

修士論文

ソーシャルネットワークにおける情報拡散力の分析

HUANG,Ziqi

主指導教員 林 幸雄

北陸先端科学技術大学院大学
先端科学技術研究科
(知識科学)

令和2年2月

Abstract

In recent years, because of the development of Social Network Service(SNS), individuals can spread information and own overwhelming power of influence. The purpose of the influencer marketing is to spread information to more people.

Among SNS users, more than 50% of SNS users perform "Spreading" of sharing information with acquaintances, and about 17% of SNS users are performed almost every day. Therefore, how to choose the most appropriate influencer is one of the important topic to many company.

The current choice of influencers is mainly focused on the attributes that correspond to the direct impact. However, even the nodes with larger degree (hubs) , it may not be at the center of the network. Therefore, it is considered necessary to compare the information spreading capabilities of seed notes based on several centrality index.

The centrality index indicates the degree to which a certain node is located at the center of the network. For different centrality indicators, the meaning of network centrality also changes.

About the information diffusion Models, Independent cascade (IC) models and linear threshold (LT) models are often used as basic information diffusion models based on probabilities. SIR model is a special case with IC model,which used for diffusion simulation of seed nodes.

We use the message propagation method to simulation, that does not require sampling of the stochastic operation in the conventional SIR model, and can greatly reduce the calculation time for equivalent operation. And we analyze how each centrality index affects information diffusion using this simulation.

目次

第1章	はじめに	1
1.1	研究背景	1
1.2	研究目的	2
1.3	論文構成	3
第2章	拡散モデルと中心性指標	4
2.1	情報拡散モデル	4
2.1.1	LTモデル (Linear Threshold Model)	4
2.1.2	ICモデル (Independent Cascade Model)	5
2.2	SIRモデル	6
2.3	中心性指標	7
2.3.1	Degree	7
2.3.2	K-core	7
2.3.3	CI (Collective Influence)	8
2.3.4	q_i : FVSの候補確率	8
第3章	メッセージ伝搬式	9
3.1	データセット	9
3.2	従来の確率動作に基づく SIRモデルによる拡散の100回平均値	10
3.3	メッセージ伝搬式について	12
3.4	メッセージ伝搬式の検証	13
第4章	メッセージ伝搬式による拡散結果の比較	21
4.1	1つの種ノードの拡散	21
4.2	複数の種ノードの拡散	25
4.2.1	提案手法	25
4.2.2	3つの種の拡散結果	26
4.2.3	5つの種の拡散結果	28
第5章	おわりに	29

目 次

1.1	ハブが中心に外れる場合	2
2.1	LT モデル	4
2.2	IC モデル	5
2.3	SIR 感染モデル	6
2.4	SIR モデルの曲線	6
2.5	Degree	7
2.6	K-core	7
2.7	CI	8
2.8	qi : FVS の候補確率	8
3.1	email 100 回平均	10
3.2	UCI 100 回平均	11
3.3	メッセージ伝搬による同期処理	12
3.4	email 100 回平均と伝搬の比較, $\beta = 0.1$	13
3.5	email 100 回平均と伝搬の比較, $\beta = 0.3$	14
3.6	email 100 回平均と伝搬の比較, $\beta = 0.6$	15
3.7	UCI 100 回平均と伝搬の比較, $\beta = 0.1$	16
3.8	UCI 100 回平均と伝搬の比較, $\beta = 0.3$	17
3.9	UCI 100 回平均と伝搬の比較, $\beta = 0.6$	18
3.10	Linux 100 回平均と伝搬の比較, $\beta = 0.3$	19
3.11	Linux 100 回平均と伝搬の比較, $\beta = 0.6$	20
4.1	email 伝搬による指標の比較	22
4.2	UCI 伝搬による指標の比較	23
4.3	linux 伝搬による指標の比較	24
4.4	UCI ネットワーク, $\beta = 0.01$	26
4.5	email ネットワーク, $\beta = 0.01$	27
4.6	UCI $\beta = 0.005$	28

表 目 次

1.1	SNS での情報拡散経験の割合 (%) [2]	1
4.1	email $\beta = 0.1$	22
4.2	email $\beta = 0.6$	22
4.3	UCI $\beta = 0.1$	23
4.4	UCI $\beta = 0.6$	23
4.5	linux $\beta = 0.1$	24
4.6	linux $\beta = 0.6$	24

第1章 はじめに

1.1 研究背景

近年、SNSの普及によって、個人でもコンテンツを幅広く拡散でき、圧倒的な情報リーチ力を持つ者も存在するようになり、彼らの拡散力をマーケティングに活用するインフルエンサー・マーケティングが注目されている。

SNSにおける情報拡散力の強い「インフルエンサー」と呼ばれる特別な消費者から「効率よく一般消費者にクチコミが普及する」というのがインフルエンサー・マーケティングの基本的な考え方である[1]。インフルエンサーたちが動画を撮って、ブログ記事をアップすることから、つながっている人々に影響を及ぼすことで、多くの人に情報を伝えるのがインフルエンサー・マーケティングの目標である。

総務省の平成27年の調査研究[2]によると、SNS利用者の中で、自らの情報発信より、情報を知人と共有する「拡散」は、SNS利用者の5割以上が実施しており、約17%はほぼ毎日実施している。また、年代別から見ると、表1.1で示すように年代を問わず活発な情報拡散が行われていることがわかる。

表 1.1: SNSでの情報拡散経験の割合(%) [2]

	ほぼ毎日	週1~2回程度	週1~2回未満	ない	人数
全体	17.1	13.6	24.6	44.7	1178
20代以下	21.7	17.4	22.0	38.8	327
30代	16.5	11.4	23.1	49.1	273
40代	13.3	12.5	26.6	47.7	256
50代	18.7	13.6	24.2	43.4	198
60代以上	11.3	10.5	31.5	46.8	124

そこで、企業にとって、最適なインフルエンサーの選び方が一つの重要な課題となる。現在のインフルエンサー選定には、主にネットワーク的な伝搬力というより、直接的な影響力に相当する属性だけを重視している。たとえば、通常SNSでインフルエンサーの影響力をスコアリングする時には、被リツイート数、いいね数、フォロワー数、メンション数などから算出する。

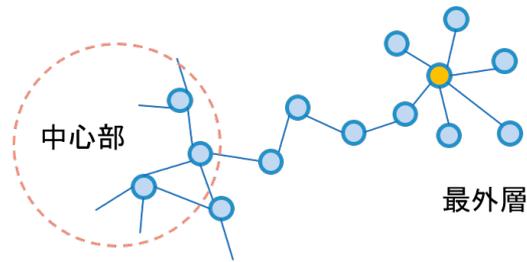


図 1.1: ハブが中心に外れる場合

しかしながら、図 1.1 に示す状況のように、次数が大きいノード（ハブ）であっても、ネットワークの中心から外れる可能性がある。そこで、いくつかの代表的な中心性に従った種ノードからの情報拡散力を比較する必要があると考えられる。

本論文は、ソーシャルネットワークにおける異なる中心性に従った種ノードからの情報拡散力を比較分析した。

1.2 研究目的

ソーシャルネットワークにおけるインフルエンサーの選定には、異なる中心性指標から抽出するノードも異なる。それぞれの指標からの拡散結果を比較して、抽出したインフルエンサーの拡散の特徴を明らかにするのが必要だと考えられる。

本研究は、後述する口コミ拡散モデルの SIR モデルに基づいて、メッセージ伝搬式の利用可能性を調べる。また、数多くの中心性の測定指標から、Degree、K-core[4]、CI (Collective Influence[5]) と q_i : FVS の候補確率 [6] の 4 つの中心性を測る指標を利用して、メッセージ伝搬式による各ノードの拡散力を求め、比較する。さらに一つのインフルエンサーに限らず、複数のインフルエンサーの拡散力およびお互いの位置と合わせて分析し、新たなインフルエンサー抽出法についても検討する。

1.3 論文構成

本論文の構成を以下に示す.

