

ネットワーク成長によるメール型ウイルスの再流行と 重点的なハブの免疫化の効果

林 幸雄[†] 箕浦 正人[†] 松久保 潤[†]

電子メールの送受信数の実測分布に基づく接触関係を考えた、コンピュータウイルスの拡散モデルを提案する。その基本特性として、ウイルスの実行率や発見率に対する、総感染 PC 数（潜伏から増殖状態に移行した数）、総感染メール数（ウイルス感染が発見されたものを含む）、絶滅までの時間をシミュレーション解析する。特に、感染の鎮静化後の再流行現象が新規ユーザの加入によって起こることを、開放系の本モデルは示唆する。経験上典型的なこの現象は、従来モデルでは説明できないものである。さらに、重点的なハブ頂点のアンチウイルスによる免疫化の効果を明らかにする。

Recoverable Prevalences of Viruses via E-mails in a Growing Computer Network and Effectiveness of the Prior Immunization for Hubs

YUKIO HAYASHI,[†] MASATO MINOURA[†] and JUN MATSUKUBO[†]

We propose a model of computer viruses via electronic mails, based on the contacts in real distributions of the sent and receive mails. As the fundamental properties for the rates of execution and detection, we analyze the numbers of total infectious computers (executed from the hidden to the propagatable state), total infected mails (including detected viruses), and the extinction time by simulations. In particular, this model in an open system suggests that a growing computer network with new mail users causes the recoverable prevalences from a low level of disease. This phenomenon is typical in experiences but not explained by the conventional models of computer viruses. Moreover, we verify the effectiveness of the prior immunization by anti-viruses for hubs in the scale-free network.

1. ま え が き

インターネットの急速な普及にともなって、コンピュータウイルスの浸透度、脅威とも年々増大し、ますます深刻な社会問題となっている。なかでも、電子メールによるウイルス感染は被害全体の 70~80%以上で^{(6),(7),(9)}、日常的な情報の交換や共有の際の添付ファイルなどを悪用するため、フロッピなどを介した感染経路とは比べものにならないほど強い感染増殖力を持っている。また、感染によるハードディスクのフォーマット（データの破壊）やファイルの削除などの直接的な被害のみならず、感染の発見や駆除、復旧作業に要する業務停滞、増殖活動による機密情報の漏洩やウイルス散布による社会的信用の低下などの間接的被害をもたらす、その損害額は想像以上に大きい。たとえば、米国における、ウイルス駆除、生産能力の損失、復旧作業による損害額の合計は年間 100~200 億ドル、

1 件のウイルスあたりでは 10~20 億ドルと推定されている。

一方、増え続けるウイルスの中で本当に問題となるのは、わずか数%の 10 種類程度にすぎない⁽⁹⁾ が、これらの時間的な挙動や被害の大きさに関する個々の観測データすら、現状では十分に得られているとはいいがたい（複合的な要因の累計が多い）。そこで、少しでも被害を抑えるためには観測のみならず、ウイルス亜種の系統解析やシミュレーションによる増殖力の予測なども用いて、対策を検討する必要がある。

コンピュータウイルスの対策に関する研究には、個々のウイルスの発見と駆除を目的としたミクロレベルの研究と、ネットワーク上のウイルスの拡散伝搬特性の解明や効果的なネットワークの免疫方法を探るマクロレベルの研究があり⁽²¹⁾、本論文では後者を扱う。

これまで、コンピュータウイルスの拡散モデルにおいて、シミュレーションによるランダムネットワーク

[†] 北陸先端科学技術大学院大学知識科学研究科
School of Knowledge Science, Japan Advanced Institute of Science and Technology

<http://itpro.nikkeibp.co.jp/free/ITPro/USNEWS/20010907/10/>

の直径と拡散速度や定常状態の感染数、完全グラフ上の理論解析による駆除率の限界が報告されている²¹⁾。また、電子メールによる拡散の特徴を考慮したモデルでは、1次元格子上の限定ケースの理論解析と、より一般的な1次元格子上のシミュレーションによる接続数と感染率の関係、アンチウイルスによる拡散抑止効果などが示されている¹⁶⁾。しかしながら、一様ランダムや完全グラフは、均一なウイルスとの接触機会を表すため現実的なモデルとはいいがたく、あくまで比較のための指標を与えているにすぎない。格子モデルには接触機会の局所的な偏りが多少あり、アンチウイルスの配置箇所の重要性が指摘されている¹⁶⁾ものの、現実的なネットワーク構造上の効果的な配置方法は検討されていない。

一方、社会的な知人関係などを表すSF (Scale-Free) と呼ばれるネットワーク構造³⁾の特徴として、絶滅のしきい値が存在しないこと、すなわち、感染率をどんなに小さくしても広がっていくことがSIS (Susceptible-Infected-Susceptible) モデルの平均場解析から最近示され¹⁹⁾、従来の疫学の常識を覆した。また、感染の発見と駆除の後に永続的な免疫を持つ、SF ネットワーク上のSIR (Susceptible-Infected-Recovered) モデルの拡散特性も解析されている^{11), 15)}。さらに、SF ネットワークにおいてはハブがアキレス腱となる(次数が大きいハブ頂点が故障するとネットワークが極度に分断される)こと¹⁾を逆に利用して、重点的なハブの免疫化によって拡散が抑えられることがSISモデルの理論解析から示されている^{4), 20)}。

上記のモデル^{11), 15), 16), 19), 21)}ではネットワークの規模は固定され(あるいは $N \rightarrow \infty$ において)、ウイルスが広く蔓延するか絶滅するか、どちらかの定常状態しか議論されてないが、実際の観測データはこれらとは少し異なる。すなわち、ネットワークは日々成長しているし、少数の感染源から拡散伝搬によって徐々に広がり、感染数の上昇下降のしばらく後に再流行するという共通のパターンがある^{8), 9), 23)}。

本論文では、実際の電子メールの送受信関係がSF ネットワーク構造となることを示し、その現実的な接触機会上の拡散伝搬の特性を明らかにする。特に、定常的な蔓延や絶滅ではなく、従来のモデルでは説明できない、いったんは感染が鎮静化した後に再流行する現象^{8), 9), 23)}を(開放系の上記のモデルにおける)新規ユーザによるネットワーク成長が引き起こすことを示唆する。さらに、この開放系のモデルにおいても、重点的なハブの免疫化によってウイルスの拡散が効果的に抑えられることを示す。

まず2章では、電子メールによるウイルスの典型的な行動パターンと、送受信数の分布に関して説明した後、その実測データに基づいた接触関係を表す具体的なネットワークモデルの構成手段を提示する。3章では、ウイルスの実行率や発見率に対する、総感染数や絶滅時間に関する基本特性を明らかにする。4章では、開放系のモデルにおける、再発的な流行現象とハブの免疫化の効果を示す。最後に5章では得られた結果をまとめ、今後の課題について述べる。

