本資料について

- 本資料は下記の論文を基にして作成されたものです。文章の内容の正確さは、保障できないため、正確な知識を求める方は原文を参照して下さい。
- 著者：Cliff C. Zou, Don Towsley, Weibo Gong
- 論文名：Email Virus Propagation Modeling and Analysis

電子メールウイルス増殖のモデル化と分析

名城大学理工学部情報工学科
渡邊研究室
040427180　三根　健司

はじめに

- 1980年代初期、ウイルスは主にフロッピーディスク交換で広がっていた
- コンピュータネットワークとインターネットが1980年代後期から人気となるにつれ、ウイルスはいろいろな手段（例えばファイルのダウンロード、電子メール、ソフトウェアのセキュリティホールを悪用する、その他）でインターネット中に広がることができるように進化
- 電子メールウイルスとは電子メールメッセージを通してそれ自体のコピーを送ることによって増殖することができる限り、どんなコンピュータープログラムでも呼ばれる

電子メールウイルスの利点

- 電子メールを通してウイルスを送ることは、コンピュータオペレーティングシステムまたはソフトウェアにおいて少しのセキュリティホールも必要ない
- コンピュータを使うほとんどの人が電子メールサービスを利用
- 多数のユーザーには電子メールウイルスについてのほとんど知識がなく、彼らが受け取る大部分の電子メール（特に友人からの電子メール）を信頼する
- 電子メールは郵便局の手紙のような私有財産であり、通信員法または規定はエンドユーザが電子メールを受け取る前に、ウイルスを検出するため電子メールの内容をチェックする許可を要求