第2章では、SNSでの情報拡散モデルのLTモデルとICモデルを説明し、また、4つの中心性を測る指標について説明する。

第3章では、メッセージ伝搬式について説明し、情報拡散を複数回した平均の結果と比較し、メッセージ伝搬式を利用する可能性を調べる。

第4章では、Degree、K-core、CIと q_i の4つの指標を用いた種ノードに対しての拡散結果を比較する。また、複数の種ノードとして拡散する場合についても検討する。

第5章では、論文の結果を整理する。

第2章 拡散モデルと中心性指標

本章において、2.1節ではSNSにおける情報の基本的な拡散モデルとしてLTモデルとICモデルについて紹介する。2.2節ではICモデルの特殊な場合であるSIRモデルを紹介する。2.3節では本研究で扱うDegree、K-core、CI(Collective Influence)と q_i : FVSの候補確率の4つの中心性を測る指標について手短かに述べる。

2.1 情報拡散モデル

一般に、情報拡散には、主に受信者主導の拡散モデルと、発信者主導の拡散モデルの2種類が考えられている。

受信者主導の拡散とは、人が大量の情報を入手して、決定を下すような状況を想定している。たとえば、ネット上で買い物時に、口コミをみて注文を決めることである。逆に、発信者主導の拡散とは、自分が発信して、他人に影響を与えるような状況を想定している。たとえば、インスタ映えのような情報発信のことである。以下に、この2種類のモデルを紹介する。

2.1.1 LTモデル (Linear Threshold Model)

LTモデルは、受信者主導モデルである。[7]

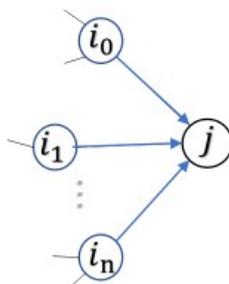


図 2.1: LTモデル

LTモデルでは以下の手順に従って拡散する。

1. 各リンク (i, j) は重み $q_{(i,j)} (> 0)$ をもつ。
2. 各ノード $j \in N$ は閾値 θ_j をもつ。
3. 各時刻 t で、すべての親ノード $i_0 \dots i_n$ からのリンクの重みの総和が θ_i 以上であれば、 j に伝える。

2.1.2 ICモデル (Independent Cascade Model)

ICモデルでは、発信者主導モデルである。[8]

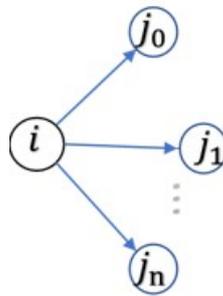


図 2.2: ICモデル

ICモデルは以下の手順に従って拡散する。

1. 各リンク (i, j) は拡散確率 $p_{(i,j)} (> 0)$ をもつ。
2. 各時刻 t で、 i はそれぞれの子ノード $j_0 \dots j_n$ に確率 $p_{(i,j)}$ で伝える。

2.2 SIR モデル

SIR モデルは IC モデルのある特殊な場合に相当する。2.1 節で述べたように、IC モデルでは、ネットワーク上のすべてのリンクはそれぞれの拡散確率 $p(i, j) (> 0)$ を持つが、SIR モデルでは固定的な拡散確率 β を持つ。SIR モデルは IC モデルを単純化した拡散モデルと見られる。

ここで、1.1 節の表 1.1 に示したように、SNS における情報拡散の経験についての総務省の調査では、全ての年代で 10~20% が情報拡散をほぼ毎日行っていることが分かっているが、ツイートやメール等で 1 日に数十件以上はやり取りがあり、それらの全てを拡散する訳ではないと考えられることから、感染率は少なくとも 10% 以下であると推測できる。よって本研究では、感染率 β を 0.1 以下に設定したシミュレーション実験を主に行っている。

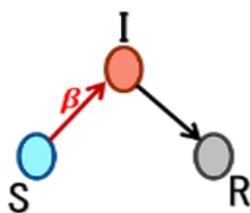


図 2.3: SIR 感染モデル

SIR モデルでは以下の手順に従って拡散する。

1. 感染者 I との接触により、感受性保持者 S が感染率 β で感染者 I となる。
2. 次の時刻で、確率 1 で感染者 I が回復し、免疫保持者 R となる。

典型的な SIR モデルの曲線として、時間に対する状態数の変化を以下に示す。

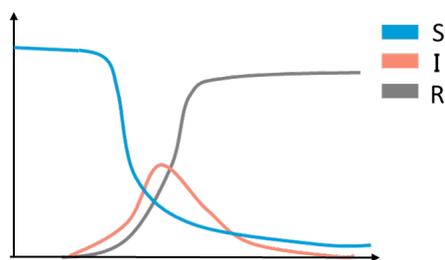


図 2.4: SIR モデルの曲線

図 2.4 のように、S 状態は単調減少で、回復の R 状態は単調増加、また感染した I 状態は山型の変化が SIR モデルに共通した特徴として見られる。

2.3 中心性指標

中心性指標は、あるノードがネットワークの中心に位置する度合を表す。但し、以下の代表的な異なる中心性指標において、ネットワークの中心性の意味も変わる。

2.3.1 Degree

ノードの次数を指す。次数とは、一つのノードが他のノードとつながっているリンクの数のことである。

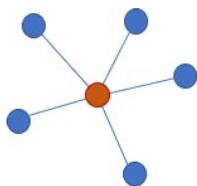


図 2.5: Degree

1つのエッジが両端の2つのノードに接続するため、エッジの片側の次数は1だけを増加するが、もう一方の次数も1だけ増加する。そのため、ネットワークの次数の総数はエッジ数の2倍になる。

2.3.2 K-core

ネットワーク内で緊密に連結された核として、K-core は部分グラフの中で、少なくとも K 個の他のノードと互いに接続されているものを指す。K-core は階層的に生成され、K 値の大きいコアは小さいコア（たとえば、(k-1)-core）に含まれる。K 値が大きいほど、コア内の連結度が高い。

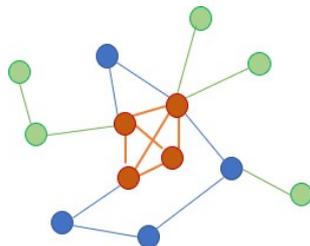


図 2.6: K-core

次数が1のノードを再帰的に削除することにより、ネットワークの K-core 分解ができる。

図 2.6 においては、赤い部分は2コア、青いノードは1コア、緑色ノードは0コアに属する。K-core を把握することによって、ネットワーク内の核ともいえる結束の強いノード集合を特定することができる。

2.3.3 CI (Collective Influence)

CI 値は、以下の式で定義され、ネットワークが分断される最小の除去ノード集合から導出される。[5]

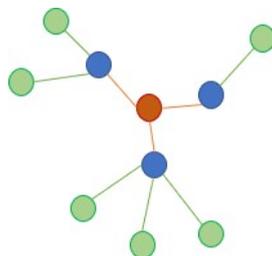


図 2.7: CI

$$CI_l(i) \stackrel{\text{def}}{=} (k_i - 1) \sum_{j \in \partial \text{Ball}(i,l)} (k_j - 1)$$

その直観的な意味としては、

$CI_l(i)$ は i から l ホップ先に拡散するリンク数の和に比例した影響力に相当する。

$\partial \text{Ball}(i, l)$ は i から l ホップ先のノード集合である。

2.3.4 q_i : FVS の候補確率

q_i 値は、FVS (feedback vertex set) の候補としての確率を示す [6]。FVS は、ネットワークからループを無くすための最小の除去ノード集合である。

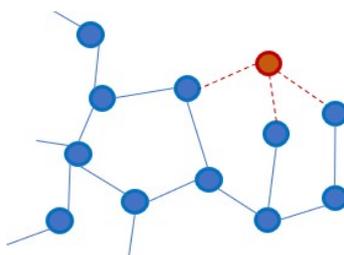


図 2.8: q_i : FVS の候補確率

q_i 値が 1 であれば FVS に属する。逆に、0 であれば FVS でない。たとえば、図 2.8 の赤いノードは q_i 値の高いノードである。

第3章 メッセージ伝搬式

本章において、3.1節は用いたデータセットについて述べる。3.2節では従来の確率動作に基づく SIR モデルによる拡散を 100 回試行した平均値を示す。3.3 節ではメッセージ伝搬式について説明する。3.4 節では結果の比較で、メッセージ伝搬式が先の 100 回平均と等価であることを数値的に検証する。

3.1 データセット

以下は、本論文で利用した 3 つの現実ネットワークである。[10][11][12]

- email (URV email) : ノード総数 $N = 1133$ 、リンク総数 $M = 5451$

電子メールネットワーク URV、Rovira i Virgili 大学 (Tarragona Sapin) のメンバー間の電子メール交換のネットワークである。

- UCI (UC Irvine online community) : ノード総数 $N = 1893$ 、リンク総数 $M = 13835$

このネットワークには、カリフォルニア大学アーバイン校の学生のオンラインコミュニティのユーザー間で送信されたメッセージが含まれている。ノードはユーザーを表す。エッジは、送信されたメッセージを表す。

- Linux (Linux mailing list) : ノード総数 $N = 24567$ 、リンク総数 $M = 158164$

これは、Linux カーネルメーリングリストの通信ネットワークである。ノードは個人 (電子メールアドレスで識別) であり、各エッジはユーザーから別のユーザーへの応答を表す。

3.2 従来の確率動作に基づく SIR モデルによる拡散の 100 回平均値

本節では、予備実験結果として、従来の SIR モデルにおける確率動作を確認するために 100 試行の平均値を示す。

ここで、email ネットワークを例として、感染率 β を 0.1、0.3、0.4、0.6 に設定した実験を行う。

図 3.1 に email ネットワークにおいて、種ノードをそれぞれの中心性指標 (deg、kc、CI、qi) が最大値となるノードを選択して、その 100 回試行を平均化した状態数の時間変化を示す。

図を見やすくするため、SIR モデルの I と R の曲線だけを表示する。また、横軸は時間で、縦軸は感染の割合である。左上の図は感染率 β が 10%、右上は感染率 β が 30%、左下は 40%、右下は 60% である。