2. 電子メールによるウイルス感染

2.1 ウイルスの行動パターン

コンピュータウイルスには、感染するターゲットによる分類で、実行ファイル型、マクロ・スクリプト型、ブートセクタ型などがあり、また、その挙動による主な分類で、他のプログラムに感染するプログラムとしての狭義のウイルス、ネットワークを介して増殖するワーム、正規ソフトを装った単体で動作する不正な(他のプログラムに感染はしない)トロイの木馬などがある。その典型的な行動パターンは、未感染(Susceptible)、潜伏(Hidden Infected)、増殖(Infectious)、発病の順に遷移する¹⁴⁾。

本論文では、電子メールによるウイルスの拡散伝搬に焦点を絞っているので、ワーム活動を行うものを対象とするが、感染したことの発見を逃れるための潜伏の仕方や、発病後の個々のコンピュータ(以後PCと呼ぶ)におけるファイル削除やメモリ浪費などの破壊的行為に関しては議論しない。また、ターゲットとしては主にファイルやマクロ・スクリプトを考え、ユーザ操作による実行や時限起動などの細かな区別をしないかわりに、アドレス帳などを介して増殖するウイルスを広く扱う。しかも、メールの受信や履歴閲覧の際に、特定のタイトルや添付ファイルなどからウイルスを発見した場合、駆除操作の後には再度感染しない免疫(Recovered, Immune)状態を考慮する。

さて、電子メールの送受信関係を表すネットワークを考える。その頂点 $i = 1, \dots, N$ は各PCを、有向辺はアドレス帳などに登録されたPC間のメールの送受信関係を表す。個々のPCの状態は、実際の挙動に近いよう、図1に従って確率的に遷移する¹⁶⁾ものとする。ここで、 λ は潜伏状態Hから感染状態Iへの実行率、 δ は感染の発見率とする。また、頂点 i におけるPCが保持する n_i 個のウイルス(感染メール)から、どれか1つでもウイルスが発見される確率は $1 - (1 - \delta)^{n_i}$ 、 n_i 個の潜伏状態からどれか1つでも増殖状態に移行する確率は $1 - (1 - \lambda)^{n_i}$ となる。これ

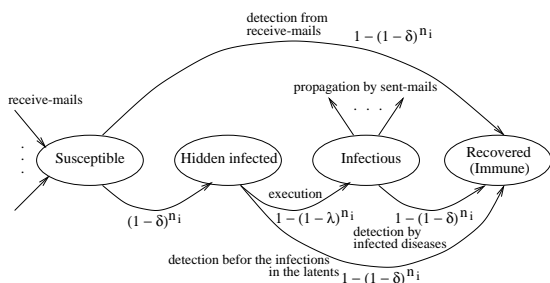


図 1 S-H-I-R 状態遷移図
Fig.1 S-H-I-R state transition diagram.

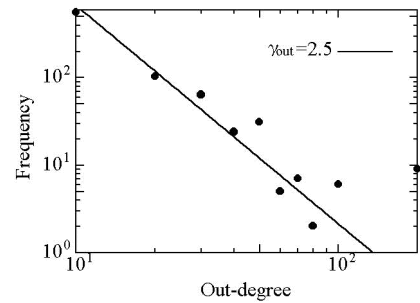


図 2 送信メール数のべき乗分布
Fig.2 Power law distribution for sent-mails.

らの状態遷移は確率的に行われるが、時間の経過ともなると各頂点はいずれは免疫状態 R となる点に注意しよう。

- 未感染状態 S の PC は、感染した相手からメールを受信した時点でウイルスに感染する。
- 受信の際、複数の相手からの受信メール (n_i 個)のうち、どれか 1 つでもウイルスが発見されれば、免疫状態 R になって、その後、再感染することはない。
- 一方、受信メールから感染しても、すぐにウイルスの増殖活動(感染メールの送信)を始めず、潜伏状態 H となる。
- 潜伏状態 H から次のステップで、 n_i 個のうちどれか 1 つでも増殖状態 I になれば、通信相手に感染メールを送る(ただし、発見されにくいよう、同じ相手には 2 度以上感染メールを送信しない)。あるいは、潜伏状態 H でウイルス感染したことが発見されて、免疫状態 R となる。
- 増殖状態 I でもウイルス感染したことが発見されれば、免疫状態 R となる。各 PC が状態 I や H で保持するウイルスの個数 n_i はステップごとに増加しうが、最大でも各頂点 i の入次数分である。

2.2 電子メール送受信数の分布

次に、実測データに基づいて、電子メールの送受信関係を表すネットワーク構造(平均辺数 \bar{k} や次数分布)を考える。

通信総研と東大社会情報研による'00年 10~11月の調査¹²⁾では、1日の送受信数の最頻度は 1~10通、平均送信数は 16.3通、平均受信数は 27.7通で、送受信

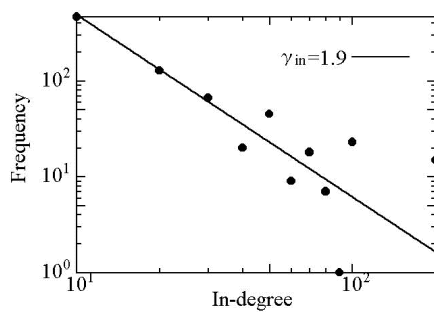


図 3 受信メール数のべき乗分布
Fig.3 Power law distribution for receive-mails.

数がある偏った分布となることが報告されている。実際、これらの送受信数の分布を両対数のグラフで表すと、図 2 と図 3 のようなべき乗則に従う。これらから、送受信数の分布のべき係数はそれぞれ $\gamma_{out} = 2.5$, $\gamma_{in} = 1.9$, 平均送受信数 \bar{k} は 5~20 通程度と推定される。

2.3 構成的な (α, β) モデル

前節の実測データに内在するべき乗則に従った送受信関係を表現する、ネットワークモデルの具体的な構成手段を述べる。

べき乗則は、インターネットのルータ、WWW のリンク、電力網、知人関係など現実の大多数のネットワークに共通する普遍特徴としてきわめて重要である²⁾。驚くことに、このような特徴を持つ SF ネットワークは以下の単純かつ自然な 2 つの生成原理に従う³⁾。

Growth: 新しい頂点を追加しながら、時間的にグラフが成長していく。

Preferential Attachment: 新しく追加された頂点よりも、すでに存在している頂点の方が、辺が張

他の調査では、情報通信総研によって'99年 4月に行われた、郵送によるアンケート結果では 1 日の平均送受信数は 6.2 通、インターネットによるアンケート結果では 14.8 通となっている。
http://www.commerce.or.jp/result/sp3/index.html
'02年 2月に発表された電通によるデジタルライフ調査では、平均送信数は 2.9 通、平均受信数は 5.1 通となっている。

http://www.dentsu.co.jp/marketing/digital_life/
各頂点の辺数を表す、次数 k を持つ頂点の頻度が、 $k^{-\gamma}$ に比例すること。係数 γ の値は対象とするネットワークごとに多少異なるが、実測値でおよそ 2 から 3 である²⁾。

表 1 α - β コインによる辺の生成
Table 1 Edge generation by the α - β coin.