電子メールウイルス増殖モデル

- ユーザの電子メールアドレス帳の中を移るだけである電子メールウイルスを考慮
- 無向グラフとして電子メールネットワークをモデル化

電子メールウイルス増殖モデル

$G =< V, E >, \forall v \in V$: 電子メールネットワーク

v: 電子メールユーザ

$\forall e = (u, v) \in E, \quad u, v \in V$: お互いのアドレスを持つユーザ$u$と$v$

$|V|$: ユーザの総数

$T_i (i = 1,2,\cdots,|V|)$: 電子メールチェック時間

$P_i (i = 1,2,\cdots,|V|)$: 添付ファイルを開く確率

$E[T_i]$: 平均電子メールチェック時間

N_t: 時間tで感染したユーザ数

N_0: 最初に感染したユーザ数

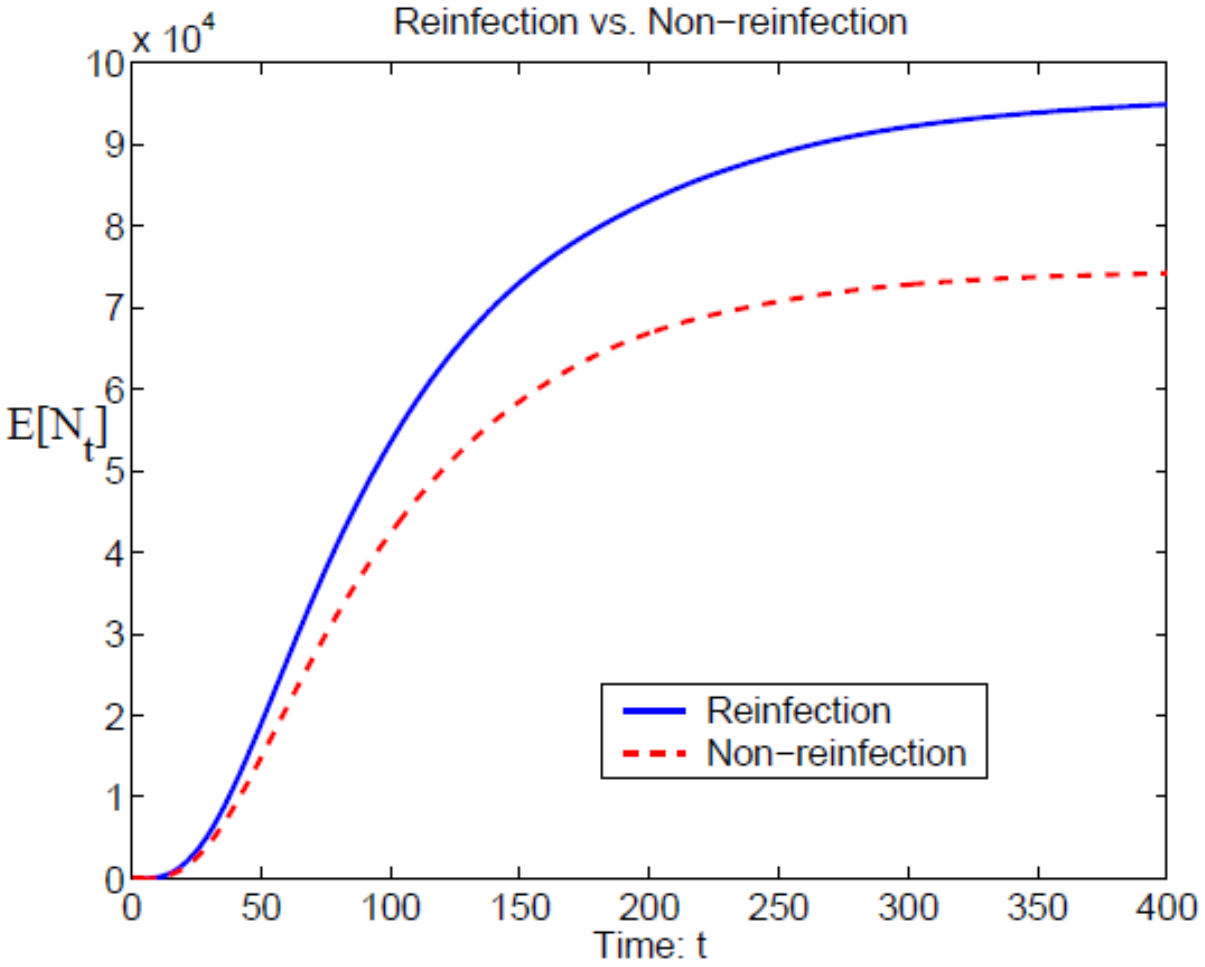
シミュレーション実験

- 100回のシミュレーションを実行
- べき乗則ネットワーク
  - 100000のノード
  - 平均度：8
  - べき乗指数 α ＝1.7
- 他のパラメータ
  - T～N（40,400）,P～N(0.5,0.09), $N_0 = 2$
- 最初に感染するノードは無作為に選ばれる

再感染VS非再感染

- 二つの異なる感染ケースを仮定
  - 再感染
  - 非再感染
- ユーザiがm個のウイルスを受け取ったとき感染する確率は $1-(1-P_i)^m$

再感染VS非再感染


Reinfection vs. Non−reinfection
x 10^4
10
9
8
7
6
5
4
3
2
1
0
E[N_t]
0
50
100
150
200
250
300
350
400
Time: t
Reinfection
Non−reinfection