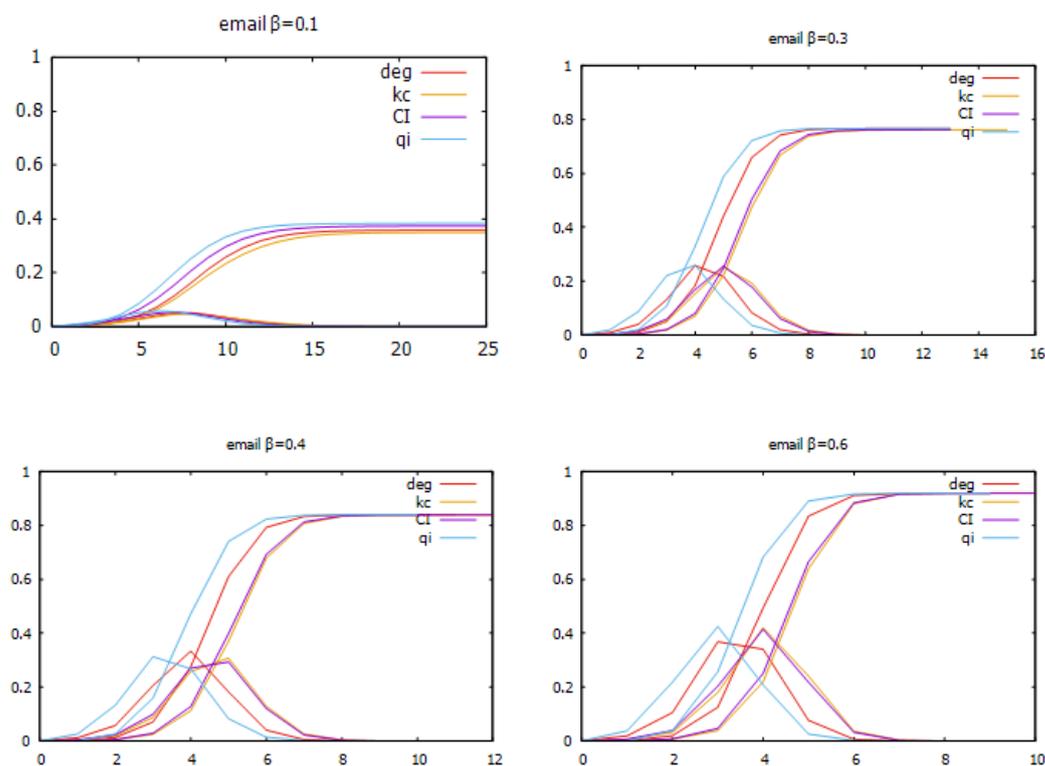


図 3.1: email 100 回平均

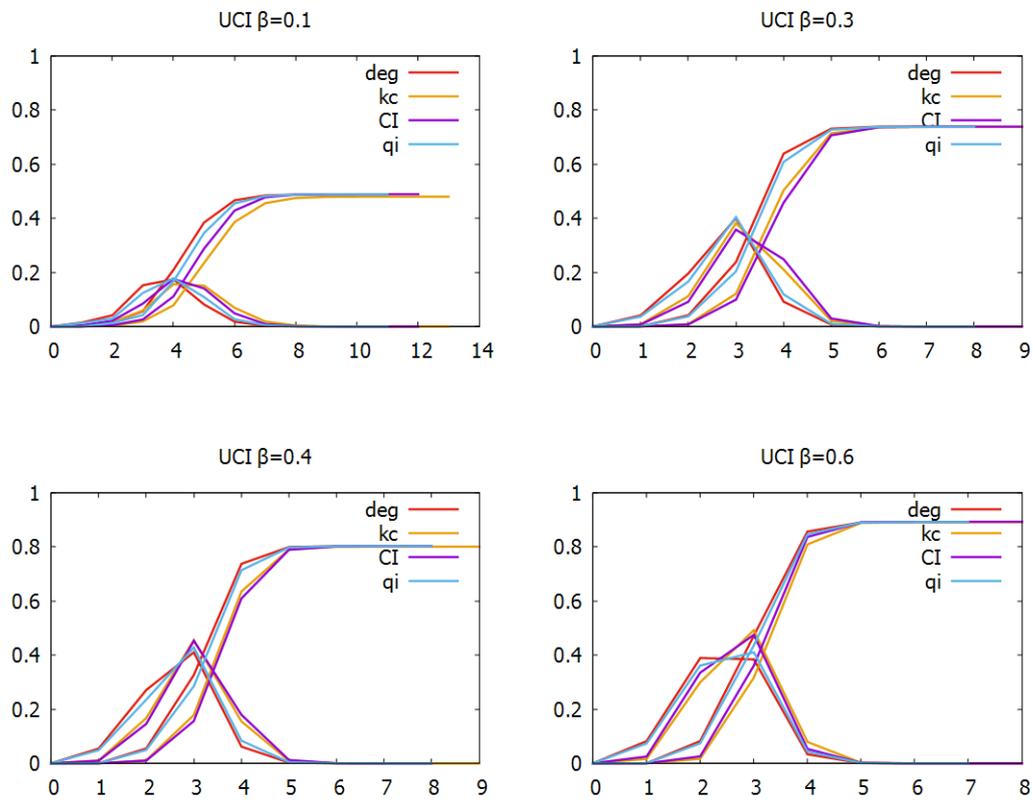


図 3.2: UCI 100 回平均

図 3.1 と 3.2 から、異なる感染率 β により、R 曲線での最終的な収束状態数の割合には差がある。 β が大きいほど、状態数 R の割合が高いことがわかる。また、同じ感染率 β の場合、4 つの指標の最終的な収束状態数 R の割合がほぼ同じである。それぞれの指標に対して、I 曲線のピーク時刻 p^* に差が出ることがわかる。

3.3 メッセージ伝搬式について

本節では、メッセージ伝搬式 [9] について説明する。メッセージ伝搬式は、前節で述べた従来の SIR モデルにおける確率動作のサンプル化が不要で、(サンプル数を無限に大きくした場合と) 等価な動作に対して計算時間を大幅に短縮できる方法である。

あるノード i に対して、各時刻 t で、隣接ノードから確率的に情報伝搬する。図 3.3 の状況を考える。

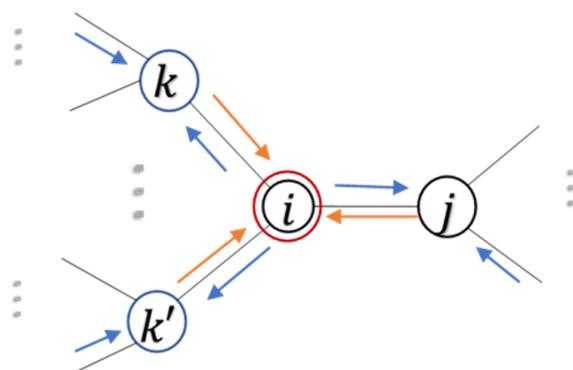


図 3.3: メッセージ伝搬による同期処理

このとき、リンク $i \rightarrow j$ で情報伝搬する確率は下の式に従う：

$$v_{i \rightarrow j} = n_i [1 - \prod_{k \in \partial i \setminus j} (1 - v_{k \rightarrow i})]$$

上式を、感染率 β を持つ SIR モデルに拡張すると、

時刻 $t + 1$ で、あるノード i が感染者 I となる感染確率 $P_I^i(t + 1)$ は以下のように定義される。

ノード $i \rightarrow j$ で情報伝搬する確率:

- $P_I^i(t + 1) = P_S^i(t) * [1 - \prod_{k \in \partial i} (1 - \beta * P_I^k(t))]$
- $P_R^i(t + 1) = P_R^i(t) + P_I^i(t)$
- $P_S^i(t + 1) = 1 - P_I^i(t + 1) - P_R^i(t + 1)$

$t + 1$ 時刻におけるノード i の感染確率 $P_I^i(t + 1)$ は、 t 時刻で未感染の S 状態で感染できる確率 $P_S^i(t)$ と、I 状態の隣接ノード $k \in \partial i$ のうち少なくとも 1 つは伝搬する確率との積となる。また、前の時刻で感染したノード i が次の時刻で免疫保持の R 状態になることから、R 状態数の割合 $P_R^i(t + 1)$ は、 t 時刻で I と R の総和となる。そして、正規化条件は、(3 状態の確率和が 1 であること) より求める。感染の I と R をひいた S 状態数の割合 $P_S^i(t + 1)$ 。

3.4 メッセージ伝搬式の検証

email ネットワークの感染率 β が 0.1、0.3、0.6 の場合、UCI ネットワークの β が 0.1、0.3、0.6 の場合、また Linux ネットワークの感染率 β が 0.3、0.6 の場合の比較結果を表す。図 3.3 から図 3.6 は、email ネットワークにおいて、2.3 で述べた 4 つの指標を用いてそれぞれの種ノードを選択し、100 回試行を平均化した状態数の割合と、メッセージ伝搬式による状態数の割合を時間変化を比較。横軸は時間で、縦軸は感染の割合である。また、左上、右上、左下、右下の順に Degree、K-core、CI、 q_i : FVS の候補確率を示す。種ノードは、それぞれの中心性指標 (deg、kc、CI、 q_i) が最大値となるノードを選定した。最大値が複数ある場合、ランダムに 1 つ選択する。また CI は 3 ホップ先の値で行う。