確率	α	$1 - \alpha$
β	新頂点の自己ループ	終点は既存で、始点は新頂点
$1 - \beta$	終点は新で、始点は既存	ともに既存の頂点

られやすく、その頻度は辺の数に比例する (rich-get-richer 現象)。

SF ネットワーク構造上のウイルスの拡散伝搬のシミュレーションを行うため、平均場近似による解析向きモデル³⁾よりも、上記の生成原理に従いつつ、べき乗の傾きが制御できる構造的な (α, β) モデル¹⁰⁾に着目する。 (α, β) モデルでは、頂点の追加 (Growth) ごとに κ 本の辺を、表 1 のようにパラメータ α, β の値

$$\gamma_{in} = \frac{1}{1 - \alpha}, \quad \gamma_{out} = \frac{1}{1 - \beta},$$

に従って確率的に生成することを、ある頂点総数 N になるまで繰り返す¹⁸⁾。ただし、生成途中で仮に作られた孤立点、自己ループ、多重辺は削除する。また、既存の頂点を追加辺の始点あるいは終点とする場合、各頂点の入次数と出次数に比例した頻度 (Preferential Attachment) でそれぞれの頂点を選ばれるものとする。

3. 拡散伝搬と絶滅時間の基本特性

確率的な実行や発見のない拡散伝搬のみの特性として、一様ランダムなネットワークでは平均辺数が 3 本以上あれば任意の初期頂点から全体の 90% 以上に蔓延するが、収束までの時間 (ステップ数) は規模 N に比例するのに対して、SF ネットワークでは平均辺数が 5 本以上で蔓延する一方、収束までの時間はほとんど規模によらないことが明らかになっている^{13), 18)}。

本章では、確率的な状態遷移に基づく拡散伝搬と絶滅状態への収束に関する特性を述べる。具体的には、前章の手順で構成した $N = 2000$ の SF ネットワーク上の SHIR モデル (図 1) において、最も伝搬しやすい出次数最大のハブを初期感染源 I とし (他のすべての頂点の初期状態は未感染状態 S)、実行率 $\lambda = 0.1, \dots, 0.9$ 、発見率 $\delta = 0.1, \dots, 0.9$ 、平均辺数 $\bar{k} = 6.8 \sim 19.7$ のそれぞれの組合せに対して、絶滅するまでの確率的な状態遷移を 100 回試行した。

図 4, 図 5, 図 6 は、総感染 PC 数 (増殖状態 I に移行した数) の $\delta = 0$ の場合の最大感染 PC 数に対する比、総感染メール数 (感染が発見されたものを含む) の比、絶滅までのステップ数 (日数に相当) についての、100 回試行の平均値をそれぞれ示す。図中に

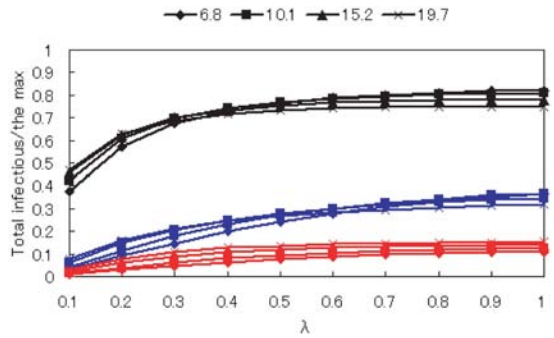


図 4 総感染 PC 数の比
Fig. 4 Rate of the total infectious PCs.

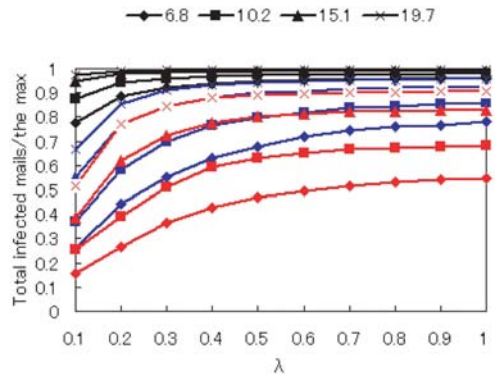


図 5 総感染メール数の比
Fig. 5 Rate of the total infected mails.

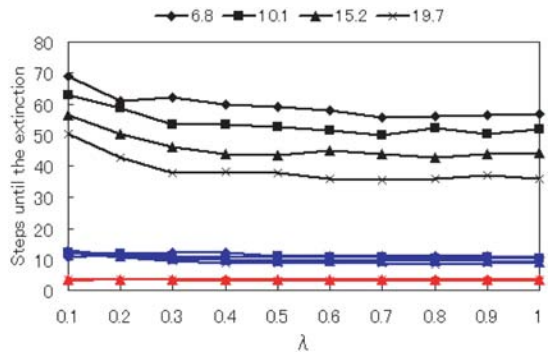


図 6 絶滅までのステップ数
Fig. 6 Number of steps until the extinction.

おける、菱形、四角、三角、 \times の各記号は平均辺数 $\bar{k} = 6.8, 10.1, 15.2, 19.7$ に、黒線、青線、赤線は発見率 $\delta = 0.1, 0.5, 0.9$ にそれぞれ対応する。

図 4 に示す総感染 PC 数は、実行率 λ が大きくなるほど増加するが、線色で区別した発見率 δ が大きくなるに従って抑えられる。ただし、各記号の平均辺数 \bar{k} には、ほとんど依存しない。一方、図 5 に示す総感染メール数は、平均辺数 \bar{k} が大きいほど増加する。実

ゆえに、平均辺数は $\bar{k} < \kappa$ となり、その値も確率的に決まる。

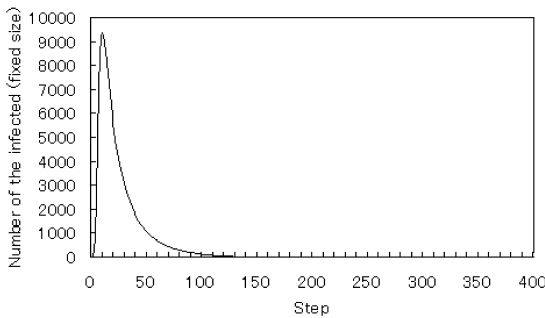


図 7 固定規模における感染数の変化

Fig. 7 Time-series of num. of the infected at a fixed size.

