修　士　論　文

SF ネットワーク構造に基づく情報伝搬と コンピュータウィルスの伝染

指導教官　　林　幸雄　助教授

北陸先端科学技術大学院大学
知識科学研究科知識システム基礎学専攻

150074　　箕浦　正人

審査委員：　林　幸雄　　助教授（主査）
中森　義輝　教授
佐藤　賢ニ　助教授
橋本　敬　　助教授

2003 年 2 月

目　次

第　1　章

1 はじめに

従来のグラフ理論やランダムネットワークとは異なる観点からネットワークに関する研究が近年盛んに行われるようになった．それらの研究[1][2]から実際のネットワークがランダムネットワークや正則ネットワークとは異なる普遍的な特徴があることがわかってきた．例えば，WWW では任意のホームページが意外に少ないリンク数でつながっている．また，リンク数が多いホームページの数が少なく，リンク数がほとんどないホームページの数がたくさんあるという冪乗分布に従う．本研究では，ネットワークに関する研究を行うためにネットワークモデルの基本特性を明らかにする．

以下本章では，ネットワークに関する先行研究や本研究の目的，方法などを述べる．

1.1 研究の動機と背景

1.1.1 ネットワーク構造

現代の我々は，様々なネットワークにかかわって生活している．ネットワークというのは，頂点とそれをつなぐ辺によって描くことができる．例えば，友人ネットワークの場合，頂点が人で，辺が友人関係となる．近年，急激に発展しているインターネットの場合は，頂点がホームページ，辺がリンクとなる．この他にも，ルータ，e-mail なども情報通信に関するネットワークである．友人などの社会ネットワークでは，見知らぬ二人が知り合いの知り合いを通じて接点があるということはよくあることである．統計的には，6 人の知人を介して結びついており、WWW では１９回のリンクを介してつながっている．このように，実際のネットワークは意外に小さいということで，スモールワールドネットワークと言われる[3].

1998 年に Watts と Strogatz は，WWW，映画俳優ネットワーク，電力網，線虫の神経回路網などのネットワークがランダムネットワークや正則ネットワークではな

く，中間的特徴をもつことを明らかにした[1]．中間的特徴というのは，任意の頂点間の距離が小さく，クラスター化されているということから，スモールワールドネットワークと呼ばれる．

さらに，Barabási と Albert は，WWWにおいて，各ドキュメントの次数分布が，冪乗分布にしたがっていることを明らかにした[4]．冪乗分布とは，頂点の数とそれが持つ次数の関係が直線的特徴を持つことで，このようなネットワークをスケールフリーネットワークと名付けた．

1.1.2 ネットワーク研究

過去数 10 年にわたってグラフ理論では，ランダムグラフや正則グラフが精力的に扱われてきたが，近年，現実のネットワーク構造はこれらとは全く異なり，すべてのネットワークに共通かつ普遍的な構造が明らかになった．その構造がスケールフリーネットワークである．

これらの研究に刺激されて，多くのネットワークに対する新しい知見が現在も次々と発見されている．それらは，情報通信だけでなく，生物，言語などの広範囲に及ぶ．多くのネットワークがスモールワールドネットワーク，スケールフリーネットワークであることが明らかになり，最近はネットワークの構造に関する研究だけでなく，ネットワークを利用した研究も行われている．例えば，WWWの検索エンジン，コンピュータウィルスの伝染，経済学における富の分布の解析，コンピュータネットワークのライフラインの設計と被害予測などが行われている．このようにネットワークを利用した研究が頻繁に行われるうになり，被害予測や耐故障性などに関するシミュレーションを行う際に，実際のネットワークの特徴を反映したネットワークモデルの生成法が重要になっている．頻繁に使われる正則ネットワークモデルの辺を数本ランダムにつなぎ合わせたスモールワールドネットワークモデルでは，スケールフリーネットワークの特徴を持たないことは明らかであるが，その逆は定かでない．したがって，スケールフリーネットワークモデルがスモールワールドネットワークの特徴を持つかどうかを調べることは非常に重要である．

1.1.3 コンピュータウィルスに関する研究

近年，インターネットが急激に拡大し，これからもその傾向はさらに大きくなることが予測される．このような状況の中，「不正アクセス」，「コンピュータウィルス」などが深刻な問題となっている．特に，コンピュータウィルスは増殖力が高く，不特定多数のシステムに被害をもたらす．コンピュータウィルスの被害を最小限におさえるために，感染予測と予防は重要な課題である．

コンピュータウィルスの感染予測に関する研究はいろいろな方法で行われている．2001 年に，岡本，石田らは SIR（S:Susceptible，I:Infected，R:Recover）モデルを使ってウィルスの広がりについて検討を行った[5]. この研究でのネットワークは１次元格子で，接続数が正規分布にしたがうネットワークである．彼らは，ウィルスが絶滅するまでに平均で７割以上の PC が感染し，ほとんど PC にウィルスが拡散することを明らかにした．また，彼らはアンチウィルスの導入割合が 8 割以上になるとウィルスを絶滅させることができることも明らかにした[6].

この他にも，M.E.J.Newman らは大学内のメールネットワークを調査して，そのネットワークを利用したウィルスの拡散現象を解析した[7]. この論文では，ネットワーク内の PC にランダムにアンチウィルスをインストールさせ免疫するよりも，ハブとなる PC に優先的に免疫する方が被害を抑えることができるということが明らかにされた．

しかし，従来の免疫学の閾値理論では，実際に起きている被害を説明することはできない．実際的な被害予測や予防対策を行うためには，従来の免疫学や研究とは異なる方法で検討を行う必要がある．

1.2 研究目的

本研究では,

1. スケールフリーネットワークモデルがスモールワールドネットワークの特徴を持つかということを調べ，決定論的な情報伝搬の特性を調べる.
2. 実際のメール送受信に基づくネットワークモデルから，コンピュータウィルスの確率論的な状態遷移に基づく伝染特性を明らかにする.

という２つの目的に対して研究を行う.

1.3 研究方法

1.3.1 ネットワークモデルの検証

論文の共著，ルータ，WWW などのさまざまなネットワークに関する実測値に基づいて，R.Kumar らによる(α，β)モデル[8]を使って，スケールフリーネットワークモデルを生成し，スモールワールドネットワークの特徴をもつかどうかを検証した.

また，そのネットワークモデルを利用して，平均辺数を変えて情報の広がる範囲を調べる.

1.3.2 コンピュータウィルスの伝染

(α，β)モデルを使って，実際のメール送受信量[9]を反映したネットワークモデルを生成し，コンピュータウィルスの伝染特性を明らかにする.

第　2　章

2 ネットワークモデルの検証

第２章では，実際のネットワークの冪級数を使って，スケールフリーネットワークモデルを生成し，精度よく近似されているかを検証する．また，そのネットワークモデルがスモールワールドネットワークの特徴である，任意の頂点間の距離が小さく，ネットワークがクラスター化されているかということも検証する．

2.1 ネットワークの種類

本研究では，表 1 に示すさまざまなネットワークの実測値[4]を用いて，次数の異なる 8 つのネットワークについて精度よく冪級数が近似されているか検証を行った．2.4 節のスモールワールドネットワークの検証，および第３章の情報伝搬にも表 1 の値を用いる．

表 1　ネットワークの冪級数と平均辺数の実測値

ネットワーク名	In-degree γ_{in}	Out-degree γ_{out}	平均辺数 $\bar{k}$
論文の引用，SPIRES	1.2	1.2	173
論文の引用，neuro	2.1	2.1	11.54
映画俳優の共演	2.3	2.3	28.78
ルータ	2.48	2.48	2.57
電力網	4.0	4.0	2.67
WWW,Kumar'99	2.1	2.38	7
WWW,Albert'99	2.1	2.45	4.51
WWW,Broder'00	2.1	2.72	7.5

2.2 (α, β)モデル

前項の表 1 に示した各パラメータに従って, 冪乗分布に従うスケールフリーネットワークを具体的に生成する手順を述べる.

スケールフリーネットワーク構造は現実のすべてのネットワークに共通し、以下の単純かつ自然な二つの生成機構に従う.

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 Attachiment:辺がたくさん張られている頂点の方が辺を張られやすく，その頻度は辺の数に比例する（rich-get-richer 現象）.

頂点結合率はWWW，ルータ，電力網などそれぞれ異なる係数をもっており，WWWでは各頂点へ入る辺と出る辺に関して違った係数を持ち，$\gamma_{in}=2.1$，$\gamma_{out}=2.38$ を持ち，ルータでは$\gamma_{in}=\gamma_{out}=2.48$ を持つ.

パラメトリックに制御可能な冪分布に従うグラフを人工的に実現する（α, β）モデル[8]を考える.（α, β）モデルでは, 時刻$t \gg 1$, 辺数$i \gg 1$ を持つ頂点の存在確率 $p_{i,t}$ が

$$In: p_{i,t} = i^{\frac{-1}{1-\alpha}}, \quad Out: p_{i,t} = i^{\frac{-1}{1-\beta}}$$

となることが近似的に導ける.