ここで、AVG-100 は 100 回平均の曲線で、mp はメッセージ伝搬式の曲線であるが、原理的に等価な動作となる両者はほぼ一致することを数値的に確認している。

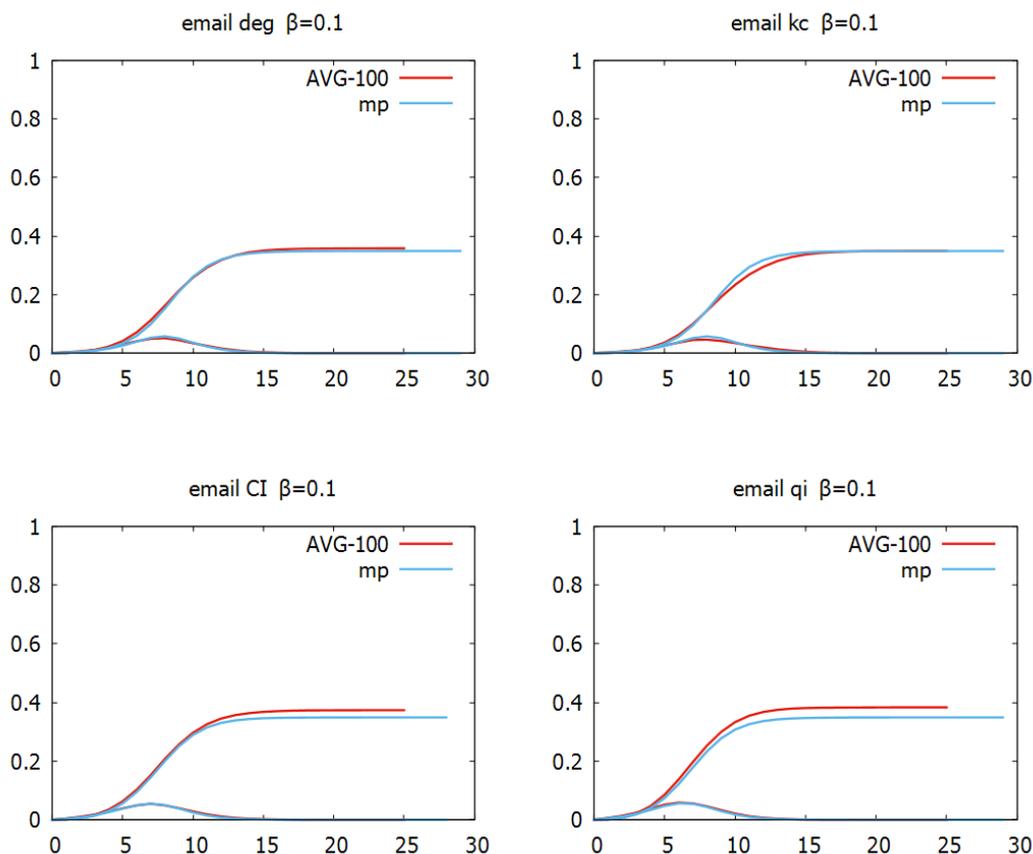


図 3.4: email 100 回平均と伝搬の比較, $\beta = 0.1$

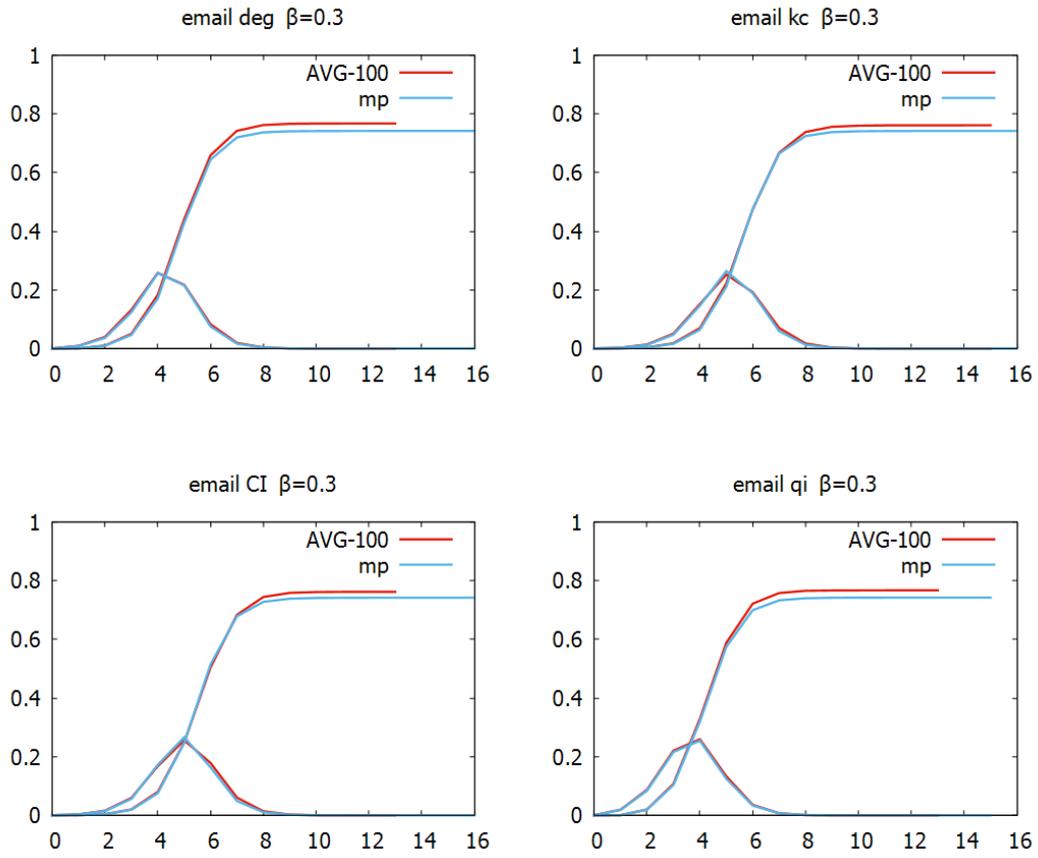


図 3.5: email 100 回平均と伝搬の比較, $\beta = 0.3$

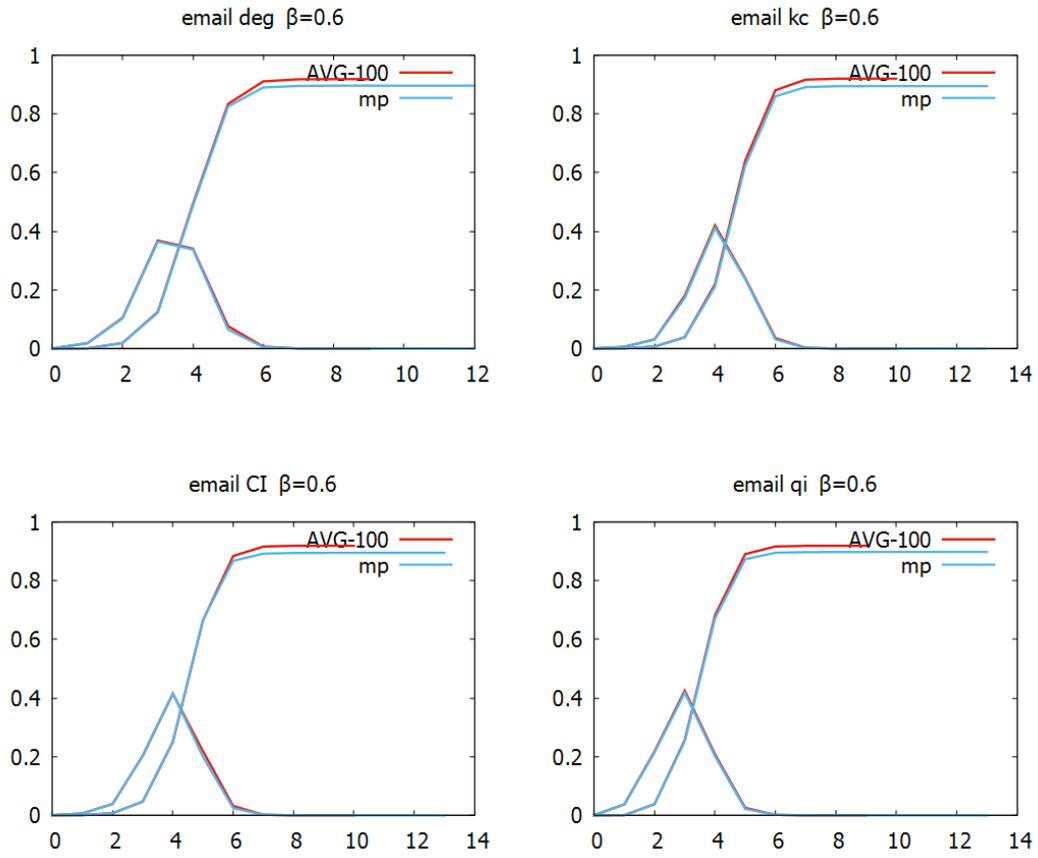


図 3.6: email 100 回平均と伝搬の比較, $\beta = 0.6$

同様に UCI と Linux についても、図 3.7 から図 3.11 に 4 つの指標 (deg、kc、CI、qi) を用いてそれぞれの種ノードを選択し、最大値となるノードを種ノードとして比較実験を行った。以下同様に、横軸は時間で、縦軸は感染の割合である。また、左上、右上、左下、右下の順に Degree、K-core、CI、qi : FVS の候補確率を示す。