行率 λ が大きくなるほど、また、発見率 δ が小さくなるほど、増加する点は総感染 PC 数と同様である。図 6 に示す絶滅までの時間には、黒線の $\delta = 0.1$ の場合以外は、平均辺数 \bar{k} に対する依存性はほとんど見られず、緑色で区別した発見率 δ の値のみに依存したほぼ一定値となっている。発見率が低い $\delta = 0.1$ のときは、実行率 λ が大きくなるに従って絶滅までの時間が減少し、平均辺数 \bar{k} が大きいほど早い段階で収束している。 λ や \bar{k} が大きいほど早く絶滅するのは、いったん広がった複数の感染メールから発見されやすくなるためと考えられる。

図 7 は、ネットワーク規模 N を固定した場合の感染数 $\sum_{i=1}^N n_i$ (各時刻の状態 H+I の保有ウイルス数) の典型的なパターン (感染が広がった後、比較的緩やかに絶滅状態に収束すること) を示す。ここで、次章に示すネットワーク成長モデルとの比較のため、 $N = 18934$, $\lambda = 0.1$, $\delta = 0.04$, $\bar{k} = 6.6$ とし、感染源 I をランダムに選んだ 5 個の頂点とした。

4. 開放系のウイルス伝搬モデル

4.1 確率的 SHIR モデル

初期規模 $N = 400$ で構成した SF ネットワークから、ウイルスの拡散伝搬をとまなう確率的な状態遷移の各ステップごとに、2.3 節で述べた手続きに従って、線形成長: +50 個ずつの頂点と対応する辺を追加、指数成長: $\times 1\%$ ずつの頂点と対応する辺を追加、のどちらかの方法で成長していく開放系のモデルを考える。その際、パラメータなどは図 7 と同じ条件とし、ステップ数 400 回で線形成長と指数成長のネットワーク規模 N がそれぞれ同程度 (20350 と 18934) となる。以後、400 ステップまでで議論するが、図 8、図 9 はそれ以降を含めた、線形成長と指数成長にお

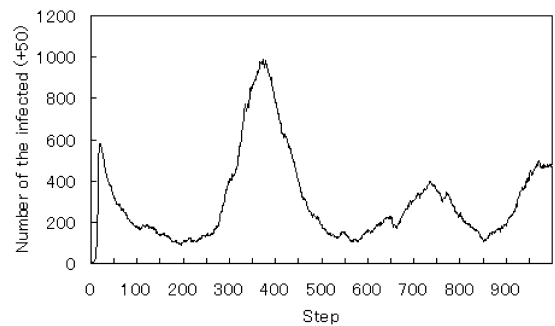


図 8 線形成長における感染数の変化

Fig. 8 Time-series of num. of the infected in a linear growth.

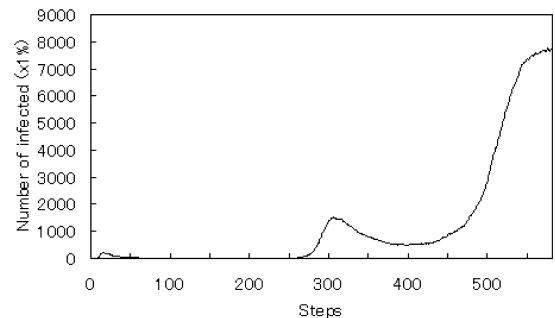


図 9 指数成長における感染数の変化

Fig. 9 Time-series of num. of the infected in an exponential growth.

る再流行現象の典型例を示す。再流行の原因は、絶滅しかけたほんのわずかな感染 PC が未感染の新規ユーザ PC を介して、再度増殖する機会を得ていくためと考えられる。

次に、重点的なハブの免疫化の効果を図 10、図 11 に示す。グラフの上から順に黒線: ウイルス発見のみによる Normal な駆除、緑線: ネットワーク成長に追従した頂点の免疫率 10%、青線: 20%、赤線: 30% に対する、100 回の試行の平均感染数 (状態 H+I の数) の変化をそれぞれ表す。ただし、ウイルス発見によって免疫状態になった頂点とは別に、30 ステップごとに成長した規模に対する比率で、出次数が大きい (ハブ) 順にあるいはランダムに頂点を免疫化する。

図 10、図 11 の線形成長と指数成長ともに、ハブ頂点の免疫率を上げるに従って感染数が減少し、免疫率 30% (赤線) ではほぼ絶滅させるまでに至っている

ランダムな 10 種類のネットワーク生成 $\times 10$ 回の確率的な拡散伝搬による 100 回の試行。

計算時間の節約以外、30 という数自体には日数程度の意味しかないが、アンチウイルスのインストールや更新などの作業が必要な免疫化よりも (毎ステップごとの) ウイルスの状態遷移の方が進行が速いことを表す意味で、妥当と考えられる。

実際は、線形成長と指数成長の中間的な成長の仕方と考えられる。

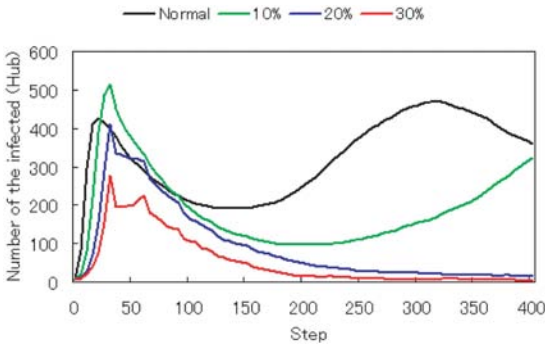


図 10 線形成長におけるハブ免疫化
Fig. 10 Hub immunization in a linear growth.

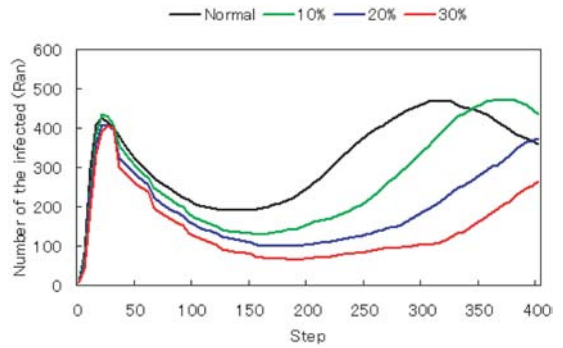


図 12 線形成長におけるランダム免疫化
Fig. 12 Random immunization in a linear growth.