具体的な生成手順を以下に示す.

[Step 0] $t=0$ として， 1 個だけの頂点を考える.
[Step 1] 新頂点を追加する.
[Step 2] $\alpha-\beta$ コインに従って k 本の辺を生成する.
[Step 3] $t=t+1$ とする.
[Step 4]定められた頂点 n になるまで Step 1 と 2 を繰り返す.
[Step 5]自己ループを削除する.

表 2　$\alpha-\beta$ コインによる辺の生成

確率	α	$1-\alpha$
β	新頂点の自己ループ	新頂点から既存の頂点
$1-\beta$	既存の頂点から新頂点	既存の頂点から既存の頂点

（Step2 の補足：辺の張り方は表 2 のように 4 通りの確率でそれぞれ選択される．その際既存の頂点を選択する場合は，入次数，出次数に比例した頻度に従う．）

2.3 冪級数の近似精度

前項の表 1 に示した先のネットワークに対して，モデルによる冪級数の近似の制度を調べた．冪級数は頂点とそれがもつ辺数の冪分布の傾きである．例えば[論文の引用，neuro]の場合では，冪分布は図 1，図 2 のようになる．この冪級数は図 1，図 2 の近似直線の傾きを表している．表 3 にすべてのネットワークモデルと実際のネットワークの冪級数を比較した．[論文の引用，SPIRES]と[電力網]については若干異なる値となった．しかし，次数の大きものと小さいものを除くと実際のネットワークと近い値となるので，実際の結合関係を表している．他の 6 つのネットワークでは，非常に近い値となっている．

表 3　実測とモデルの冪級数の比較

ネットワーク名	In-degree γ_{in}		Out-degree γ_{out}	
	実測	モデル	実測	モデル
論文の引用, SPIRES	1.2	1.80	1.2	1.74
論文の引用，neuro	2.1	2.11	2.1	2.07
映画俳優の共演	2.3	2.32	2.3	2.39
ルータ	2.48	2.51	2.48	2.50
電力網	4.0	2.85	4.0	2.72
WWW,Kumar'99	2.1	2.21	2.38	2.29
WWW,Albert'99	2.1	2.17	2.45	2.48
WWW,Broder'00	2.1	2.21	2.72	2.75

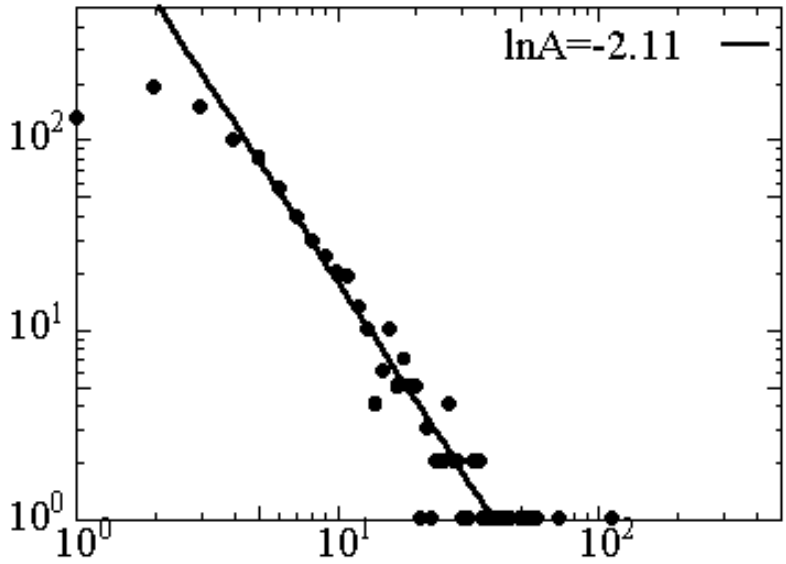


図 1　[論文の引用，neuro]　indegree

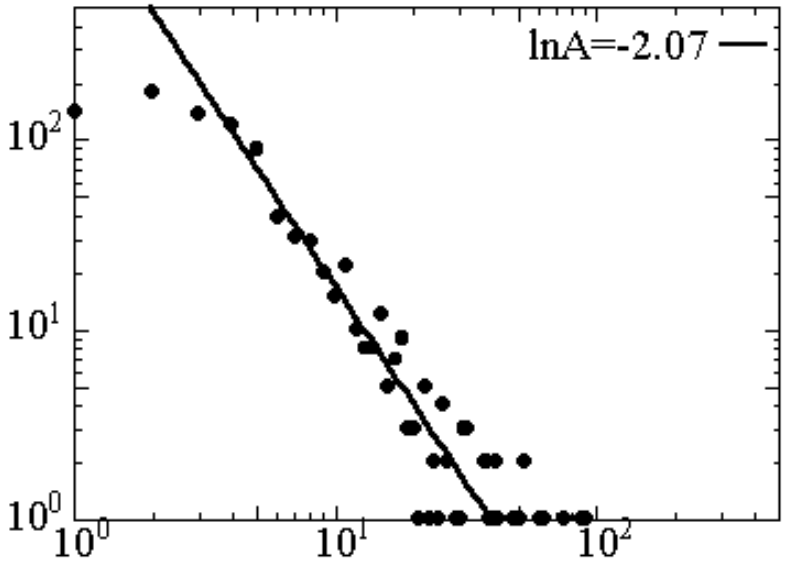


図 2　[論文の引用，neuro]　outdegree

2.4 スモールワールドネットワークとの共存

スモールワールドネットワークはランダムネットワークのように任意の頂点間の距離が小さく，正則ネットワークのようにクラスター化されているという２つのネットワークの中間的特徴をもっている．したがって，スモールワールドネットワークは小さな平均経路長と大きなクラスター係数を持つことが明らかになっている．つまり，スモールワールドネットワークの平均経路長とクラスター係数は

$$L_{sw} \sim L_{rand} << L_{reg}, \quad C_{sw} \sim C_{reg} >> C_{rand}$$

L:平均経路長 C:クラスター係数

sw:スモールワールドネットワーク

rand:ランダムネットワーク　reg:正則ネットワーク

の関係を満たしている．したがって，先に生成したスケールフリーネットワークモデルが上記の関係を満たし，スモールワールドネットワークの特徴を持つかどうかを明らかにする．

2.4.1 平均経路長の比較

2.4.1.1 平均経路長について

平均経路長というのは任意の頂点間の距離d_{ij}の平均である．ここでd_{ij}というのは頂点 i と頂点 j を最短でつなぐのに必要な辺数である．それらの平均を計算することによって頂点 i の平均経路長L_iを求めることができる．したがってすべての頂点にたいしてL_iを求めそれらを平均することによって，ネットワークの平均経路長Lが得られる．以上の操作は以下の式にまとめられる[3].

$$L \equiv \frac{1}{N}\sum_{i=1}^{N} L_i, \qquad L_i \equiv \frac{1}{N-1}\sum_{j \neq i} d_{ij}$$

例えば図 3 では，頂点 A からの最短でつながる辺数は表 4 のようになる．

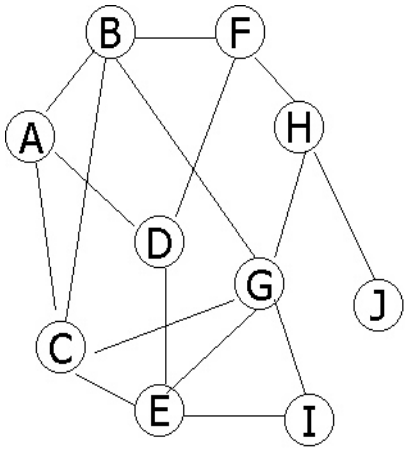


図 3　ネットワークの例

表 4　頂点 A からの最短で結ぶ辺の数

	B	C	D	E	F	G	H	I	J
d	1	1	1	2	2	2	3	3	4

表 4 より L_A=19/9 になる．同様にすべての頂点も計算でき、表 5 のようになる．

表 5　各頂点の平均経路長

	A	B	C	D	E	F	G	H	I	J
L	19/9	15/9	16/9	17/9	15/9	16/9	13/9	16/9	19/9	21/9

これらを平均して L =167/81=2.06 となる．

2.4.1.2 平均経路長の結果

平均経路長の比較では，実際のネットワークの平均辺数に近いネットワークモデルで比較を行った．ランダムネットワークと正則ネットワークはモデルのサイズと平均辺数を使って理論的に求めた値です．平均経路長を求める式は以下のようになる[3].

$$L_{rand} \sim \frac{\log N}{\log \bar{k}}, \qquad L_{reg} = \frac{N(N+\bar{k}-2)}{2\bar{k}(N-1)} \sim \frac{N}{2\bar{k}}$$

