝播に関する研究は数多く行われている。これらを大きく分けると特殊なネットワーク上で情報伝播を汎用的な情報伝播モデル (SIR モデルなど) でシミュレートしているものと、汎用的なネットワーク上で特殊な情報伝播モデルを用いてシミュレートしているものに分かれる。まず特殊なネットワークをシミュレートしているものを紹介する。風間は Twitter ネットワークの繋がり方やその構造がどのようになっているか、またそのネットワークが特定の情報伝播時にどのように変化していたかを解析している [風間 2012]。Meeyoung らは Flickr ネットワークにおける伝播域の広さと情報伝播の速度、情報伝播におけるロコミの役割について解析・研究している [Cha 2009]。Yong らはオンラインのソーシャルネットワークと物理的なネットワークを結合した多層ネットワークモデルを提案しており、このネットワークに SIR モデルを適用した場合の情報伝播・拡散について研究している [Zhuang 2016]。Liang らは伝染病の流行モデルを応用してモジュラーネットワークにおける情報伝播現象に関する調査を行っている [Huang 2006]。Zijie らは高速道路上の車両間で形成されるアドホックネットワークにおける情報伝播の過程について検討している [Zhang 2011]。Bernhard らは現実の人間のネットワークとしてグループの社会構造を含むモデルを提案し、モデルを用いてシミュレートした結果の調査を行っている [Voelkl 2010]。次に特殊な情報伝播モデルを用いてシミュレートしているものを紹介する。

竹内らはソーシャルネットワーク上で情報共有を実現するための情報伝播モデルを提案し、モデルを用いた社内コミュニケーション支援システムを作成、評価を行っている [竹内 2005]。Jichang らはソーシャルネットワークにおける情報伝播をシミュレートするための情報拡散モデルの提案とその実証を行っている [Zhao 2012]。このように情報流通に関する研究には、ネットワーク構造に着目したものや、特殊な条件の情報伝播モデルを提案するものが存在するが、本研究のように人々の情報の受容と発信を浸透現象と捉えて確率的に再現して分析した例はない。本稿では 2 章でパーコレーション理論の説明をした上で、類似研究について紹介と本研究におけるモデルの応用を説明する。3 章では単純なモデルの構築、実装と単純な実験を行う。4 章では複数の情報を用いた情報伝播モデルに関する実験を行い、5 章ではモデル全体に対する考察を行う。

2. パーコレーション理論と情報伝播モデル

2.1 パーコレーション理論

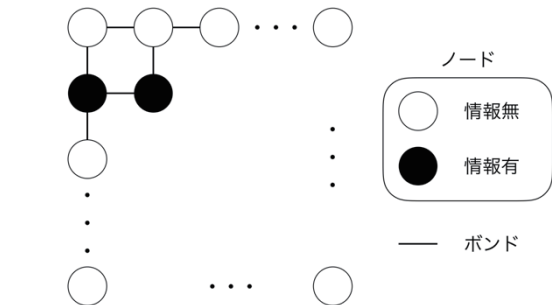
パーコレーション理論とは 1957 年に Broadbent と Hammersley [Broadbent 1956] によって導入された統計理論の一つであり、様々な構造を持ったネットワークにおいてノードからノードへの浸透を確率で行うことで浸透を模倣している。この理論は物理学をはじめとして様々な分野で利用されており、情報科学においても応用されている。

2.2 本研究における情報伝播モデルの定義

本稿ではコミュニティの人の情報伝播をパーコレーションとみなす。Fig.1 にネットワークの構造の概要を示す。正方格子状に配置した人間をノードとし、人同士の繋がりをボンド、これらで構成されるコミュニティをネットワークと定義する。情報を得ているノードが得ていないノードに対してボンドを通じて浸透を試みることで情報の伝播を模倣する。通常、伝播はボンドの確率によって制御されるが、本モデルでは人の持つパラメータが伝播を左右すると考え、ノードが伝播確率を持つ。Fig.2 に 3×3 のノードを持つネットワークにおける伝播のイメージを示す。中心の伝播済ノードからステップ毎にボンドを通じて周囲のノードに伝播を行う。この時、ノードの持つ伝播確率によって伝播の可否が決定されるため、必ずしも伝播が行われるとは限らない。伝播されたノードは次のステップから伝播済ノードとなり、周囲に対して伝播を始める。情報伝播自体のシミュレーションを行うモデルの研究は非常に少なく、