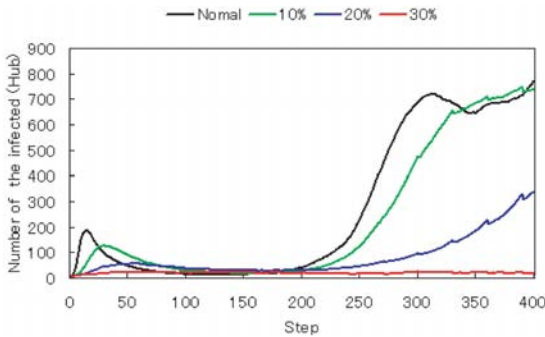


図 11 指数成長におけるハブ免疫化
Fig. 11 Hub immunization in an exponential growth.

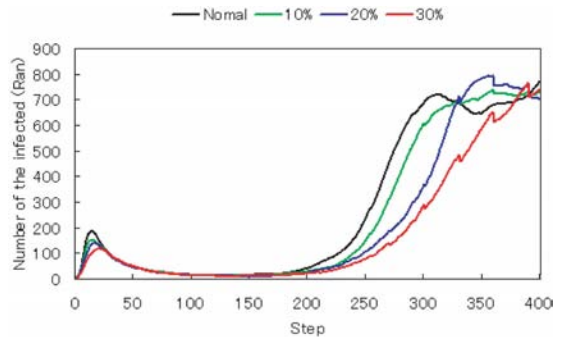


図 13 指数成長におけるランダム免疫化
Fig. 13 Random immunization in an exponential growth.

(ただし、完全絶滅ではなく数個のウイルスは存在する)。一方、図 12、図 13 に示すランダムな頂点の免疫化では、免疫率を上げるに従って感染数が若干減少するものの、30%でも絶滅から程遠い。

ここで、ウイルスの実行や発見は確率的に行われるため、なかには 400 ステップまでに完全に絶滅する場合も存在する¹⁷⁾。図 10~13 の線形成長と指数成長の場合において、100 回の試行中で完全に絶滅しなかった頻度を表 2 にそれぞれ示す。指数成長におけるハブに対する免疫率 10%や 20%で不完全絶滅頻度が Normal よりも増え、免疫化が逆効果のようにも見えるが、図 14、図 15 に示すように 400 ステップまでの (100 試行の平均の) 総感染数は、免疫率を上げるに従って大きく減少している。図中、Hub と Ran はそれぞれハブとランダムに選んだ頂点に対する免疫化、括弧内の記号は (H や I の) 状態数を表し、Ex は前記の再流行以外の完全に絶滅した場合を示す。完全に絶滅していない (再流行する) 場合の総感染数においても、せいぜい半分に減少のランダムな免疫化より、1/6 以下の減少を示すハブの免疫化に効果がある。

上記の免疫率 30%の線形成長の場合で、30 ステップごとに免疫化された頂点数とウイルス発見数の 100

表 2 線形成長 (上) と指数成長 (下) の不完全絶滅頻度
Table 2 Frequency of non-complete extinctions in a linear (top) and exponential (bottom) growth.

Immune	Normal	10%	20%	30%
Hub	100	92	35	6
Random	100	96	92	70

Immune	Normal	10%	20%	30%
Hub	50	77	73	10
Random	50	38	29	23

回試行の平均値を図 16、図 17 に示す。図 16 では、免疫化されたハブの頂点数 (三角印) がウイルス発見による状態 R の頂点数 (四角印) よりも圧倒的に多く支配的で、特にその傾向はステップ数が大きくなるほど顕著になっている。ここで、点線は免疫化するハブとして (図 17 ではランダムに) 選ばれた頂点がすでに状態 R になっていた場合を表し、ステップ数によらず免疫化されたハブ全体の 1 割以下 (図 17 では半分程度) となっている。一方、ランダムな免疫化の場合の図 17 では、ステップ数を大きくするに従ってその比率は減少するものの、ウイルス発見数 (四角印) がランダムに免疫された頂点数 (三角印) と同程度になっている。

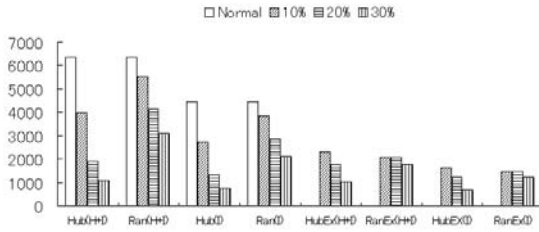


図 14 線形成長における総感染数の比較

Fig. 14 Comparison of the total infected in a linear growth.

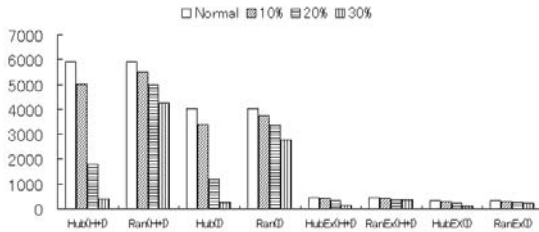


図 15 指数成長における総感染数の比較

Fig. 15 Comparison of the total infected in an exponential growth.

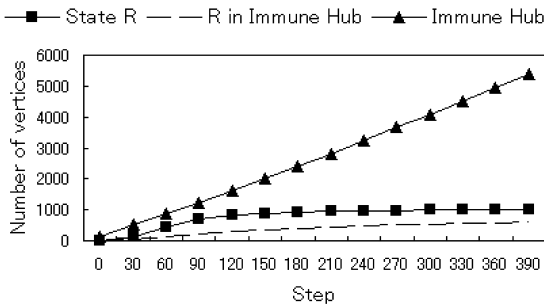


図 16 免疫状態の頂点数 (ハブ免疫化 30%)

Fig. 16 Number of vertices in state R (hub immunization 30%).

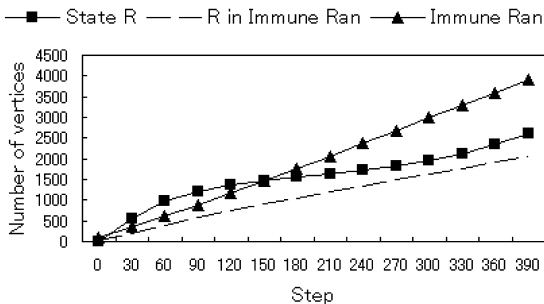


図 17 免疫状態の頂点数 (ランダム免疫化 30%)

Fig. 17 Number of vertices in state R (random immunization 30%).