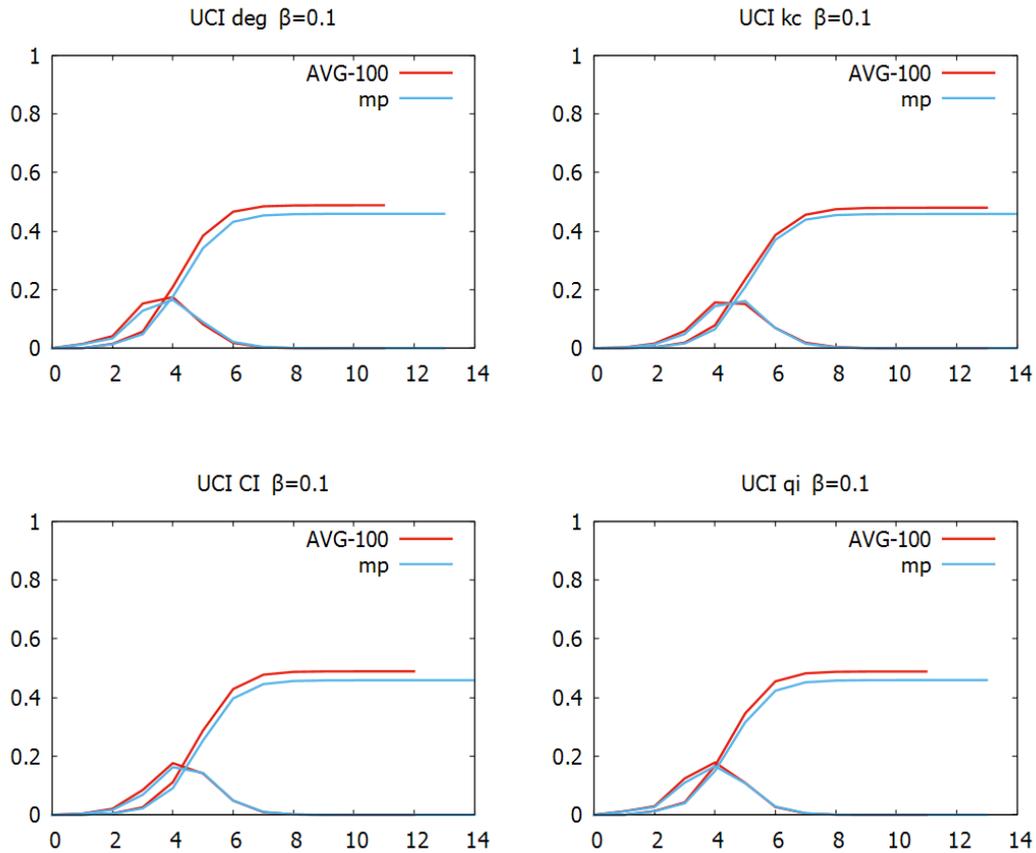


図 3.7: UCI 100 回平均と伝搬の比較, $\beta = 0.1$

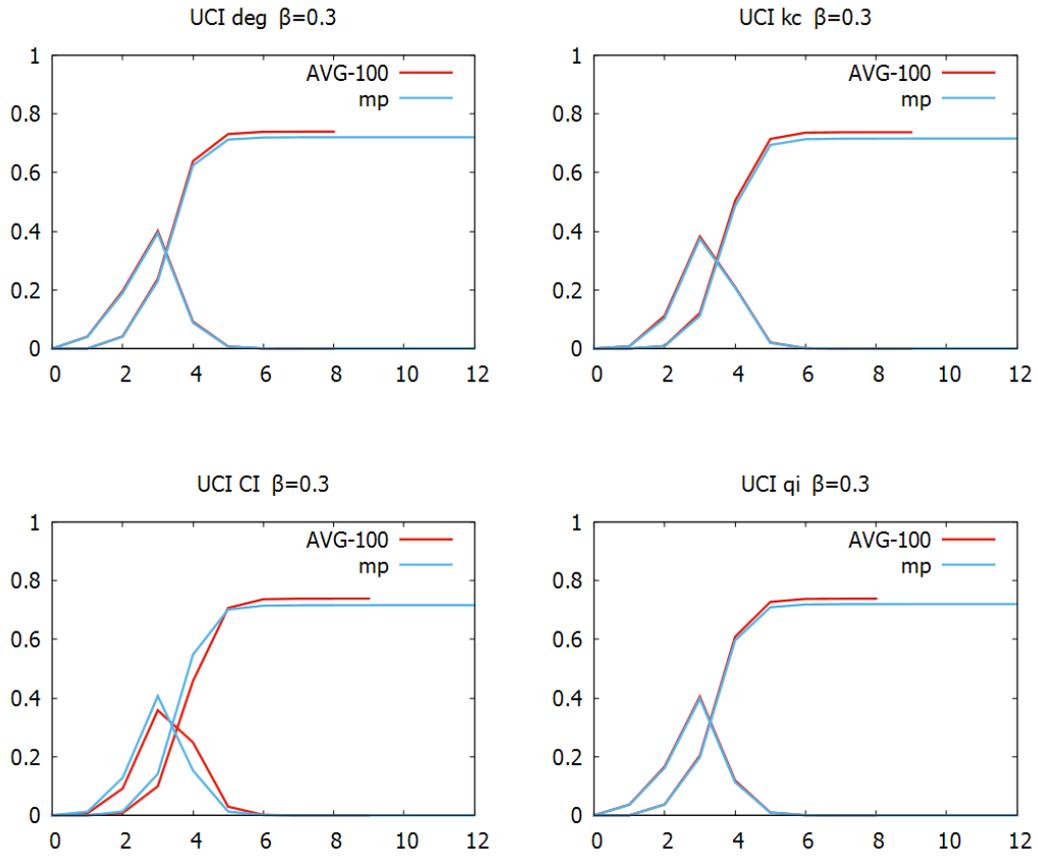


図 3.8: UCI 100 回平均と伝搬の比較, $\beta = 0.3$

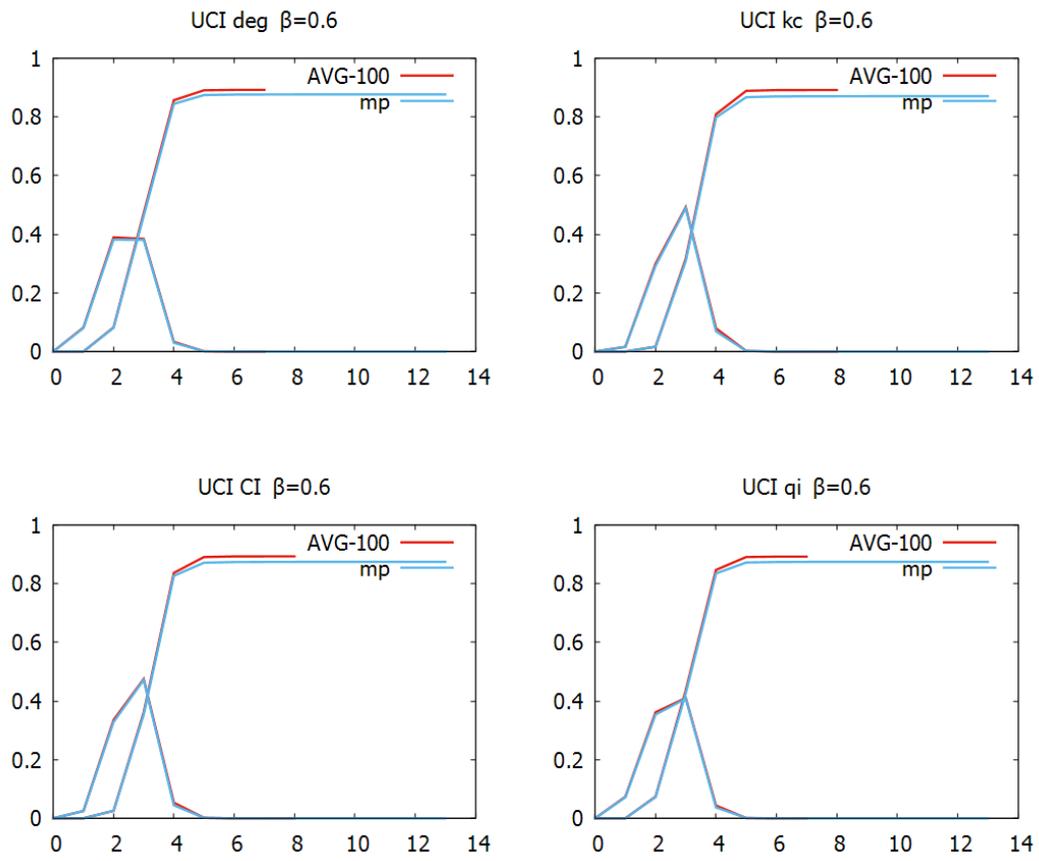


図 3.9: UCI 100 回平均と伝搬の比較, $\beta = 0.6$

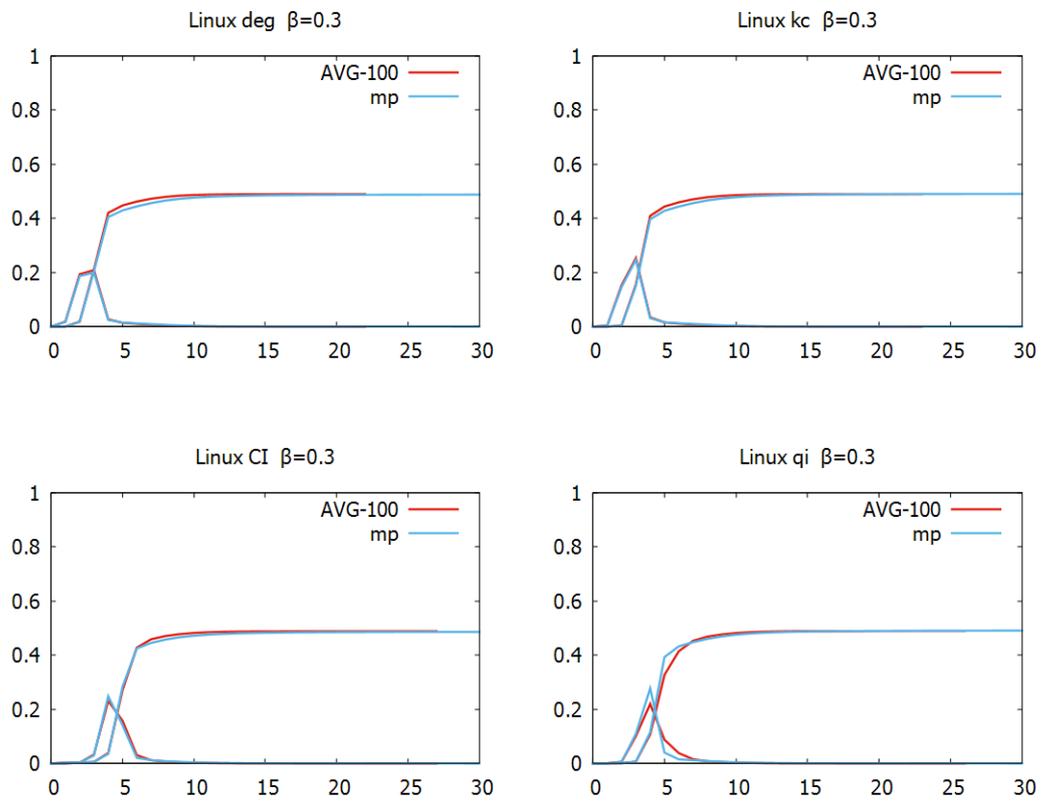


図 3.10: Linux 100 回平均と伝搬の比較, $\beta = 0.3$

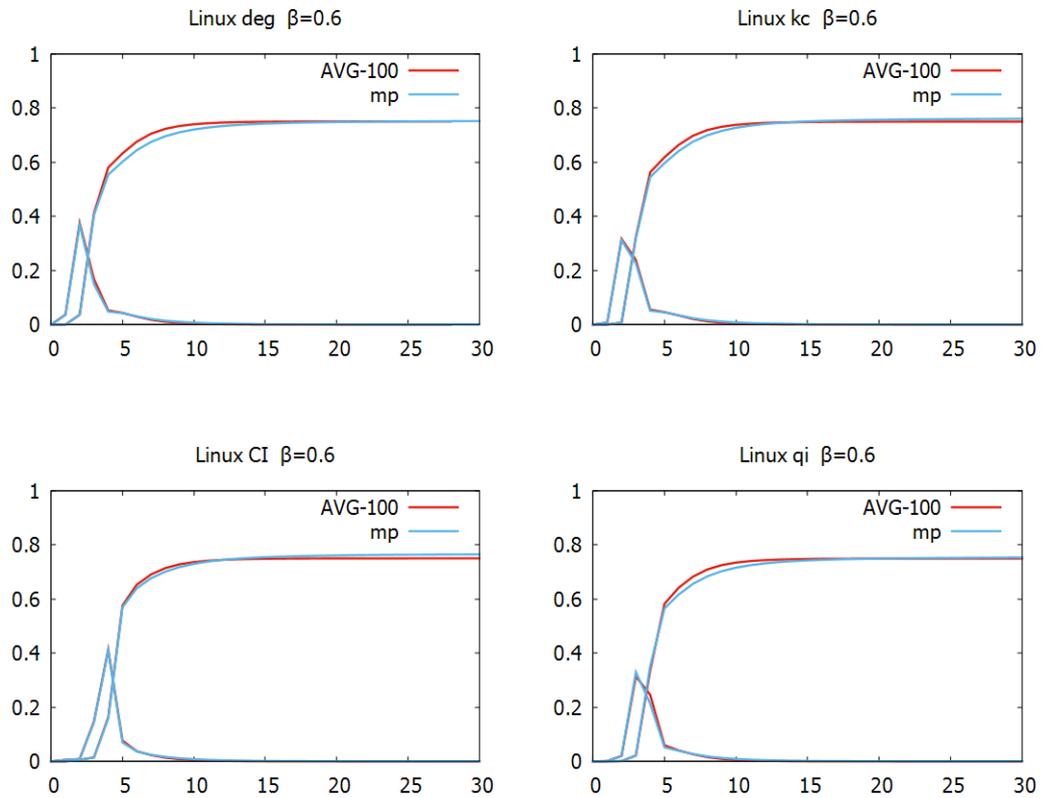


図 3.11: Linux 100 回平均と伝搬の比較, $\beta = 0.6$

図 3.3 から図 3.11 に共通した結果として、異なる指標と指標値に関係なく、100 回試行を平均化した R 状態数の割合とメッセージ伝搬式による R 状態数はほぼ一致することが分かる。若干差が出てるのは、100 回平均が未だ十分とは言えないと考えられる。

第4章 メッセージ伝搬式による拡散 結果の比較