（α, β）モデルによるスケールフリーネットワークモデル（SF），ランダムネットワーク，正則ネットワークのそれぞれの平均経路長を表 6 に示す．

表 6　平均経路長の比較

ネットワーク名	SF	ランダム	正則
論文の引用，neuro	2.98	2.82	42.9
ルータ	5.42	6.70	127
電力網	5.55	6.61	134
WWW,Kumer'99	3.47	3.45	66.6
WWW,Albert'99	4.14	4.40	97.8
WWW,Broder'00	3.46	3.38	64.5

表 6 の値を比較すると，正則ネットワークより非常に小さいく，ランダムネットワークに近い値である．したがって，

$$L_{sf} \sim L_{rand} << L_{reg},$$

となる．このように，スケールフリーネットワークモデルはランダムネットワークのように，小さな平均経路長を持ち，スモールワールドネットワークの１つの特徴を満たす．

2.4.2 クラスター係数の比較

2.4.2.1 クラスター係数について

クラスター係数はネットワークの集まり具合を表すものである．ここでは，クラスターというのは，3 つの頂点がそれぞれ 1 本の辺を介してつながっている場合である．つまり，三角形の各頂点に頂点がある場合を示す．頂点 i のクラスター係数というものは，実際に存在しているクラスターの数E_iを，可能なクラスターの数で割ったもので，クラスター係数C_iで定義される．そして，すべての値を求め平均したものがネットワークのクラスター係数Cである．可能なクラスターの数は頂点 i が接続しているすべての頂点から 2 つ選びだす組み合わせの数である．以上の操作は以下の式のようにまとめられる[3].

$$C \equiv \frac{1}{N}\sum_{i=1}^{N} C_i, \qquad C_i = \frac{2E_i}{k_i(k_i - 1)}$$

k_i:頂点 i がもつ辺数

例えば図 3 の場合，各頂点のクラスター係数は以下の表のようになる．

表 7　各頂点のクラスター係数

	A	B	C	D	E	F	G	H	I	J
C_i	1/3	1/6	1/2	0	1/3	0	1/5	0	1	0

これらを平均してC=76/300=0.253 となる．

2.4.2.2 クラスター係数の結果

平均経路長と同様に実際のネットワークに近い平均辺数を持ったネットワークモデルで比較を行った．ランダムネットワークと正則ネットワークも平均経路長と同様に理論的に求めた．クラスター係数を求める式は以下のようになる[3].

$$C_{reg} = \frac{3(\bar{k}-2)}{4(\bar{k}-1)}, \qquad C_{rand} \sim \frac{\bar{k}}{N}$$

ネットワークモデル，正則ネットワーク，ランダムネットワークのそれぞれのクラスター係数を表 8 に示す．

表 8　クラスター係数の比較

ネットワーク名	SF	正則	ランダム
論文の引用，neuro	0.508	0.679	0.0116
ルータ	0.135	0.292	0.0039
電力網	0.094	0.310	0.0037
WWW,Kumer’99	0.377	0.632	0.0075
WWW,Albert’99	0.252	0.549	0.0051
WWW,Broder’00	0.392	0.638	0.0078

表 8 の値を比較すると，スケールフリーネットワークはランダムネットワークよりも非常に大きく，正則ネットワークと近い値である．したがって，

$$C_{sf} \sim C_{reg} >> C_{rand}$$

となる．このようにスケールフリーネットワークモデルは正則ネットワークのように大きなクラスター係数を持ち，スモールワールドネットワークの１つの特徴を持つ．

2.5 ネットワークモデルの検証のまとめ

(α, β) モデルから実測値に対応した冪乗分布を持つネットワークモデルが生成できることが明らかになった．特に$\gamma_{in}, \gamma_{out}$が２～３の間では，非常に良い結果を示した．

また，(α, β) モデルによって生成したスケールフリーネットワークモデルはランダムネットワークのように平均経路長が小さく，正則ネットワークのようにクラスター係数が大きいことが明らかになった．したがって，スケールフリーネットワークモデルがスモールワールドネットワークの特徴を持つことが明らかになった．

第　3　章

3 情報伝搬

全章で，(α, β) モデルを用いれば，実際のネットワークの特徴であるスケールフリーネットワークとスモールワールドネットワークを持ったネットワークモデルが生成できることが明らかになった．本章では，そのネットワークモデルを用いて，ネットワークの基本特性となる情報の広がりについて検討を行う．特に，１つの頂点からいくつの頂点に情報を伝搬すればネットワーク全体に情報が蔓延するか（広がりきるか）を調べた．

3.1 ネットワークモデル

ネットワークモデルは前章の冪級数の近似で用いたネットワークモデル 8 種類で行った．基本となるパラメータは前章の表１に示す．

3.2 情報伝搬の方法

表１の８種類のネットワークに対して，情報伝搬特性の解析を行う．ここで，総頂点数は 1000 として，各頂点を未伝搬状態と伝搬済み状態の 2 値として，伝搬済み状態の頂点から隣接しているすべての頂点に情報が伝搬していき，情報が伝搬された頂点は未伝搬状態から伝搬済み状態になる．伝搬済み状態の頂点の増加の様子やハブとなる頂点１つからどれだけの頂点に情報が伝搬されるかを解析する．総頂点の 90%以上の頂点が伝搬済み状態になったときに，ネットワーク内に情報が蔓延したとする．

3.3 情報伝搬の結果

それぞれのネットワークについて，平均辺数を変化させてどれだけの頂点に情報が伝搬するか調べた．ネットワークは確率的に生成されるため，平均辺数$\bar{k}$が同じになることはないので，(1.3~1.5)，(2.5~3.0)，(4.6~5.0)，(5.8~6.2)の 4 つの範囲で頂点数の増え方を比較した．グラフの値は 100 回の情報伝搬を平均したものである．図 4 から図 7 のすべての範囲において，[論文の引用，SPIRES]だけは伝搬済み状態の頂点の数が極端に少ない．また，入次数が$\gamma_{in}=2.1$で出次数がことなるネットワークでは，出次数が大きいほど伝搬済み状態の頂点は少ない．