図 18, 図 19 に示す免疫率 10% の場合は, ウィルス発見数と免疫化された頂点数の大小関係が逆転する. 図 20, 図 21 に示す免疫率 20% の場合は, 10% と

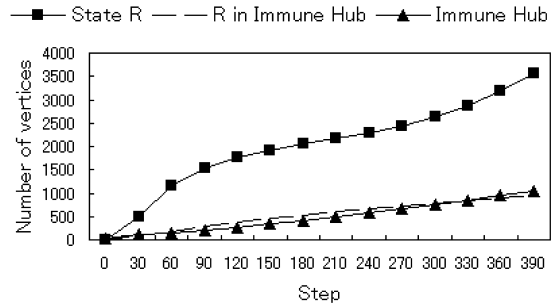


図 18 免疫状態の頂点数 (ハブ免疫化 10%)

Fig. 18 Number of vertices in state R (hub immunization 10%).

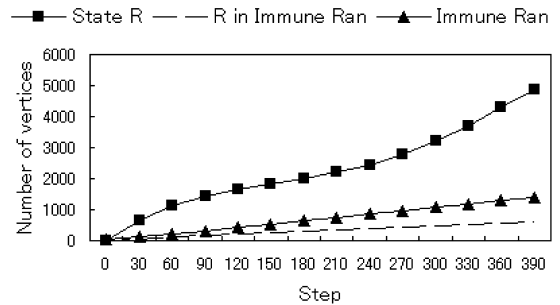


図 19 免疫状態の頂点数 (ランダム免疫化 10%)

Fig. 19 Number of vertices in state R (random immunization 10%).

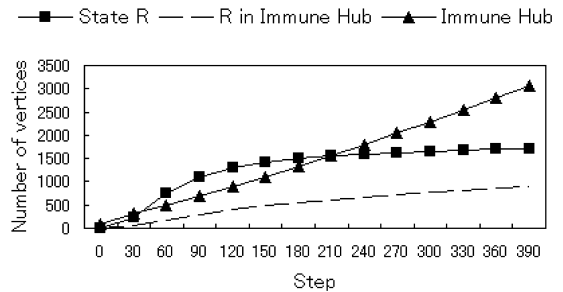


図 20 免疫状態の頂点数 (ハブ免疫化 20%)

Fig. 20 Number of vertices in state R (hub immunization 20%).

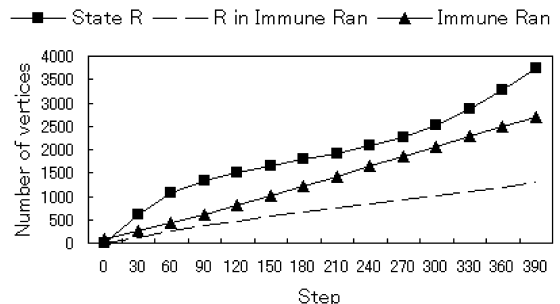


図 21 免疫状態の頂点数 (ランダム免疫化 20%)

Fig. 21 Number of vertices in state R (random immunization 20%).

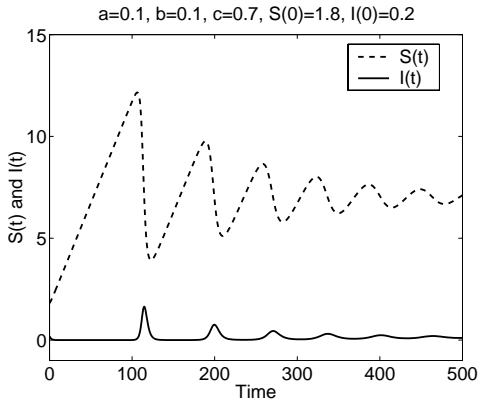


図 22 線形成長における $S(t)$ と $I(t)$
Fig. 22 $S(t)$ and $I(t)$ in a linear growth.

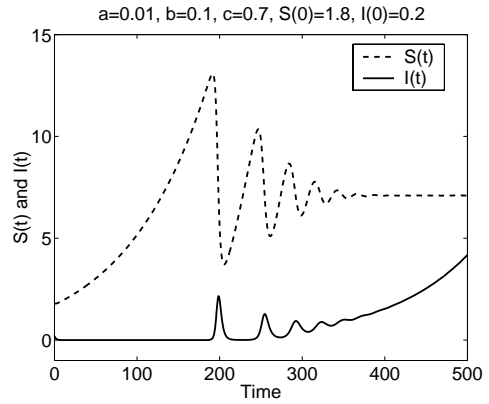


図 24 指数成長における $S(t)$ と $I(t)$
Fig. 24 $S(t)$ and $I(t)$ in an exponential growth.

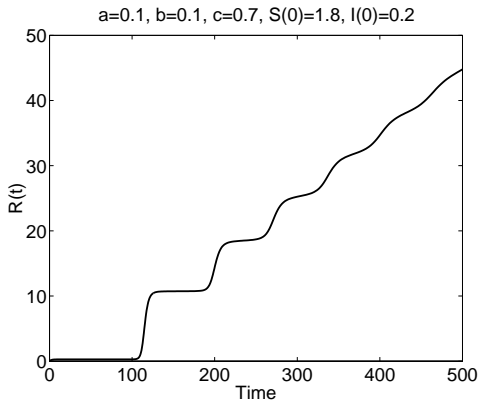


図 23 線形成長における $R(t)$
Fig. 23 $R(t)$ in a linear growth.

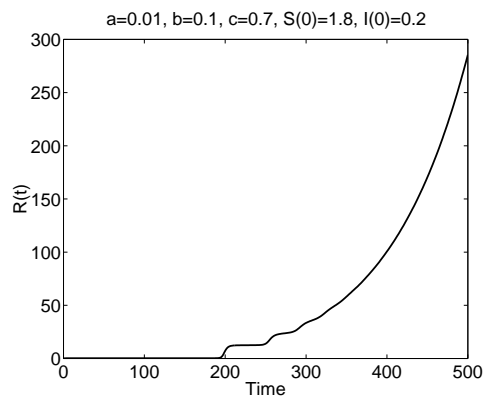


図 25 指数成長における $R(t)$
Fig. 25 $R(t)$ in an exponential growth.