連絡先: 氏名, 所属, 住所, 電話番号, Fax 番号, 電子メールアドレスなど

既存の研究では前述の通り非常に限定的な環境や特徴的なシチュエーションに合わせたモデルが多数を占めている。本研究のモデルはこれらに対しコミュニティの人の情報伝播という抽象的な事象をモデル化している。さらにパラメータの追加によって現実的な環境に合わせたシミュレーションが可能となる。これは既存の研究ではできない事象であり、他の研究にも応用可能な



方法となり得る。

Fig.1 ネットワークの構造概要

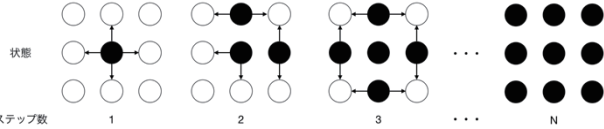


Fig.2 3×3 のノードの伝播イメージ

3. モデルの実装と単純な数値実験

3.1 一種類の情報の伝播の場合の実装概要

まず一種類の情報をネットワーク上で伝播させた場合について考える。さらに作成したモデルに対してパラメータを追加することでより現実的なモデルとして扱うことができるか実験を行う。51×51 のノードからなるネットワークを用意し、様々なパラメータを追加した状態で実験を行うことでモデルの評価を行う。実験では最初に初期化を行い、伝播を実験毎に定めたステップ数繰り返す。次節にパラメータと実験内容について記述する。実験は MacBook Pro 2017 macOS High Sierra 10.13.5 で行い、言語は python anaconda3-4.4.0 を使用した。51×51 の配列を 2 つ用意しそれぞれに伝播確率、状態を格納する。伝播確率には実数値、状態には 0 と 1 (伝播なしと伝播済) の状態を格納することとした。この配列を用いて伝播を繰り返す。伝播では全てのノード (配列の全要素) を走査し、状態が伝播済なものについて周囲への伝播判定を行う。初期状態としてネットワーク中心のノードに情報が伝播された状態とし、以降の条件を追加した状態で実験する。

忘却

情報伝播において発生し得る、忘却に関するパラメータを追加する。伝播確率は全てのノードで一定とし、各ノードに対して忘却確率を設定する。これによりコミュニティにおける伝播と忘却の関係性を観測することが可能である。伝播確率を p 、忘却確率を q として変動させた場合、集団がどのようになるのか実験する。0.1 から 0.9 の範囲 (0.1 刻み) で p と q を変動させ、100 試行後に一番回数が多かったものを結果とする。このとき蔓延は総ノード数のうち 90% が伝播済の状態、消失を全てのノードが伝播なしの状態を、均衡を前述の二つに当てはまらない状態を指すこととした。伝播処理の後に忘却処理を行う。忘却処理では伝播済ノード全てを走査し、ノードに対して忘却確率を用いて忘却処理を行う。忘却が行われたノードは状態が伝播前となる。

伝播確率の分布設定

各ノードの伝播確率を分布として与えることで、人間が持つ個人差を反映させることを考える。ネットワーク内のどの程度のノードがどのような伝播確率を持つ、といった設定をすることでネットワークの特性を変更することが可能である。伝播確率の確率分布を複数用意し、どのような分布が伝播しやすいのかを実験する。このとき、全ての分布は確率の期待値が同時値になるようにする。これにより全てのノードはそれぞれが確率分布に従った確率を 1 つ持つようになる。Table 1 に使用する分布とそれぞれの比率、排出する伝播確率を記述する。それぞれの分布全体の伝播確率の期待値は 0.05 とした。ノードの初期化時に確率を設定する。この時に全てのノードに対して初期化を行うが、伝播確率の決定を関数化し、関数内で分布を作成することでノード全体の伝播確率を制御する。