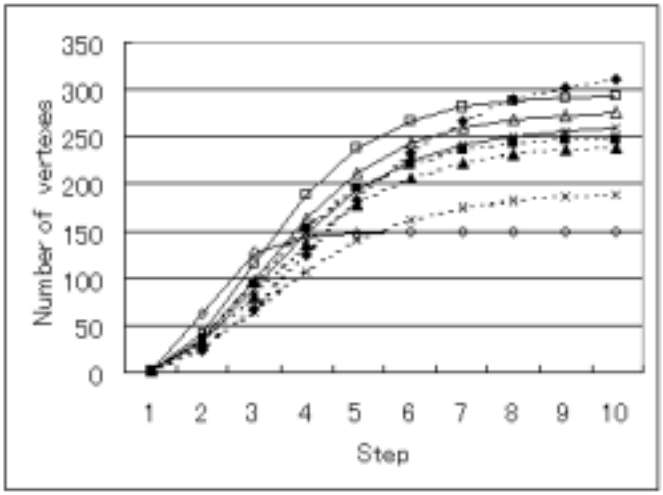


図 4 平均辺数$\bar{k}$ = 1.3~1.5 に対する頂点数の増加

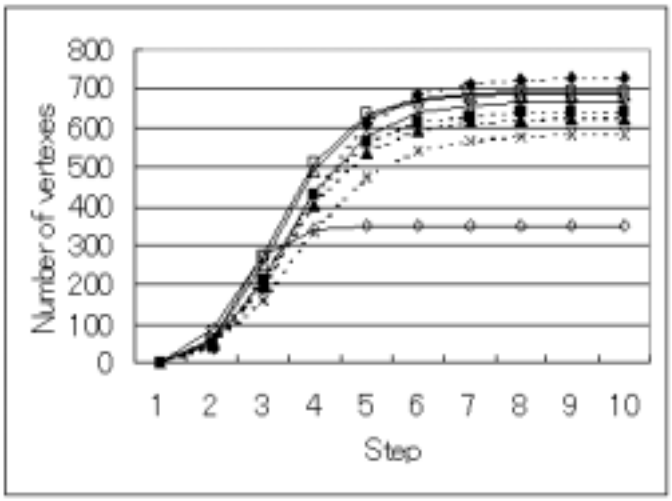


図 5　平均辺数$\bar{k}$＝2.5~3.0 に対する頂点数の増加

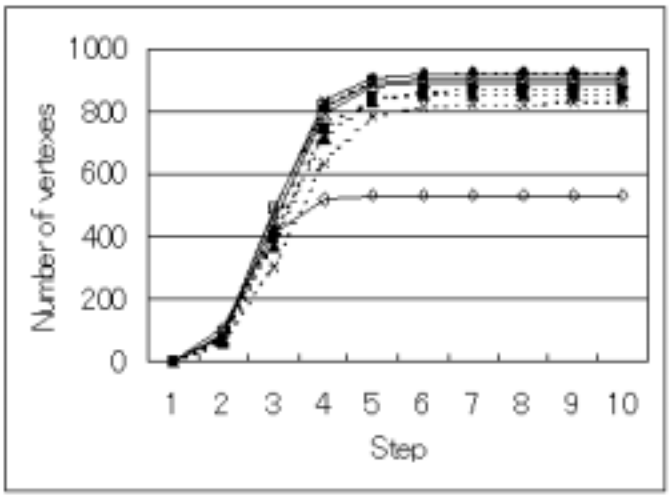


図 6　平均辺数$\bar{k}$＝4.6~5.0 に対する頂点数の増加

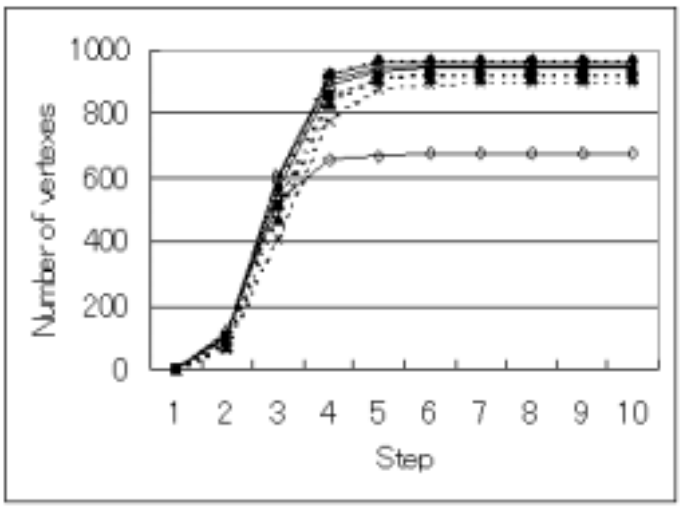


図 7 平均辺数$\bar{k}$＝5.8~6.2 に対する頂点数の増加

ランダムネットワークと WWW ネットワークの場合において，蔓延状態になるときの平均辺数が明らかにされており，ランダムネットワークでは平均辺数 3 本，web ネットワーク($\gamma_{in}=2.1$，$\gamma_{out}=2.38$)では平均辺数 4.8 本で情報が広がりきり，90%以上の頂点が伝搬済み状態になった[9]．本研究では，蔓延状態になる確率が 90%以上になるときの平均辺数を調べた．それぞれのネットワークにおける平均辺数を表 9 に示し，その時の頂点数の増加の様子を図 8 に示す．モデルの平均辺数の値，グラフの値は 100 回の情報伝搬の平均値である．

表 9　蔓延率 90 以上の時の平均辺数と実際のネットワークの平均辺数

ネットワーク名	平均辺数		蔓延確率
	実測	モデル	
論文の引用，SPIRES	173	11.3	100
論文の引用，neuro	11.54	5.23	100
映画俳優の共演	28.78	4.91	98
ルータ	2.57	4.63	95
電力網	2.67	4.40	99
WWW,Kumar'99	7	5.64	99
WWW,Albert'99	4.51	5.56	97
WWW,Broder'00	7.5	5.94	96

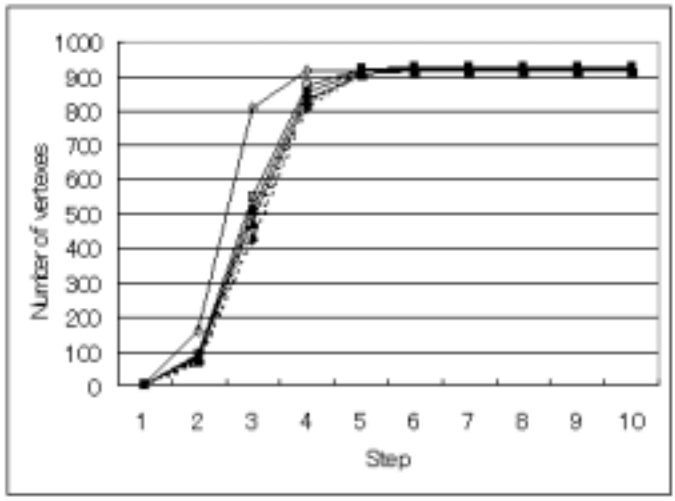


図 8　蔓延率が 90 以上のときの頂点数の増加

表 9 のように，[論文の引用，SPIRES]以外は平均辺数が 4.5～6.0 本であれば情報が広がりきることが明らかになった．[論文の引用，SPIRES]，[映画俳優の共演]などはモデルの平均辺数の方が小さく情報が広がりきらないのに対して，[ルータ]，[電力網]などはモデルの方が大きく情報が広がりきることが明らかになった．

伝搬済みの頂点の増加については，[論文の引用，SPIRES]だけが 3 ステップで蔓延状態になるが，他のネットワークは 4 ステップで蔓延状態になる．これは，平均辺数の違いによるものではないかと思われる．

3.4 情報伝搬のまとめ

[論文の引用，SPIRES]を除いて，冪級数が違っても情報伝播する頂点数やステップ数にあまり変化は見られなかった．[論文の引用，SPIRES]の場合，冪級数が小さいのが原因ではないかと思われる．また，平均辺数が 4.5～6.0 本あればネットワーク全体に情報が蔓延し，蔓延するにはたった 3，4 ステップだけで十分である．

第 4 章

4 コンピュータウィルスの伝染

第２章，３章でネットワークモデルの構造や基本特性となる決定論的情報伝搬に関して検証を行った．第４章では，(α, β) モデルを用いて生成したネットワークモデルを使って，決定論的な情報伝搬を応用した確率論的な状態遷移に基づく，コンピュータウィルスの伝染に関して検証を行う．その際, 実際のメールネットワークを使って，ウィルスが伝染する範囲，消滅するまでにかかる時間などを調べる．

4.1 メール送受信ネットワークモデル

先行研究では，ランダムネットワーク，正則ネットワークなどのさまざまなネットワークでウィルスの伝染に関する研究が行われている．しかし，このようなネットワークは実際のウィルスが伝染するネットワークとは接触関係を表している．したがって，実際の接触関係に近いネットワークモデルを生成する必要がある．

本研究では，e-mail によって広がるウィルスの伝染について検討を行うため，実際の接触関係を考慮したネットワークモデルを生成する．1 日のメール送受信頻度を図 9 に示す[9]．それらのメール送信数，受信数の両対数プロットを図 10，11 に示す．両図からメール送受信数は冪乗分布に従うことが判明し，受信の冪級数は－1.9，送信の冪級数は－2.5 となった．

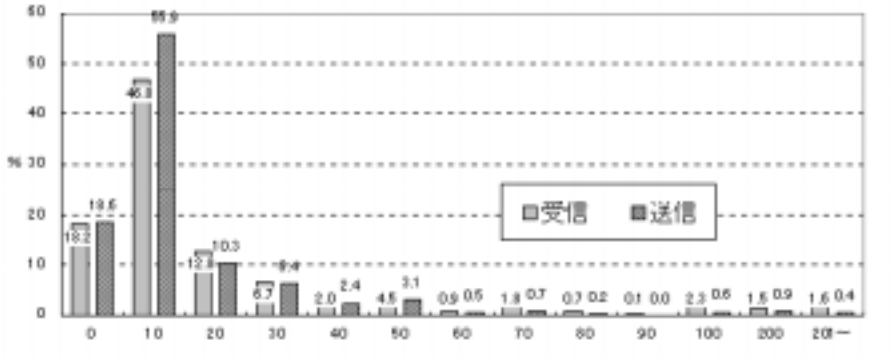
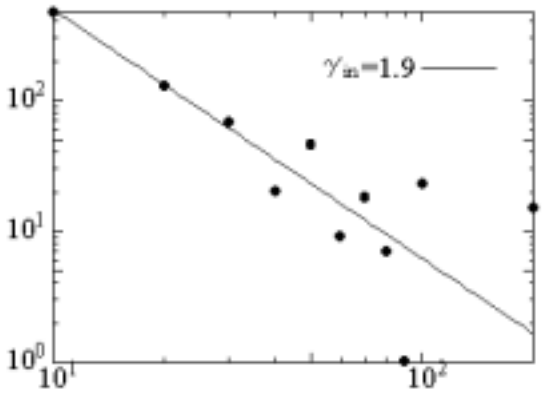


図 9　1 日のメール送受信頻度



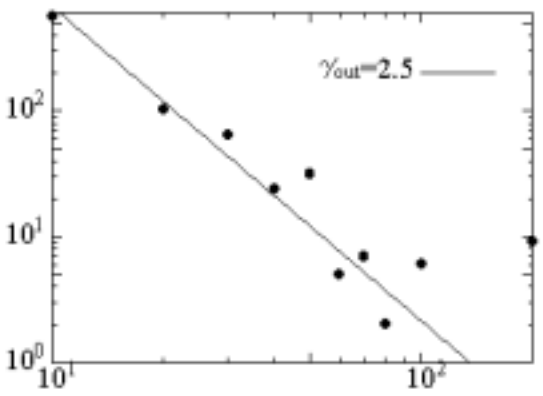
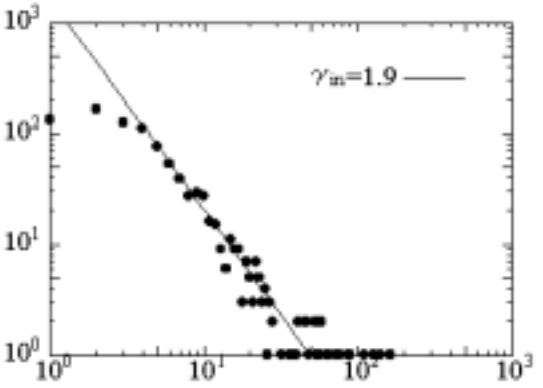


図 10　メール受信頻度　　**図 11　メール送信頻度**

この冪級数に従うメール送受信ネットワークモデルを(α, β)モデルを使って生成し, 図 12，13 に示す.