30%の中間的な結果となった。指数成長の場合も上記と同様で、特にウイルス発見数に対する免疫化されたハブの頂点数の比率がさらに大きくなる。これら図 16、図 17 を先の結果 (図 10~15) と合わせて考えると、感染数の激減や絶滅にはウイルス発見よりもハブの免疫化が効果的であることがより鮮明になる。

4.2 決定論的 homogeneous 近似

伝染病における理論解析のための、決定論的な Kermack-McKendrick モデルにおいて、規模を固定した場合は感染率があるしきい値以上ならいったん広がった後に絶滅する挙動しか表現できないが、出生 (ネットワーク成長) を考慮すると再発的な流行現象が起こることが知られている²²⁾。ただし、出生率と死亡率を同一とした総人口が一定の場合や、競合項 (縄張り争い) を持つモデルが主に議論され、特に何の規制もなく成長し続けているコンピュータネットワークにはそのままではあてはまらない。

一方、接触関係の空間的な局所性の観点から、再発

的な振動現象が例示されている⁸⁾が、例示以外ではどんなトポロジーに起因するのか明らかになっていない。

本節では、前節の確率論的モデルに対応する、ネットワーク成長をとまなう決定論的な Kermack-McKendrick モデルについて議論する。その際、潜伏状態 H を省略し、接触関係を homogeneous に近似して単純化することで、再流行現象がネットワーク構造より成長に起因することを示唆する。

まず、線形成長の SIR モデルは、

$$\frac{dS(t)}{dt} = a - bS(t)I(t), \tag{1}$$

$$\frac{dI(t)}{dt} = bS(t)I(t) - cI(t), \tag{2}$$

$$\frac{dR(t)}{dt} = cI(t) \tag{3}$$

と表される。ここで、未感染数 $S(t)$ 、感染数 $I(t)$ 、免疫数 $R(t)$ 、ネットワーク規模 $N(t) = S(t) + I(t) + R(t)$ 、係数 $a, b, c > 0$ 、初期値 $S(0), I(0) > 0, R(0) = 0$ とする。

上記の連立微分方程式 (1) ~ (3) を、ステップ幅 $\Delta t = 10^{-4}$ とした 4 次の Runge-Kutta 法によって求めた結果を、図 22, 図 23 に示す。 S, I が $dI/dt = 0$ と $dS/dt = 0$ にそれぞれ対応する平衡解 $S^* = c/b$ と $I^* = a/c$ に収束し、 $R > 0$ は減衰振動しながら線形に増加している。

また、解 $N(t) = N(0) + at$ を持ち、 $S(t)$ と $I(t)$ の級数展開は、

$$S(t) \approx 1.8 + 0.08t + 0.00856t^2 - 0.0014501t^3 + 0.0002008t^4 - 0.0000250t^5 + O(t^6),$$

$$I(t) \approx 0.2 - 0.104t + 0.02784t^2 - 0.0050458t^3 + 0.0006821t^4 - 0.0000704t^5 + O(t^6),$$

となる ($R(t)$ は $N(t) - S(t) - I(t)$)。

一方、指数成長の SIR モデルは、

$$\frac{dS(t)}{dt} = aN(t) - bS(t)I(t), \quad (4)$$

$$\frac{dI(t)}{dt} = bS(t)I(t) - cI(t), \quad (5)$$

$$\frac{dR(t)}{dt} = cI(t) \quad (6)$$

と表される。同様に式 (4) ~ (6) を 4 次の Runge-Kutta 法によって求めた結果を図 24, 図 25 に示す。 S のみが $dI/dt = 0$ に対応する平衡解 $S^* = c/b$ に収束し、 $I, R > 0$ は減衰振動しながら指数的に増加している。

また、解 $N(t) = N(0)e^{at}$ を持ち、

$$S(t) \approx 1.8 - 0.16t + 0.0962t^2 - 0.0173206t^3 + 0.0002487t^4 - 0.0000308t^5 + O(t^6),$$

$$I(t) \approx 0.2 - 0.104t + 0.02688t^2 - 0.0045339t^3 + 0.0005457t^4 - 0.0000453t^5 + O(t^6).$$

前節の確率論的 SHIR モデルとの対応では、 $t \approx 300$ で線形成長と指数成長が同程度の規模となり、図 22, 図 24 の実線の $I(t)$ は $0 \leq t < 300$ において、それら確率論的モデルにおける図 8, 図 9 と定性的に同様な再流行現象を表している。一般に確率論的モデルの平均値は決定論的モデルに対応するものの、個々の挙動では必ずしも大小関係が一致しないこと¹⁷⁾ や、初期感染数は 0 でなく $I(0) = 0.2$ であることに注意されたい。また、決定論的モデルにおける再流行のタイミングや大きさが、確率論的モデルにおける +50 や $\times 1\%$ などのネットワーク成長率に対応する係数 a 、ウイルスの実行率 λ や発見率 δ に対応する係数 b や c (平均辺数 \bar{k} の影響は b に含まれる) に依存する点も同様である。

5. む す び

コンピュータウイルスの被害の中で、電子メールによるものは 70 ~ 80% 以上で^{6), 7), 9)}、その損害も大きい。そこで、個々の駆除のためのアンチウイルスの開発のみならず、その効果的なネットワーク上の配置から拡散伝搬による被害を抑えることが望まれる。

従来のコンピュータウイルスの拡散モデルでは、解析に適した一様ランダムや完全グラフ²¹⁾、一次元格子などのネットワーク構造が扱われ、電子メールによるウイルス感染の接触機会の偏りが指摘されている¹⁶⁾ もの、実際の送受信数に基づいたネットワーク構造や、アンチウイルスの配置方法は明らかでなかった。一方、社会的な知人関係などを表す SF ネットワーク構造上の SIS モデルでは、絶滅のしきい値が存在しないこと¹⁹⁾ や、重点的なハブの免疫化によって拡散が抑えられることが、平均場解析から最近示された^{4), 20)}。しかしながら、これらのモデルではネットワーク規模が固定され (あるいは $N \rightarrow \infty$)、ウイルスの蔓延か絶滅の定常状態しか議論されておらず、観測データに見られる再流行現象^{8), 9), 23)} を説明できなかった。

本論文では、電子メールによる拡散モデル¹⁶⁾ を基に、

- 実際の送受信関係¹²⁾ がべき乗分布に従う SF ネットワーク構造となることを示し、その現実的な接触関係上の拡散伝搬の基本特性を明らかにした。特に、増殖状態への実行率 λ や平均辺数 \bar{k} が大きいほど、いったん広がった複数の感染メールからウイルスが発見されやすくなって結果的に早く絶滅することが分かった。ただし、 \bar{k} は今後さらに増えると思われる平均送受信数に相当するので、一時的な被害範囲の拡大が起こりうる。
- また、新規ユーザによるネットワーク成長によって、再流行現象が引き起こされることを、確率論的および決定論的な双方のモデルから示唆した。
- さらに、重点的なハブの免疫化によって、再流行する可能性があるウイルスの拡散伝搬も効果的に抑えられることを示した。