Table1 伝播確率の分布

分布名	比率	確率
基準	1:8:1	0.02,0.05,0.08
同一	1	0.05
均一	1:1:1:1:....:1	0.005,0.015,...,0.095
二値1	850:49	0.001,0.9
二値2	1:1	0.005,0.095

3.2 解析結果

忘却に関する実験の結果

伝播確率 p と忘却確率 q の値をそれぞれ 100 ステップ動かし、100 試行した際の値がどのようになったかを Fig.3 に示す。 q の値を 0.7 以上にしてしまうと p のなく値に関係消失してしまうことがわかった。本モデルに忘却のパラメータを追加した場合、その影響は非常に大きく、パラメータ q によって結果が大きく変わると言える。また、伝播確率 p と忘却確率 q が均衡する点が $q \approx 2p$ の近傍に存在することがわかる。この実験の結果は人間の忘却の強さを示していると言える。大抵の出来事は忘却の影響が強く出てしまい、最終的に再度誰かが提唱されない限り消えてしまう。また、一度均衡した情報は長期間の間コミュニティに止まり、全体には広がらないものの消えもしない。

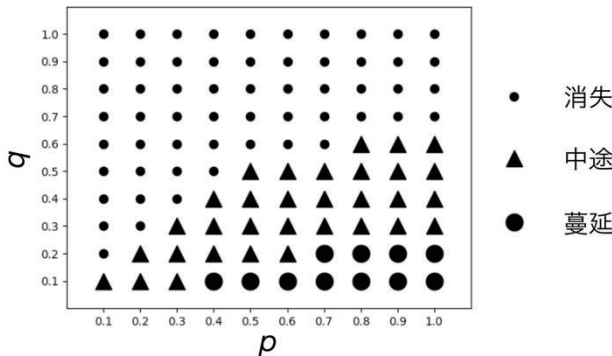


Fig.3 伝播確率と忘却確率の関係性

分布に関する実験の結果

前述した分布において、全てのノードが何ステップで伝播済になったかを示す。基準分布は 204.47, 同一分布は 198.31, 均一分布は 576.06, 二値分布 1 は 833.72, 二値分布 2 は 257.87 となった。数値は全てのノードが伝播済になるまでにかかったステップ数の平均値である。実験から同一分布が一番伝播する能力が高いことが観測がきた。これはコミュニティの性質によって情報

伝播の能力が変わるという社会での実例と符合していると考えられる。

3.3 考察

忘却確率における均衡する点について

Fig.3 より、均衡する点はおおよそ $q = 2p$ となる線分の周囲であると言える。しかし、この関数はボンド数の変化があった場合に対応することができない。なぜなら、伝播確率はボンド数に依存しているが、忘却確率はノード数に依存しているためである。このことより、ボンド数を考慮した均衡点の式は $q = p \times x/2$ であると予想する。ここで x は 1 つのノードがもつボンドの数とする。

どのような分布の伝播力が高いのか

結果より分布の伝播力の高さの順列を確認することができた。ここからどのような分布の伝播力が高いのかを考える。各分布の確率の分散を計算し、結果との関連を確かめた。それぞれの分散の値は基準分布が 0.00018、同一分布が 0、均一分布が 0.000825、二値分布 1 が 0.04165、二値分布 2 が 0.002025 となった。同一分布が一番分散の値が小さいこと、また分布に関する実験の結果とほぼ同時順番で並べられることがわかる。このことより伝播確率の分布は分散が低い方の伝播能力が高いと言える。