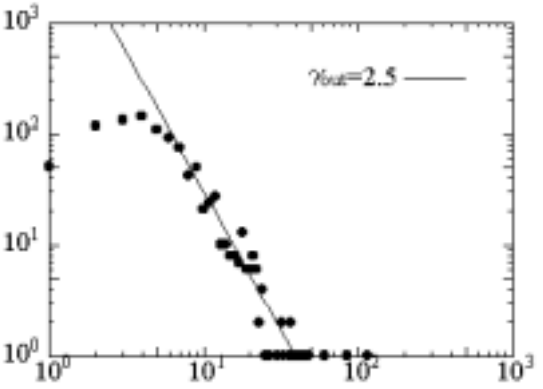


図 12　モデル受信分布　　**図 13　モデル送信分布**

メール送受信数の平均値はいろいろな研究所, アンケートで調査されている[9]. 本研究では, 情報通信総合研究所による 1999 年 4 月に行われたアンケート結果[10]の 6.2, 14.8 に近くなるようにした.

4.2 ウィルスの伝染モデル

本研究で用いたウィルス伝染モデルは，未感染状態，潜伏状態，増殖状態，免疫状態の４つの状態を遷移する SHIR モデルである．ネットワークにおける頂点を PC，辺をメールの送受信とする．それぞれ PC の状態の特徴は次に示す．

[未感染状態](Susceptible)
ウィルスが送られていない状態．また，PC の初期状態でもある．隣接する PC１つが感染するとウィルスが１つ送られてくる．ウィルスが届いたときにウィルスを発見して，駆除すれば免疫状態になる．もし，ウィルスが来たことに気づかなければ潜伏状態になる．

[潜伏状態](Holder)
ウィルスが存在するが，感染はしていない状態．ウィルスを発見して駆除して免疫状態になるか，ウィルスを誤って実行して増殖状態になるまで潜伏状態のままである．隣接している PC が新たに増殖状態になるとウィルスが届き，ウィルスがたまっていく．しかし，ウィルスの総数はメール受信する PC の数が最大である．

[増殖状態](Infected)
潜伏状態からウィルスを実行して，感染してしまった状態．感染した時に，隣接している PC にウィルスを１つづつ伝染させる．ウィルスを伝染させた後は，ウィルスが発見されて駆除されるまで増殖状態のままである．

[免疫状態](Recover)
ウィルスを駆除して，ウィルスが存在しな状態．一度免疫状態になると，ウィルスを完全に遮断することができ，PC の最終状態である．

各 PC における遷移状態を図 14 に示す．１つのウィルスに対してウィルスを実行して感染させてしまう確率λ（実行確率），ウィルスを発見して消滅させる確率δ（発見確率）とする．ウィルス保有数ｎは最大で近接する PC の台数とする．

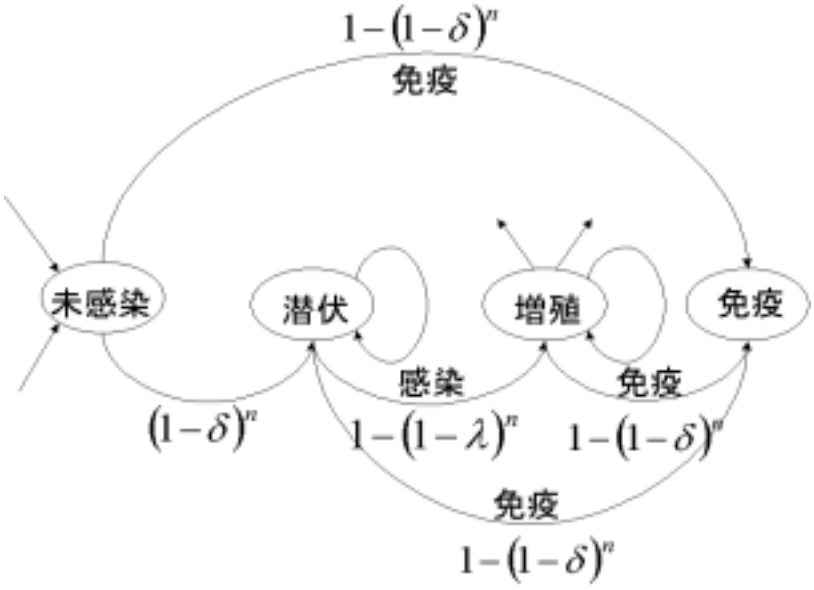


図 14　SHIR モデルの状態遷移

4.3 サイズと平均辺数による違い

SHIR モデルを使って，平均辺数とネットワークのサイズの違いによって，次の 3 つの項目について違いがあるかを調べた.

1．感染する PC の総数,
2．ウィルスが届けられた PC の総数
3．ウィルスが完全に消滅するまでにかかる時間

4.3.1サイズによる違い

ネットワークの総頂点数 N を(500,1000,2000)の３つの場合で比較を行った．このときの平均辺数$\bar{k}$はそれぞれ(6.7, 6.8, 6.9)である．それぞれのサイズにおいて, (λ, δ)を 10%づつ変化させたときに，３つの項目がどのように変化するかを図に示す．図 15 に感染する PC の割合について，図１６にウィルスが訪れたＰＣの割合，図 17 に消滅するまでにかかったステップ数を示す. 感染割合の縦軸は感染した総数をδ=0 のときの感染数で割った値，ウィルスが訪れたＰＣの割合の縦軸は感染数と同様に，ウィルスが訪れた総頂点数をδ=0 のときのウィルスが訪れた総頂点数で割った値とした．すべてのサイズで，発見確率が大きくなるにつれて被害は減少し，50%になると 10%の 3 分の 1，90%になると 6 分の１なる．サイズよる感染割合，ウィルスが訪れる割合の差はあまりみられない．ステップ数は図 17 のように，実行確率に依存しない．消滅するまでにかかるステップ数は発見確率が小さいときはサイズが大きいほどステップ数が大きくなるが発見確率が大きくなるとほとんど差がなくなる．このようにサイズによる差はほとんどなかった．