本章において、4.1節ではSIRモデルによるメッセージ伝搬式に従って、Degree、K-core、CIと q_i の4つの指標を用いた種ノードに対しての拡散結果を比較する。また、4.2節では、複数の種ノードとして拡散する場合について検討した。

4.1 1つの種ノードの拡散

図4.1は、emailネットワークにおいて、4つの指標を用いてそれぞれの最大値である種ノードを選択し、メッセージ伝搬式による状態数時間変化の結果を比較する。左側の図は感染率 β が10%の結果で、右側は60%を示す。ここで、種ノードは、それぞれの指標の最大値であるノードで、最大値が複数ある場合には、ランダムで1つを選定する。CIは3ホップ先の値で行う。

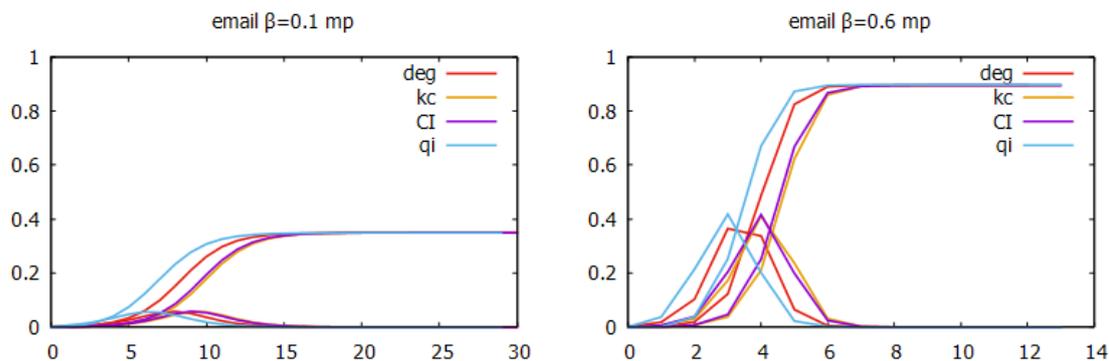


図 4.1: email 伝搬による指標の比較

表 4.1: email $\beta = 0.1$

	p^*	T	R_∞
Degree	8	29	0.349064
K-core	9	30	0.34926
CI	9	30	0.349248
qi	6	29	0.348952

表 4.2: email $\beta = 0.6$

	p^*	T	R_∞
Degree	3	12	0.89603
K-core	4	13	0.894467
CI	4	13	0.894597
qi	3	13	0.897426