4. 複数種類の情報の伝播

次に複数種類の情報をネットワーク上で伝播させた場合について考える。初期状態として各情報毎にランダムな場所のノードを一定数伝播後にし、次項以降の条件を追加した状態で実験する。なお本研究では、一度伝播されるノードは伝播するノードからの情報を受け取り、以降は受け取った情報を伝播することとした。また、一度情報が伝播されたノードは他の情報を受け付けないものとする。

4.1 実験内容

初期配置に差のある場合

初期配置に差がある場合、伝播力にも差が出るのか考える。初期配置をする際に、集団として配置する。具体的な配置方法を Fig.4 に示す。このような初期配置を 1 つの情報に対して適用する。その後全てのノードがどちらかの情報を伝播済になるまで伝播させる。この時どちらの伝播力が高いかを伝播済のノード数を用いて比較する。なお伝播確率の分布は 1:8:1 の比率で 0.2,0.5,0.8 とした。初期化時にはノード数の調整とブロック化を行う必要がある。ノード数の調整に関しては、 2×2 、 3×3 の情報について初期化を行い、同様の数のノードを伝播済とする。ブロック化に関しては、ランダムで決定した伝播予定のノードの周囲 (2×2 の場合 3、 3×3 の場合 8) マスを同時に伝播済とした。

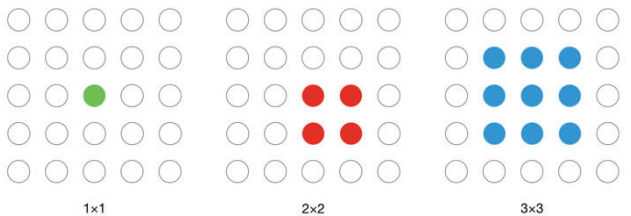


Fig.4 2×2 と 3×3 の初期配置

伝播確率が低い場合

前項にて記述した実験の伝播確率を極端に低くした場合にどのような変化が起こるか実験する。使用する伝播確率の分布は 1:8:1 の比率で 0.02,0.05,0.08 とした。

伝播分布が違う場合の比較

前節にておこなった、伝播分布による伝播力の違いを再度比較する。1 つの情報につき 1 つの伝播分布を適用し、同時に 2 種類の情報を伝播させる。この時どちらの伝播力が高いか、また、どの程度の差が各分布に存在するのか伝播済ノード数を用いて比較実験する。分布の組み合わせ毎に 100 試行ずつを行い、片方の情報が伝播されたノード数を記録する。この場合伝播分布を複数持つ必要があるため、伝播確率を格納する配列を一つ追加する。伝播処理を行う際に参照する配列を変更することで、複数の伝播分布を使い分けることが可能となる。

4.2 解析結果

初期配置に差がある場合の結果

Fig.5,6 に試行した結果の一例を示す。緑が伝播前、赤と青がそれぞれの情報が伝播済であることを示す。 2×2 の場合は $1721 : 880$ 、 3×3 の場合も $1940 : 661$ と同様に伝播力が低くなってしまっている。これによりコミュニティ内において集団で固まってしまうと、情報伝播が行われにくくなってしまい、コミュニティ全体に対しての影響力が下がってしまうと言える。