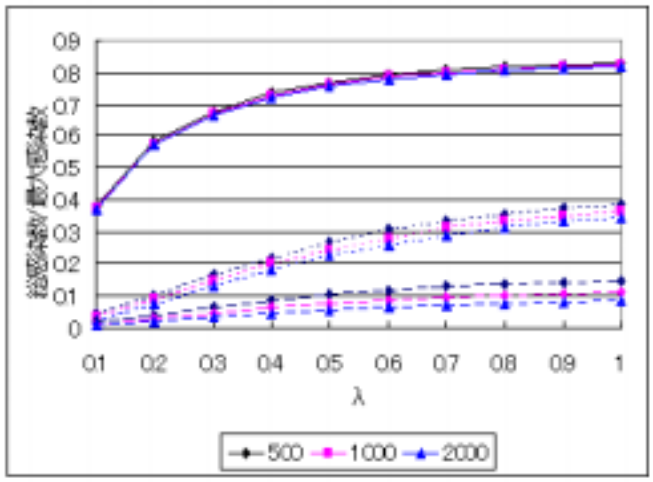


図 15　頂点数 N に対する感染割合

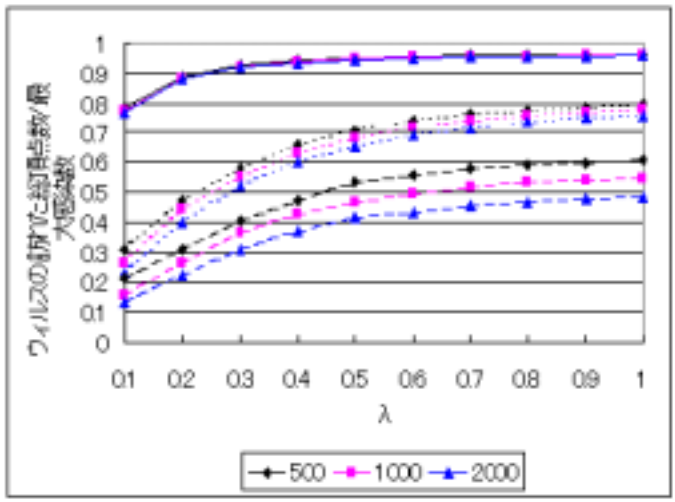


図 16　頂点数 N に対するウィルスが訪れた頂点の割合

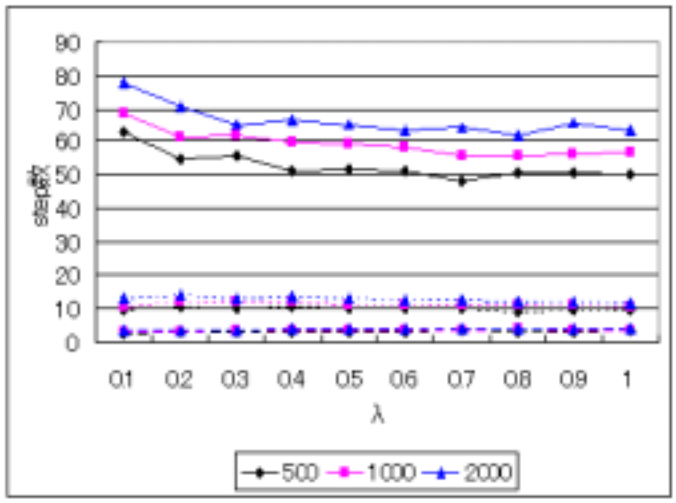


図 17　頂点数 N に対する Step 数

（直線：δ=10%，点線：δ=50%，破線：δ=90%）

4.3.2　平均辺数の違い

ネットワークの総頂点 N を 1000 にして，平均辺数$\bar{k}$を(6.8，10.2，15.1，19.7)の４つの場合について比較をした．それぞれの平均辺数における図を示す．感染割合に関しては，ほとんど差がなかった．しかし、ウィルスの訪れた頂点の割合は平均辺数が 19.7 の場合，発見確率が高くなっても頂点の割合は高く，発見確率が 90%のときは，6.7 よりも 30%以上多かった．発見確率が大きい場合は差が見られないが，小さい場合は，平均辺数が大きい方が早く消滅した．平均辺数が大きい場合は早くウィルスが広まって，早く消滅することが明らかになった．

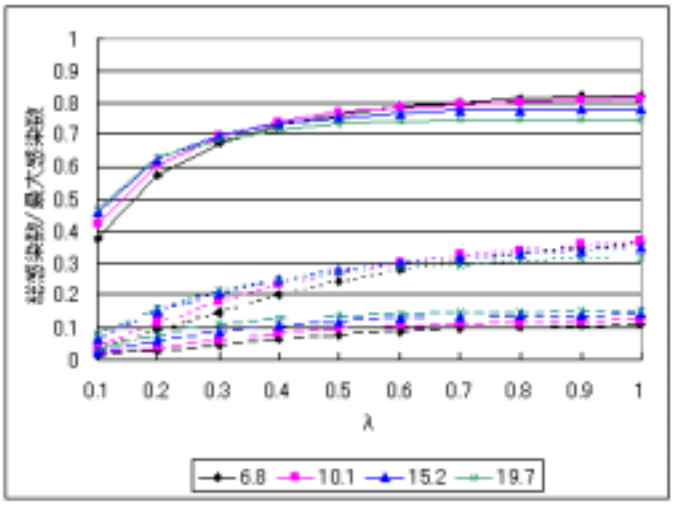


図 18　平均辺数$\bar{k}$に対する感染割合

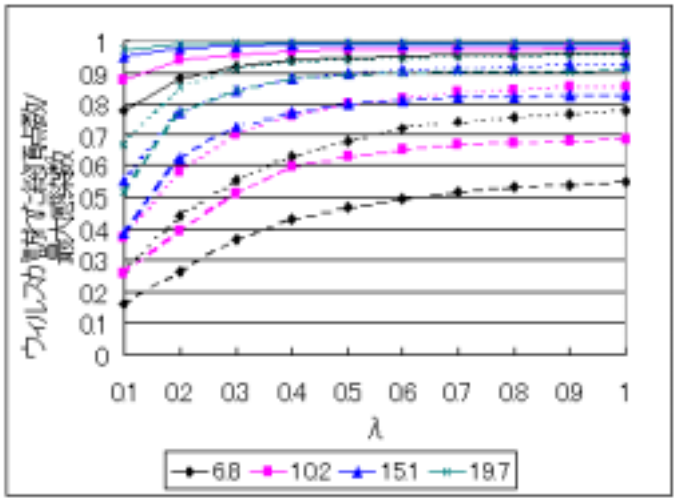


図 19　平均辺数 $\bar{k}$ に対するウィルスが訪れた頂点数

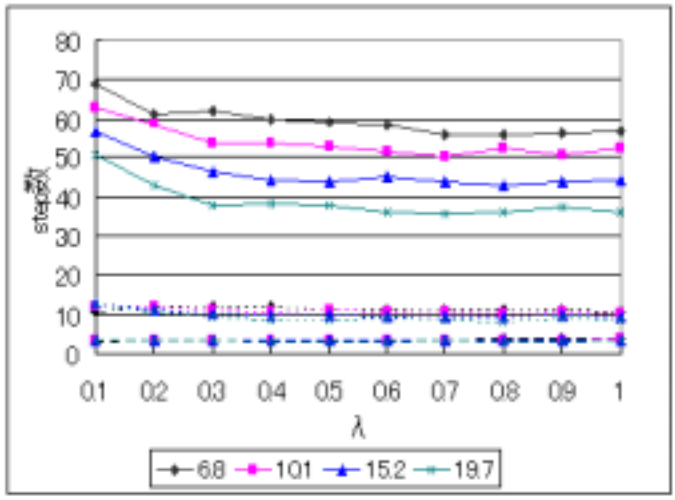


図 20　平均辺数 $\bar{k}$ に対する Step 数

4.4 成長するネットワークによる解析

実際の社会では，e-mail を使う人は増加しており，これからも増加していくと考えられている．したがって本研究でも，ウィルスの伝染と共に成長するネットワークを考え，ウィルスの伝染について調べた．

本研究では，ウィルスが１回伝染する度に，1%ずつ増加する指数成長のネットワークと 50 個ずつ増加する線形成長のネットワークについて，ウィルスの伝染を調べた．初期のネットワーク内の PC を 400 として，400step ごとにウィルスの伝染を行うとネットワーク規模は，指数成長で 18934，線形成長で 20350 となる．その際，初期感染 PC をランダムに５つ選んでウィルスの伝染を行った．

4.4.1 ウィルスの広がり

指数成長，線形成長，成長しないネットワークの step ごとのウィルスが存在する PC の数の推移を図 21 に示す. 成長しないネットワークは一端増加してから単調減少する．しかし，成長するネットワークは一端増加して，ほぼウィルスが消滅してからまた増加するという再流行現象を起こすことが明らかになった．再流行現象はサイズが固定された既存のモデルでは，説明することがすることができなかったが，現実には広く共通して観測される．

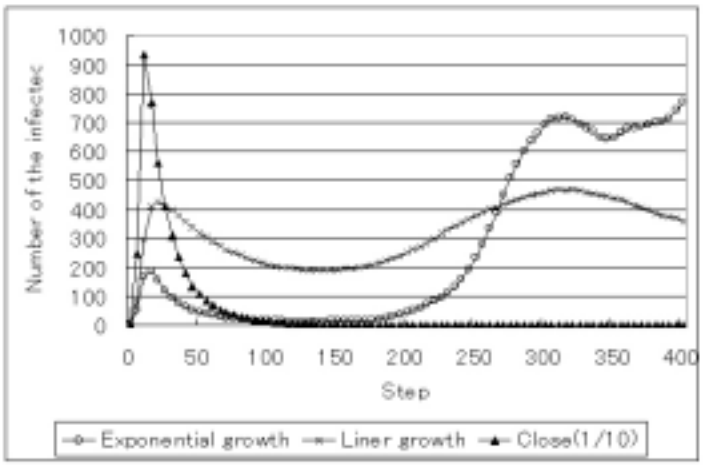


図 21　ウィルスが存在する PC の推移

4.4.2 （λ，δ）の違い

成長するネットワークにおいては，図 21 のようにウィルスの存在する PC の数は推移することがある．このような再流行現象が起こる実行確率λと発見確率δについて調べた．

発見確率δを 4%，実行確率λを，10%，20%，30%の３つの場合について比較を行った．１つのネットワークに対して，初期の感染した PC が異なる場合を 10 回，それを 10 個のネットワークで行い，初期の感染した PC とネットワークの異なる 100 個の場合について，再流行現象がおきた回数を調べ，その結果を表 10 に示す．

表 10　実行確率λに対する回数

	10%	20%	30%
回数	50	44	47

10%～30%ではあまり差がなく，ほぼ 50%の割合で図のような現象が起こる．

次にλを 10%，δを 4%，5%，6%の３つの場合について比較を行った．この場合もλの比較と同様に 100 個の場合についての回数を調べ，その結果を表 11 に示す．

表 11　発見確率δに対する回数

	４%	5%	6%
回数	50	13	1

δの場合は 1%増加するだけで回数が激減した．4%のときが 50%であったのに，６%になるとたった 1%ほどになってしまう．再流行現象が起きるには，ウィルスを発見する確率が非常に重要になる．

4.5 ハブとランダムの免疫化

M.E.J.Newman らによる成長しないネットワークモデルによるウィルスの伝染による解析で，ランダムに PC に免疫するよりも，ハブとなる PC に優先的に免疫した方がウィルスの被害が抑えられることが明らかにされている[7].