図 4.1 から図 4.3 は、感染率 β が 0.1 と 0.6 の場合における、email ネットワーク、UCI ネットワーク、Linux ネットワークにおける 4 つ指標の比較結果を表す。図表の結果から見ると、収束した R 状態数の割合 R_∞ のちに若干差があるものの、図 4.1 の最終的な R 値がほぼ一致することから、異なる指標が最終的な拡散効果に影響が少ないことがわかる。但し、R 曲線の収束時刻 T と I 曲線のピーク時刻 p^* に差があることから、異なる指標が拡散の速さに影響があることは確認できる。

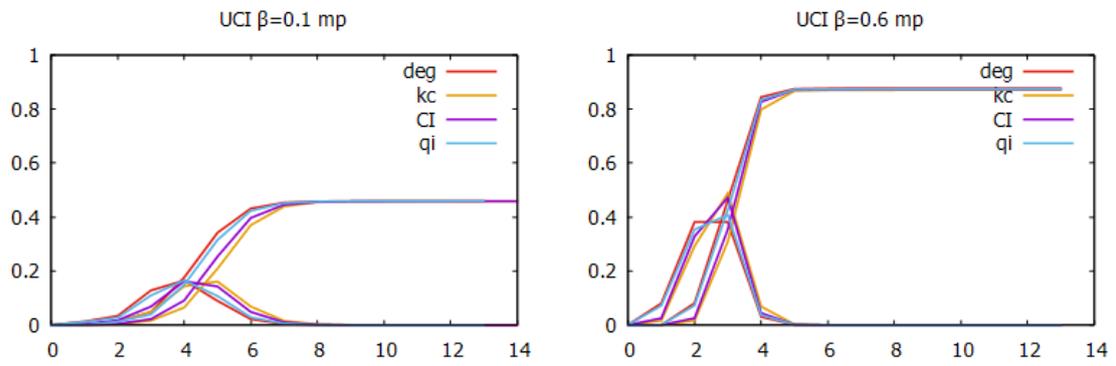


図 4.2: UCI 伝搬による指標の比較

表 4.3: UCI $\beta = 0.1$

	p^*	T	R_∞
Degree	4	13	0.459219
K-core	5	14	0.45868
CI	4	14	0.458665
qi	4	13	0.459198

表 4.4: UCI $\beta = 0.6$

	p^*	T	R_∞
Degree	2	13	0.876398
K-core	3	13	0.870315
CI	3	13	0.873851
qi	3	13	0.873947

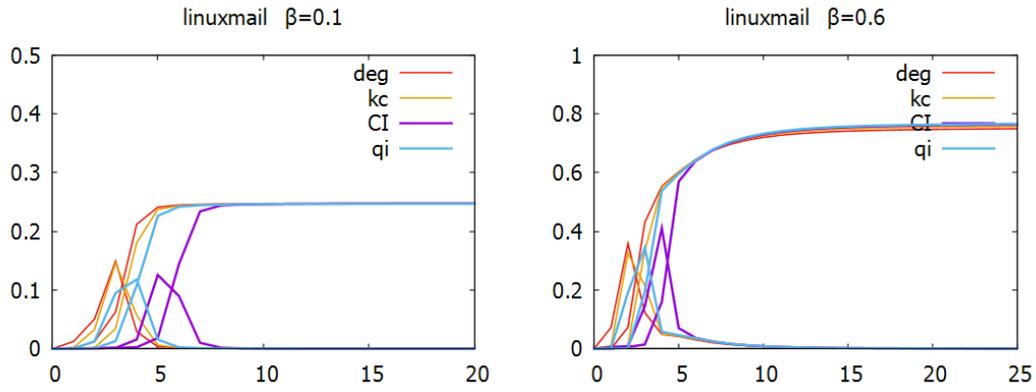


図 4.3: linux 伝搬による指標の比較

表 4.5: linux $\beta = 0.1$

	p^*	T	R_∞
Degree	3	23	0.247603
K-core	3	24	0.246938
CI	5	25	0.246876
qi	4	23	0.246720

表 4.6: linux $\beta = 0.6$

	p^*	T	R_∞
Degree	2	88	0.752208
K-core	2	88	0.760173
CI	4	89	0.766913
qi	3	88	0.768891

4.2 複数の種ノードの拡散

現実的なインフルエンサー・マーケティングは通常、複数のインフルエンサーを選定して、同時に情報の拡散を行う。以下、複数の種ノードを抽出するには、ノードの属性だけでなく距離も考えることを提案する。

4.2.1 提案手法

ノードの抽出は以下のように行う。

1. ノード間の距離を考慮しない複数種ノードの抽出：

それぞれの指標により、最大値 K_A であるノードから、降順でノードを選択して $i_A, i_B, i_C...$ とする。

2. ノード間の距離を考慮して複数種ノードの抽出：

Step 1: それぞれの指標により、最大値 K_A である1つの種ノード i_A を選択する。最大値ノードが複数ある場合では、ランダムに一つを選択して i_A とする。

Step 2: i_A 以外のすべてのノード i_B から、A までの距離 $l_{B \rightarrow A}$ (ステップ) を計算し、 $l_{B \rightarrow A} \times K_B$ を計算する。積が最大のノードが i_B となる。(複数がある場合では、ランダムに一つを選択する。)

Step 3: i_A と i_B 以外のすべてのノード i_C から、 i_B までの距離 $l_{C \rightarrow B}$ を計算して、 $l_{C \rightarrow A} \times l_{C \rightarrow B} \times K_C$ の積を計算する。積が最大のノードが i_C となる。(複数がある場合では、ランダムに一つを選択する。)

Step 4: (3つ以上の種ノードを抽出する場合、以上同様に計算し、 $i_D, i_E...$ を抽出できる。)

4.2.2 3つの種の拡散結果

図4.3は、email ネットワークにおいて、4つの指標を用いてそれぞれの種ノードを選択して、100回試行を平均化した状態数の割合とメッセージ伝搬式による状態数の時間変化を比較する。以下同様に、横軸は時間で、縦軸は感染の割合で、0から0.01の範囲で表す。また、左上、右上、左下、右下の順に Degree、K-core、CI、 q_i : FVSの候補確率を示す。

2.2節で述べたように、現実的な感染率 β は10%以下であると推測できる。そこで、感染率 β は1%に設定したシミュレーション実験を行っている。N=3は距離を考慮した3つの種、N=3randは距離を考えない、最大順位の3つの種、N=1maxは最大値の1つの種、N=1minは最小値の1つの種とする。