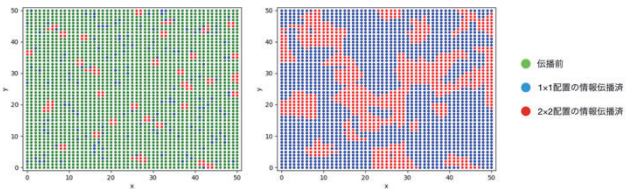


Fig.5 2×2 との伝播力の比較結果

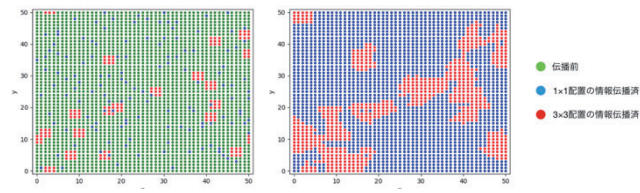


Fig.6 3×3 との伝播力の比較結果

伝播確率が低い場合

前項に比べ $2161:440$ と伝播力の差が広がってしまっているのがわかる。これより伝播確率が低ければ低いほど、伝播力の差が顕著になっていくことがわかる。

伝播分布が違う場合の結果

Table2 に試行結果を示す。表の数値は総ノード 2601 個のうち A 側の分布を用いた情報が何個になったかの平均値を表す。一種類の情報の伝播を行なった場合と同じく、分散が小さい伝播確率の分布を用いた方が伝播力が高いと言える。これに関してもコミュニティの性質によって情報伝播の能力が変わるという社会での実例と符合していると考えられる。値は 2×2 の場合は $1287 : 1314$ 、 3×3 の場合は $1301 : 1300$ となった。このことより情報の伝播において大事なことは伝播の候補をいかにして増やすかであることがわかる。

Table2 分布による伝播速度の違い

A\B	基準	同一	二値2	二値1	均一
基準	x	-1393.04	-743.91	-272.17	-1169.78
同一	1393.04	x	-698.15	-251.77	-1144.03
二値2	743.91	698.15	x	-746.63	-1798.66
二値1	272.17	251.77	746.63	x	-2291
均一	1169.78	1144.03	1798.66	2291	x

4.3 考察

初期配置によって生まれた差を埋めることができるか

前節より、ノードをブロック状にまとめて配置した場合伝播力が低くなってしまう現象があった。これはノードの伝播可能なボンドの数が減ってしまうことが原因であると考えられる。ここでまとめて配置した場合に均衡を取るためにはノード数をどの程度増やせば良いのかを考える。伝播可能なボンド数を一致させることで伝播力が同一になると仮定する。次の式で 1 つのブロックで伝播可能なボンド数を求め、その差を埋めるノードを初期に配置することを考える。 2×2 の場合伝播可能なボンド数は $8, 3 \times 3$ の場合は 12 となる。これに対し同時個数のノードが持つ伝播可能なボンド数は 16, 36 となる。つまり初期配置するブロックの数をそれぞれ 2, 3 倍にすることで釣り合いが取れるはずである。この条件で前述した実験を行う。実際に初期配置数を調整した結果の例は次の Fig.7 のとおりである。

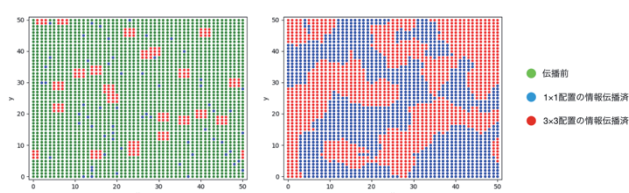


Fig.7 3×3 の初期配置数調整後

伝播確率によるモデル特性の変化

前節より伝播確率が低い状態で複数の情報を伝播した場合、伝播力の差が顕著になるという結果が出た。なぜこのような差が起きたかを考える。伝播力は要素を考えると伝播できるボンド数とステップ数、伝播確率などがある。伝播が進むにつれて伝播済ノードが増え、それに伴い伝播可能なボンドも増加する。伝播確率が低い場合ということは新たな伝播済ノードが増えるスピードが遅いということである。つまり伝播可能なボンド数に差が発生するとその差の影響を受けているステップが多くなる。また伝播確率が非常に低い状態では伝播可能なボンド数が伝播力に大きな影響を及ぼす。これにより差が出やすくなってしまうと考えられる。実際にどの程度の差が生まれるのかを実験する。 3×3 の、伝播確率が低い場合の実験を用いて追実験を行う。具体的には初期配置数を変化させることで、どの程度の初期位配置数で釣り合いが取れるようになるかを検証する。まずは考察にて行った初期配置数の調整を適用する。初期配置数を単純に 3 倍にした場合、終了後の数は 975 : 1626 と差がある状態となってしまう。さらに補正をかけ、配置数を 4 倍程度にした場合、終了時の数は 1342 : 1259 とある程度同数となっている。このように伝播確率が低い場合、伝播力にこのような影響を及ぼすことが確認できた。比率が同じ場合であっても結果に差異が出るため、モデルの拡張時にはこの点に留意する必要があると考えられる。