したがって，成長するネットワークについても，ハブに免疫する場合の方がランダムに免疫する場合よりもウィルスの伝染を抑えられるかどうかを調べた．この時のλは 10%、δは４%とし，λやδを変化させたときと同様に 100 個の場合について調べた.

それぞれの免疫化については 30step ごとに免疫割合に従って，免疫を行った。ハブに免疫する場合，出次数が多い順で免疫を行う．例えば免疫割合が 10%であれば，すでに免疫状態になっている場合が考えられるが，免疫状態になっている場合も含めて出次数が多い順の 10%が必ず免疫状態になるようにした．一方，ランダムに免疫する場合は免疫しようとした PC の合計が 10%になるようにした.

4.5.1 指数成長による比較

指数成長におけるハブに免疫した場合と，ランダム免疫した場合の 400step たったときにウィルスが生き残った回数を調べた結果を表１2 に示す．生き残った場合の総感染数とウィルスが訪れた総頂点，ハブに免疫する場合とランダムに免疫する場合のウィルスが存在する PC の推移をそれぞれ図に示す.

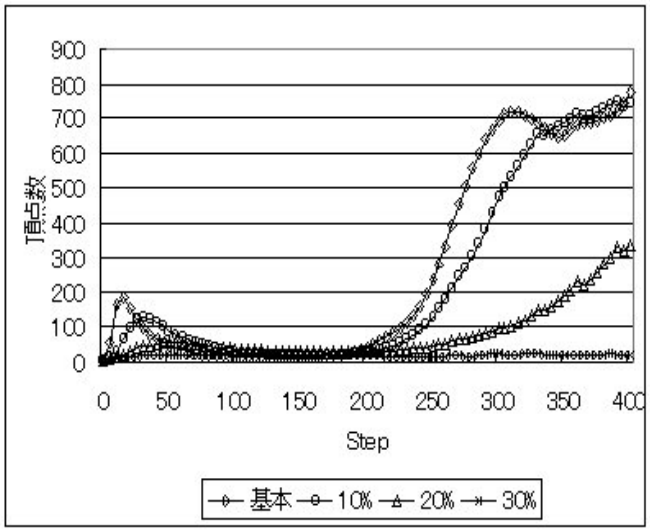


図 22　ハブ免疫化によるウィルスが存在する PC の推移

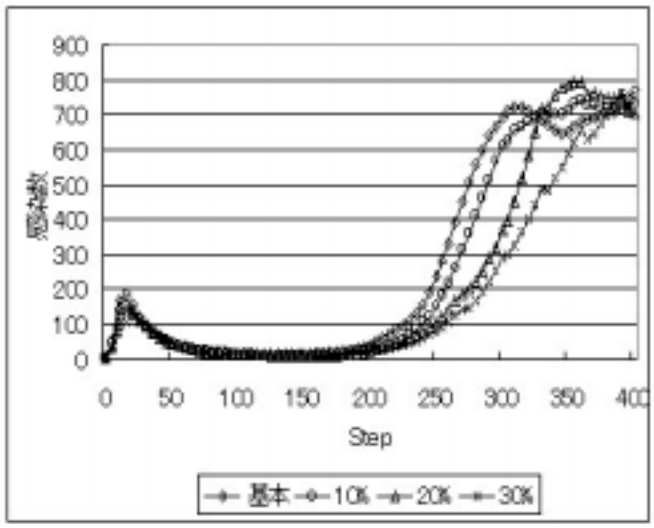


図 23　ランダム免疫化によるウィルスが存在する PC の推移

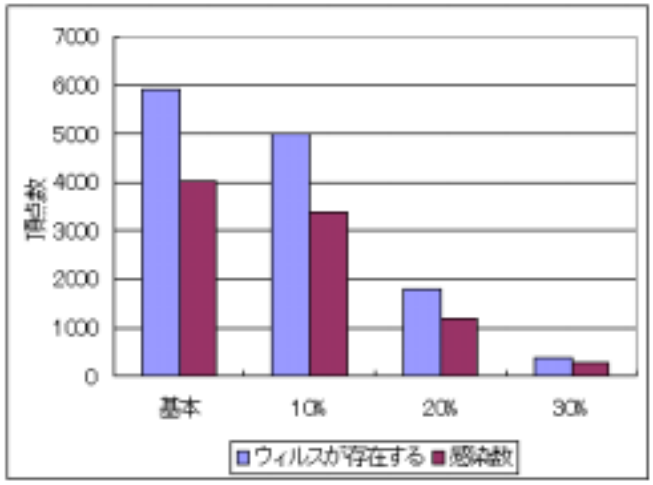


図 24　ハブ免疫化による総頂点数

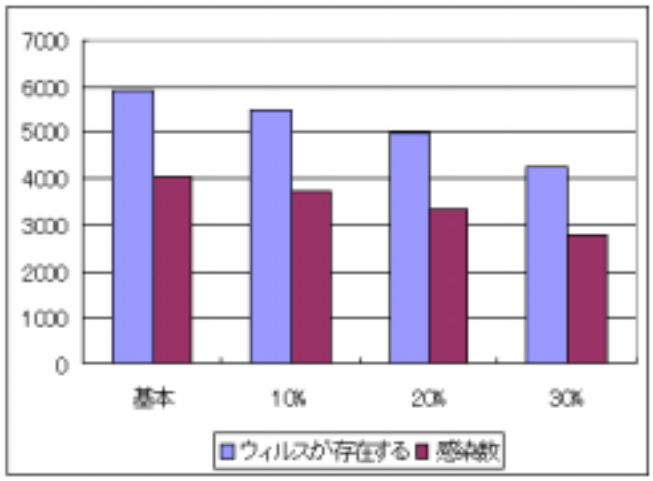


図 25　ランダム免疫化による総頂点数

表 12　指数成長の生き残り回数

	基本	10%	20%	30%
ハブ	50	77	73	10
ランダム	50	38	29	23

ハブに免疫する場合はウィルスの発見だけによる免疫（基本）よりウィルスが生き延びてしまう回数が増えている．しかし，これはハブに免疫するとウィルスの広がりを抑えてしまったために長く生き残ってしまったと考えられる．ハブに免疫する場合

は，ウィルスが長く生き残ってしまうが，図 22 や図 24 のように免疫割合が 20%になると免疫しないもの 4 分の 1 から 3 分の 1 に抑えることができ，30%になると 10 分の 1 以下に抑えることができる．ランダムに免疫する場合は，短い時間でウィルスを消滅させることができる．これは一端増加した後，ウィルスが存在する PC が数個しかないので，偶然その頂点を免疫してしまったり，その回りを免疫させてしまうことではないかと考えられる．しかし，図 23 や図 25 のようにウィルスが生き残ってしまった場合は免疫しないものと被害はあまり変わらない．

4.5.2 線形成長による比較

指数成長と同様に回数を表に，生き残った場合の総感染数とウィルスが訪れた総頂点，ハブに免疫する場合とランダムに免疫する場合のウィルスが存在する PC の推移をそれぞれ図に示す．