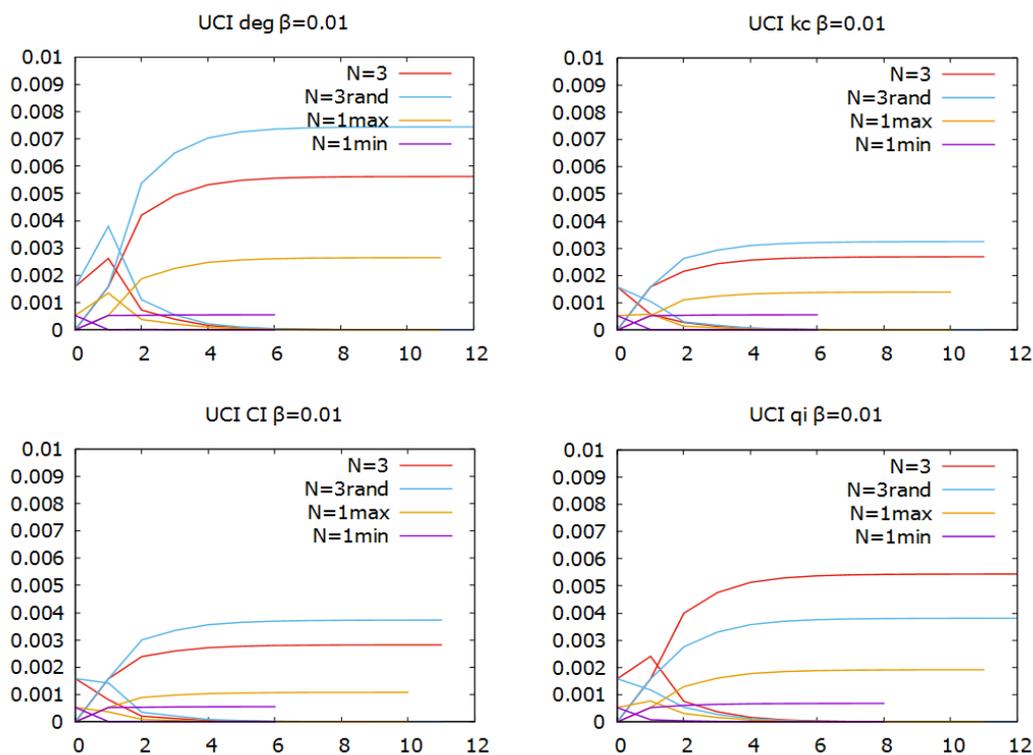


図 4.4: UCI ネットワーク, $\beta = 0.01$

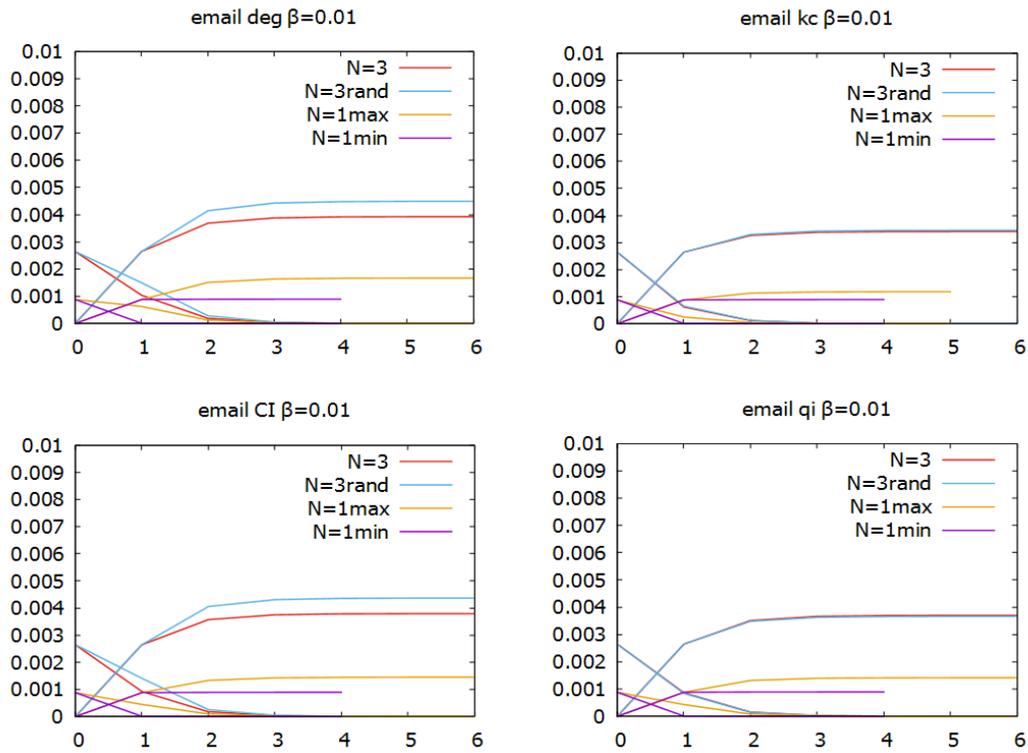


図 4.5: email ネットワーク, $\beta = 0.01$

図 4.3 と図 4.4 は、感染率 β が 0.01 の場合、email ネットワーク、UCI ネットワークの 4 つ指標のそれぞれの比較結果。図表の結果から見ると、距離を考える $N = 3$ の曲線とランダムで選んだ曲線が少し差があるが、UCI の場合、 q_i 以外に、最大値からランダムで選んだ種ノードの拡散範囲（R 状態数の割合）が大きい。email のほうから見ると、両者の差が小さい。距離を考えることから、拡散効果に影響があるとは言えない。

4.2.3 5つの種の拡散結果

現実の情報拡散率が10%より小さいと考え、感染率 β 値を0.005として設定して、5つの種の場合について、K-core、Degree、CIの結果をそれぞれ比較した。

図4.6では、N=5は距離を考慮した5つの種、N=5 rand: 距離を考えない、最大値の中からランダムに選んだ5つの種である。4.2.2節と同じように、両者の差が小さい。距離を考慮することから、拡散効果に影響があるとは言えない。

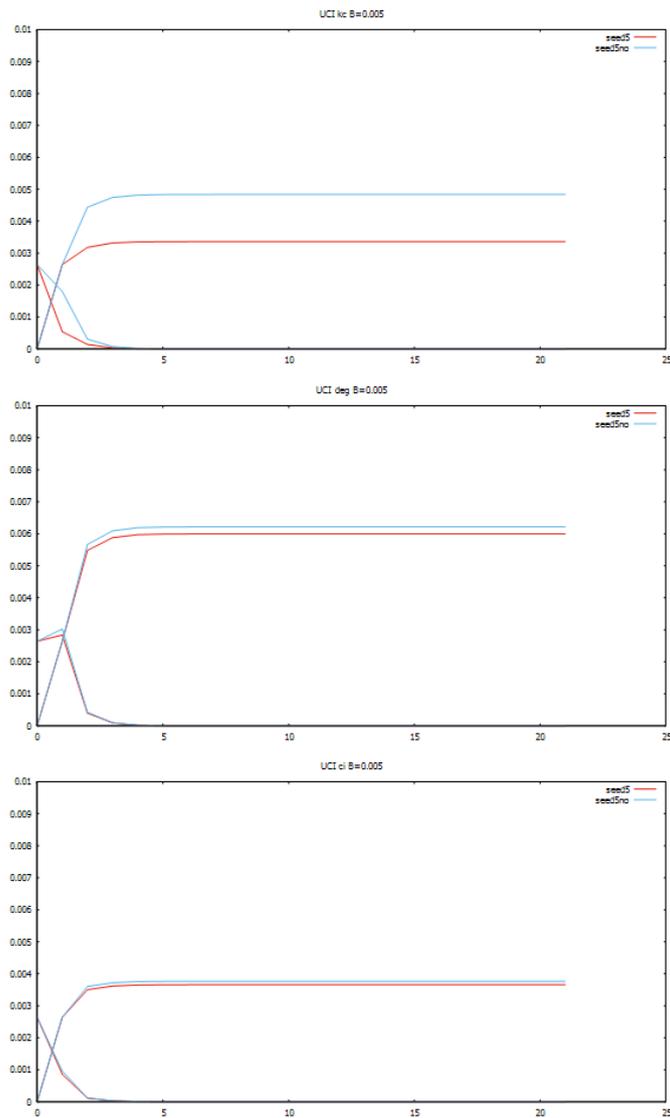


図 4.6: UCI $\beta = 0.005$

第5章 おわりに

3.4 節で口コミ拡散モデルの SIR モデルに基づいて、メッセージ伝搬式の利用可能性を検証した。4.1 節で数多くの中心性の測定指標から、Degree、K-core、CI (Collective Influence) と q_i (FVS) の 4 つの中心性を測る指標を利用して、メッセージ伝搬式によるシミュレーションで比較した。また、4.2 節で、複数種ノードの抽出手法を提案し、一つのインフルエンサーに限らず、複数のインフルエンサーの拡散力およびお互いの位置と合わせて分析した。

本論文では、

- 異なる指標を用いた最大値の拡散には、最終的な拡散範囲 (R 状態数の割合) の差があまり出ないことを確認した。インフルエンサーの選定には、4 つの指標のどちらを利用しても、大体同じ効果が出た。
- 拡散範囲以外、拡散の速さを考慮する場合、指標の選定には意義があると考えられる。
- また、異なるネットワークがそれぞれのノード数、繋がり方、規模の違いから、指標の結果も差が出る。ですので、インフルエンサーの選定には、そのネットワークの繋がり構造の特徴に依存する。

現実的なソーシャルネットワークと比べて、本論文で利用した 3 つのネットワークがノード規模が小さく、今後の課題として、特に 4.2 節の複数種ノードの抽出に対しては、抽出するノードの数と新たな距離の測り方について検討が必要だと考えられる。

参考文献

- [1] インフルエンサーは本当に使えるのか?-なぜあの人のクチコミは影響力があるのか (2) ,”<https://toyokeizai.net/articles/-/37622?page=4>”.
- [2] 総務省「社会課題解決のための新たな ICT サービス・技術への人々の意識に関する調査研究」,2015.
- [3] A.Namtirtha, A.Dutta, , and B.Dutta, Identifying influential spreaders in complex networks based on kshell hybrid method, *Physica A* 499, 310-324, 2018.
- [4] M.Kitsak , L.K.Gallos, S.Havlin, and F. Liljeros, Identification of influential spreaders in complex networks, *Nature Physics* volume 6, 888-893 ,2010.
- [5] F.Morone,and H.A.Makse, Collective Influence Algorithm to find influencers via optimal percolation in massively large social media, *Nature* 524, 65-68, 2015.
- [6] H. Zhou, Spin glass approach to the feedback vertex set problem, *The European Physical Journal B*, 86:455, 2013.
- [7] D.Kempe, J.Kleinberg, and É.Tardos,Maximizing the Spread of Influence through a Social Network,*theoryofcomputing* 11,105-147,2015.
- [8] J.Kleinberg,Cascading Behavior in Networks: Algorithmic and Economic Issues,*Algorithmic Game Theory*, 613-632 ,2007.
- [9] F.Morone, HA.Makse , Influence maximization in complex networks through optimal percolation, *Nature*.Aug 6;524(7563), 65-8, 2015.
- [10] Alex Arenas Website-E-mail network URV, ”<http://deim.urv.cat/~alexandre.arenas/data/welcome.htm>”.
- [11] The Koblenz Network Collection-UC Irvine messages,”<http://konect.uni-koblenz.de/networks/opsahl-ucsocial>”.
- [12] The Koblenz Network Collection-Linux kernel mailing list replies,”<http://konect.uni-koblenz.de/networks/lkml-reply>”.