5. 全体考察

5.1 パラメータの追加

受容確率

人と人との情報伝播において重要なのは話すことと聞くことの 2 つである。話し手が話す確率を伝播確率としているが、その情報を聞き手が受け入れるか否かを考慮するパラメータが存在しない。これを追加することでさらに現実に即したモデルになると考える。

近傍数の変化

近傍の数についても変化させることが簡単に可能となる。例として少数のノードの近傍、つまりボンドが非常に多い状態を考える。このようなノードをインフルエンサとして定義することで、現実におけるインフルエンサの持つ情報伝播能力の高さをさらに表現しやすくなる。

5.2 モデルが表せるもの

ここまでモデルの拡張性とその効果の測定を行ってきた。このモデルでは人のコミュニティにおいて発生する情報伝播を模倣することができる。このようなモデルは現在非常に少ない。また、コミュニティや人を現実的に考えると、個人差やコミュニティの性質を考慮する必要がある。このモデルではこれらを確率関数として定義し、パラメータとして追加することで状況に合わせる事が可能となる。このモデルを応用することで特殊な状況下におけるシミュレーション (i.e. 選挙戦のシミュレーション) を行うことができる。と考える。

6. おわりに

本研究ではパーコレーション理論を応用した情報伝播モデルの作成を試みた。簡素な状態でも様々な状況を模倣することができ、パラメータを追加していくことで非常に多くの状況をシミュレートできるモデルを作成することができた。今後は更なる追加パラメータの考案やその実験、効果測定などを行なっていくことで作成したモデルの有用性を模索していく。

参考文献

- [Broadbent 1956] Broadbent, S. R. and Hammersley, J. M: Percolation processes, I. Crystals and mazes, Proc. Cambridge Philos. Soc., 1956
- [堀 1973] 堀素夫: 浸透の理論とその応用, 日本統計学会誌, 日本統計学会, 1973
- [加藤 2002] 加藤孝明: パーコレーション理論を用いた市街地の防災性評価, オペレーションズリサーチ, 日本オペレーションズ・リサーチ学会,
- [風間 2012] 風間一洋: Twitter における情報伝播, 人工知能学会誌, 人工知能学会, 2012
- [Cha 2009] Meeyoung Cha, Alan Mislove and Krishna P. Gummadi: A measurement-driven analysis of information propagation in the flickr social network, WWW '09 Proceedings of the 18th international conference on World Wide Web, International World Wide Web Conference, 2009
- [Zhuang 2016] Yong Zhuang and Osman Yagan: Information Propagation in Clustered Multilayer Networks, IEEE Transactions on Network Science and Engineering, IEEE, 2016
- [Huang 2006] Liang Huang, Kwangho Park and Ying-Cheng Lai: Information propagation on modular networks, Phys. Rev. E, Phys, Physical Review, 2006
- [Zhang 2011] Zijie Zhang, Guoqiang Mao and Brian D. O. Anderson: On the Information Propagation Process in Mobile Vehicular Ad Hoc Networks, IEEE Transactions on Vehicular Technology, IEEE, 2011
- [Voelkl 2010] Bernhard Voelkl and Ronald Noë: Simulation of information propagation in real-life primate networks: longevity, fecundity, fidelity, Behavioral Ecology and Sociobiology, Springer, 2010
- [竹内 2005] 竹内 亨, 寺西 裕一, 春本 要, 下條 真司: ソーシャルネットワークを活用した情報伝播モデルに基づいた社内コミュニケーション支援システム, 電気学会論文誌 C, 電気学会, 2005
- [Zhao 2012] Jichang Zhao, Junjie Wu, ZXu Feng, Hui Xiong, Ke Xu: Information propagation in online social networks: a tie-strength perspective, Knowledge and Information Systems, Springer, 2012