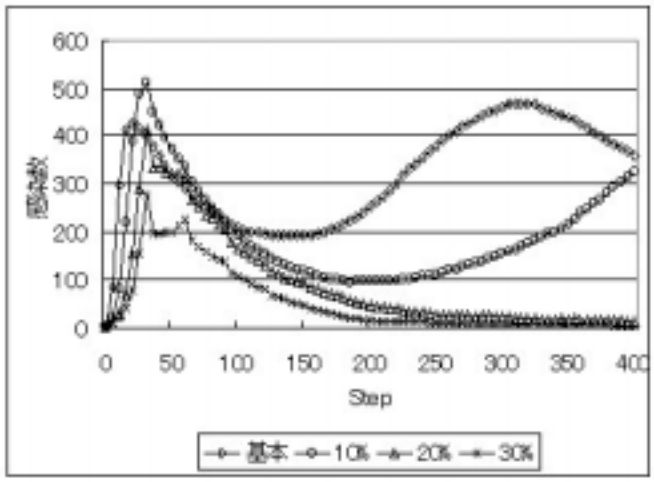


図 26　ハブ免疫化によるウィルスが存在する PC の推移

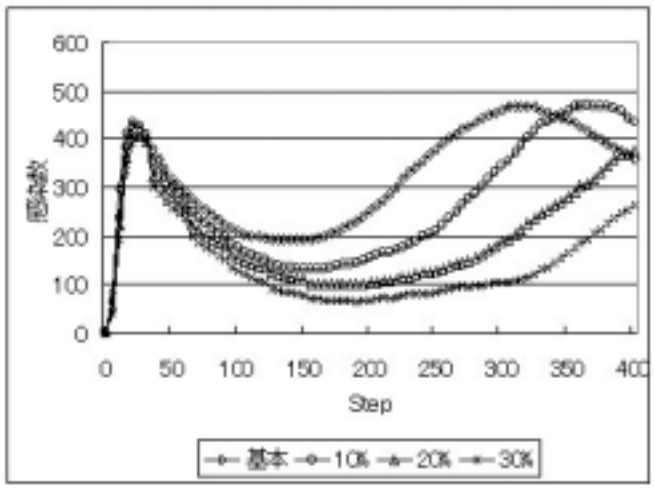


図 27　ランダム免疫化によるウィルスが存在する PC の推移

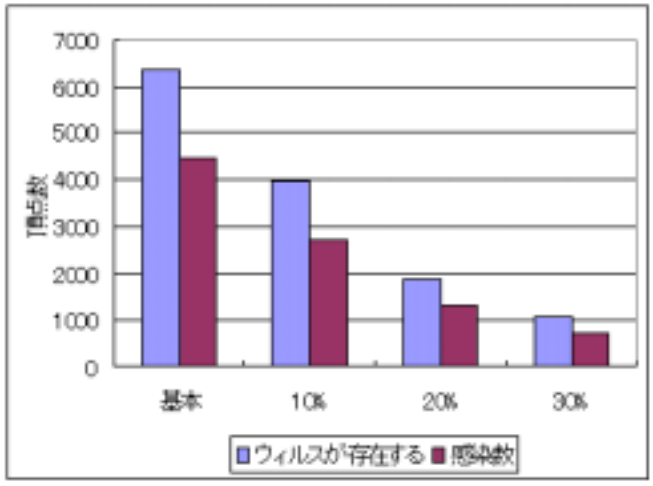


図 28　ハブ免疫化による総頂点数

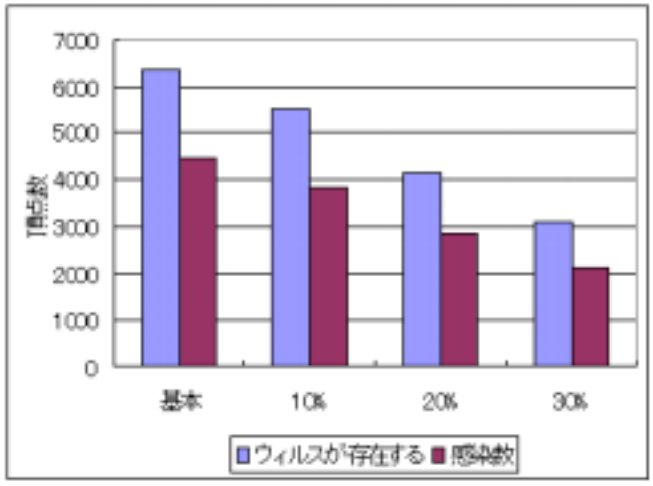


図 29　ランダム免疫化による総頂点数

表 13　線形成長の生き残り回数

	基本	10%	20%	30%
ハブ	100	92	35	6
ランダム	100	96	92	70

線形成長の場合，ハブに免疫する方がランダムに免疫するよりもウィルスを絶滅させることができ，被害を抑える事ができる．これは線形成長の場合，ウィルスが存在する PC は一端増加した後，半分ぐらいにしか減少しないので，ランダムに免疫する場合では有効に免疫できないと考えられる．

4.6 コンピュータウィルスの伝染のまとめ

ネットワーク内のメール送受信量が増加すると，ウィルスは発見されやすくなり，消滅するまでにかかる時間が短くなるが，被害が大きくなってしまうことが明らかになった．

また，新規ユーザを伴う成長するネットワークでは，従来の免疫学，研究では説明することができなかった再流行現象が起きることを明らかにした．再流行が起きる場合にも重点的にハブを免疫することによって被害を抑えられることが明らかになった．

第５章

5 まとめ

本研究では，実際のネットワークにおける現象を解析するために，実際のネットワーク構造に近いネットワークモデルを生成できるかどうかを検証し，そのネットワークモデルを用いて情報伝播とコンピュータウィルスの伝染特性を調べた．

ネットワークモデルは(α，β)モデルによるスケールフリーネットワークモデルを用い，そのスケールフリーネットワークモデルがスモールワールドネットワークという特徴を持つことが明らかになった．したがって，このネットワークモデルは実際のネットワークに近いモデル構造を持つ．

さらに，決定論的な情報伝搬に関しては，論文の引用，ルータ，WWW などの実測値に基づく，ネットワークについて，平均辺数が 4.5～6.0 本あればネットワーク全体に情報を広がらせることができることを明らかにした．その結果から実測値の平均辺数と比較すると実際のネットワークは状態が蔓延しやすい状態であることも明らかにした．

確率的な状態遷移に基づくウィルスの伝染特性に関しては，メールの使用頻度が多く，ウィルスが広範囲に広がる方が，ウィルスが消滅するまでに時間がかからないことが明らかになった．また，従来の免疫学における閾値理論では説明することができない，再流行現象が起こることを示した．再流行が起きる場合でも，ハブとなる PC を免疫することによって，被害を抑えることができることが明らかになった．

将来メールを使用する頻度が増加すると考えられるので，これらの結果から被害がますます増加していくものと考えられる．一方，ネットワーク内のハブとなる PC にアンチウィルスの導入やアップデートを優先的に行うことで，ネットワーク全体の被害を抑えることが可能となる．

参 考 文 献

[1] D.J.Watts and S.H.Strogatz, “Collective dynamics of ‘small-world’ networks,” Nature, vol.393, pp.440-442, (1998).

[2] A.-L.Barabási, R.Albert, and H.Jeong, “Mean-filed theory for scale-free random networks,” Physica A, vol.272, pp.173-187, (1999)

[3] 相馬亘, 下原勝憲, “スモールワールドネットワークの役割,” Technical report of IEICE, NGN2001-12

[4] R.Albert and A.Barabási “Statistical Mechanics of Complex Networks,” arXiv:cond-matt/0106096v1, (2001).

[5] 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ィルスの拡散モデルの解析,”電子情報通信学会論文誌, vol.J84-D-I, No.5, pp.474-482, (2001).

[6] T.Okamoto and Y.Ishida, “A Performance of A Mobile Anti-virus System,” AROB7th’02, January, (2002).

[7] M.E.J.Newton S.Forrest and J.Balthrop, “Email networks and the spread of computer viruses,” Phys.Rev.E 66,035101, (2002).

[8] 三上利治, “日本人とインターネット生活,”, http://sophy.asaka.toyo.ac.jp/users/mikami/info&media/WIP2001.2.28.html

[9] R.Kumar, P.Raghavan, S.Rajagopalan, and A.Tomkins, “Extracting large-scale knowledge based from the web,” Proc. of the 25th VLDB Conf., pp7-10, (1999).

[10] 大久保和彦, 林幸雄, 蜷川繁, “Web 的ネットワークにおける情報伝搬率と速度,” 電情学論, vol.J85-D-I, no.2, pp241-244, (2002).

[11] 情報通信総合研究所, “郵送アンケートと第 6 回 MIN アンケートとの比較分析結果,” http://www.commerce.or.jp/

発 表 論 文

[1] 箕浦正人, 林幸雄, “ さまざまなネットワークの情報伝播と構造の特徴, ”電気化関係学会北陸支部連合大会，2002.

[2] 箕浦正人, 林幸雄, “ 電子メールによるコンピュータウィルスの伝染, ”情報処理学会全国大会，2003.

[3] 林幸雄，箕浦正人，松久保潤, “ ネットワーク成長によるメール型ウィルスの再流行と重点的なハブの免疫化の効果, ” 情報処理学会，数理モデル化と問題解決研究会，2003.