ランダムな免疫化では再流行が防げない場合でも、ハブの免疫率を 30% 以上にすればほとんど絶滅し (表 2 より 9 割は完全絶滅)、総感染数も大幅 (図 14 より 1/6 以下) に減少することが分かった。

以上、実際の送受信関係に基づくウイルスの拡散伝搬の基本特性、新規ユーザによる再流行現象、重点的なハブの免疫化の効果が明らかになった。また、ハブ

独 Kiel 大のメールサーバのログ解析でも同様な報告がある⁵⁾。

を見つけるには、接続辺を順にたどる単純な幅優先探索ではうまくいかないことを我々は経験的に見つけ、より効率的な探索法についてもすでに検討しているが、その内容に関しては別の機会に報告したい。

本論文のモデルでは、ウイルスの実行率や発見率は時間的に一定で、さまざまな送受信相手を表現できるように確率的に生成されたネットワークが成長するものの、辺の付け換えは行っていない。一方、電子メールによるウイルスの特徴をさらに詳細にとらえれば、たとえば、送受信リスト中の相手が大きく変わったり、広く蔓延したウイルスには特に注意して対策を講じるなど発見率も変化することなどが考えられる。また、誰もがメーリングリストを利用すれば広範囲に送信可能であり、ハブとなるヘビーユーザも固定できないかもしれない。いたづらに詳細化することは好ましくないが、こうした電子メールの特徴を必要に応じてさらにふまえた検討や、感染数の観測データの整備、それらデータとの比較によるモデルの洗練化などが今後の課題として考えられる。

謝辞 本研究の一部は、文部科学省科学研究費13680404の援助を受けている。

参 考 文 献

- 1) Albert, R., Jeong, H. and Barabási, A.-L.: Error and attack tolerance of complex networks, *Nature*, Vol.406, pp.378–382 (2000).
- 2) Albert, R. and Barabási, A.-L.: Statistical Mechanics of Complex Networks, arXiv: cond-mat/0106096v1 (2001).
- 3) Barabási, A.-L., Albert, R. and Jeong, H.: Mean-field theory for scale-free random networks, *Physica A*, Vol.272, pp.173–187 (1999).
- 4) Dezsö, Z. and Barabási, A.L.: Halting viruses in scale-free networks, *Physical Review E*, Vol.65, 055103 (2002).
- 5) Ebel, H., Mielsch, L.-I. and Bornholdt, S.: Scale-free topology of e-mail networks, *Physical Review E*, Vol.66, 035103(R) (2002).
- 6) 情報処理振興事業協会：平成12年度国内におけるコンピュータウイルス被害状況調査，平成13年2月(2001)。
- 7) 情報処理振興事業協会セキュリティセンター：国内におけるコンピュータウイルス被害状況調査，平成14年3月(2002)。
- 8) Kephart, J.O. and White, S.R.: Measuring and Modeling Computer Virus Prevalence, *Proc. 1993 IEEE Comp. Soc. Symp. on Res. in Security and Privacy*, pp.2–15 (1993).
- 9) Kephart, J.O., Sorkin, G.B., Chess, D.M. and White, S.R.: コンピュータウイルス vs. デジタル免疫系, *日経サイエンス* (Feb. 1998).
- 10) Kumar, R., et al.: Extracting large-scale knowledge bases from the web, *Proc. 25th VLDB Conf.*, pp.7–10 (1999).
- 11) May, R.M. and Lloyd, A.L.: Infection dynamics on scale-free networks, *Physical Review E*, Vol.64, 066112 (2001).
- 12) 三上俊治：日本人とインターネット生活—ワールドインターネットプロジェクト(2001)。
<http://sophy.asaka.toyo.ac.jp/users/mikami/info&media/>
- 13) 箕浦正人, 林 幸雄：さまざまなネットワークにおける情報伝播と構造の特徴, 電気関係学会北陸支部連合大会, 講演論文集 C-49, p.185 (2002).
- 14) 日経 NETWORK：感染, 潜伏, 増殖, 発病—典型的な行動パターンとは, 特集ウイルス大図鑑 Part 1 生態, pp.062–069 (Aug. 2001).
- 15) Newman, M.E.J.: Spread of epidemic disease on networks, *Physical Review E*, Vol.65, 016128 (2002).
- 16) 岡本 剛, 石田好輝：電子メールにより拡散するコンピュータウイルスの拡散モデルの解析, 信学論 D-I, Vol.J84-D-I, No.5, pp.474–482 (2001).
- 17) 大久保明：生態学と拡散, 緒論, 筑紫書館(1975).
- 18) 大久保和彦, 林 幸雄, 蛭川 繁：Web 的ネットワークにおける情報伝搬率と速度, 信学論 D-I, Vol.J85-D-I, No.2, pp.241–244 (2002).
- 19) Pastor-Satorras, R. and Vespignani, A.: Epidemic dynamics and epidemic states in complex networks, *Physical Review E*, Vol.63, 066117 (2001).
- 20) Pastor-Satorras, R. and Vespignani, A.: Immunization of complex networks, *Physical Review E*, Vol.65, 036104 (2002).
- 21) 千石 靖, 岡本栄司, 満保雅浩, 植松友彦：コンピュータウイルスの拡散と消滅の大域的振る舞いについて, 情報処理学会論文誌, Vol.37, No.4, pp.579–587 (1996).
- 22) 重定南奈子：侵入と伝播の数理生態学, 第8章 感染症の侵入と伝播, 東大出版会(1992).
- 23) White, S.R., Kephart, J.O. and Chess, D.M.: Computer Viruses: A Global Perspective, *Proc. 5th Virus Bulletin Int. Conf.* (1995).

(平成15年2月2日受付)

(平成15年3月18日採録)



林 幸雄(正会員)

昭和 37 年生。昭和 60 年豊橋技術科学大学工学部電気電子工学科卒業。昭和 62 年同大学院修士課程修了。富士ゼロックス(株) (株)ATR 視聴覚機構研究所(株)ATR 人間情報通信研究所等の研究員を経て、現在、北陸先端科学技術大学院大学知識科学研究科助教授。工学博士。数理工学, 情報幾何学, 力学系, Web 生態系, ニューラルネットの研究に従事。日本応用数学会, 電子情報通信学会, 日本神経回路学会, INNS 等各会員。



箕浦 正人

昭和 54 年生。平成 13 年名古屋工業大学工学部応用化学科卒業。平成 15 年北陸先端科学技術大学院大学知識科学研究科前期課程修了。



松久保 潤

昭和 51 年生。平成 9 年北九州高等工業専門学校電子制御工学科卒業。平成 11 年同校専攻科制御専攻卒業。平成 13 年北陸先端科学技術大学院大学知識科学研究科前期課程修了。現在, 同後期課程在学中。電子情報通信学会学生会員。