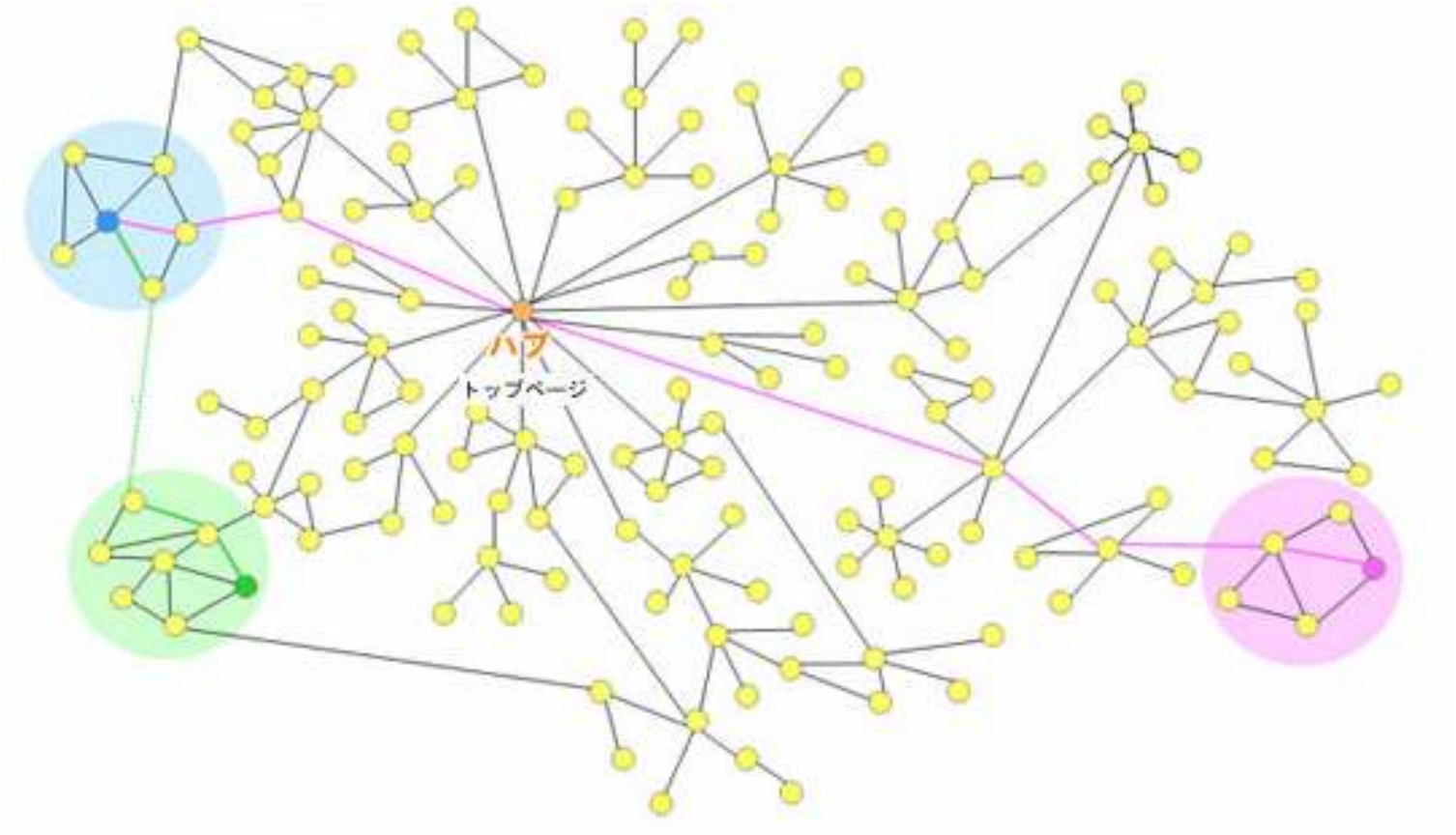
トポロジの影響

- どのようなトポロジがウイルスの増殖に影響を与えるのか調査
- トポロジ
  - べき乗則ネットワーク
  - スモールワールドネットワーク
  - ランダムネットワーク

トポロジの影響

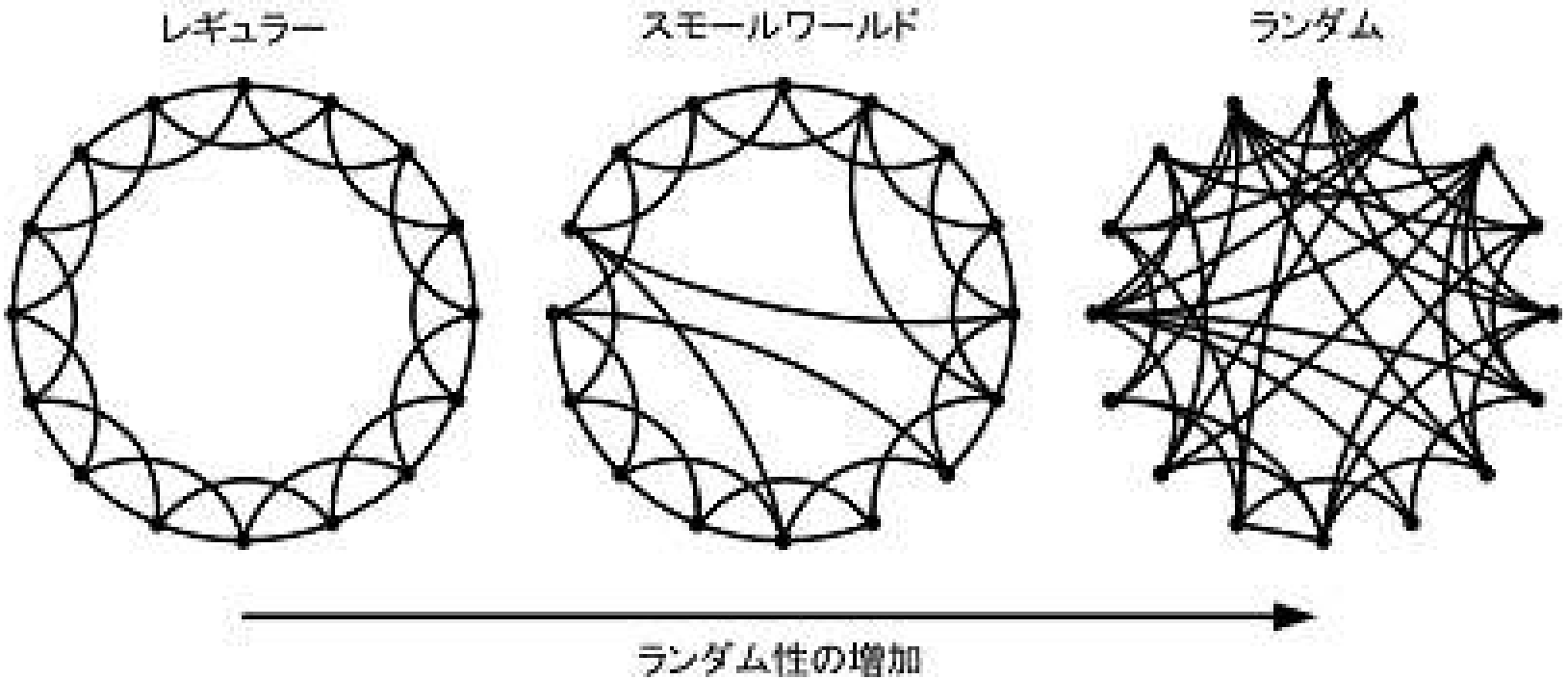
- べき乗則ネットワーク
  - 一部のノードが膨大なリンクを持つ一方で，ほとんどはごくわずかなノードとしか繋がっていないようなネットワーク
- スモールワールドネットワーク
  - あるノード（ネットワークの要素）からほかの任意のノードにたどり着くのに、少数の中継ノードを経由するだけでよいネットワーク
- ランダムネットワーク
  - ノードとノードの間のリンクが志向性もなく、規則性もなく、ランダムに張られているネットワーク

トポロジの影響



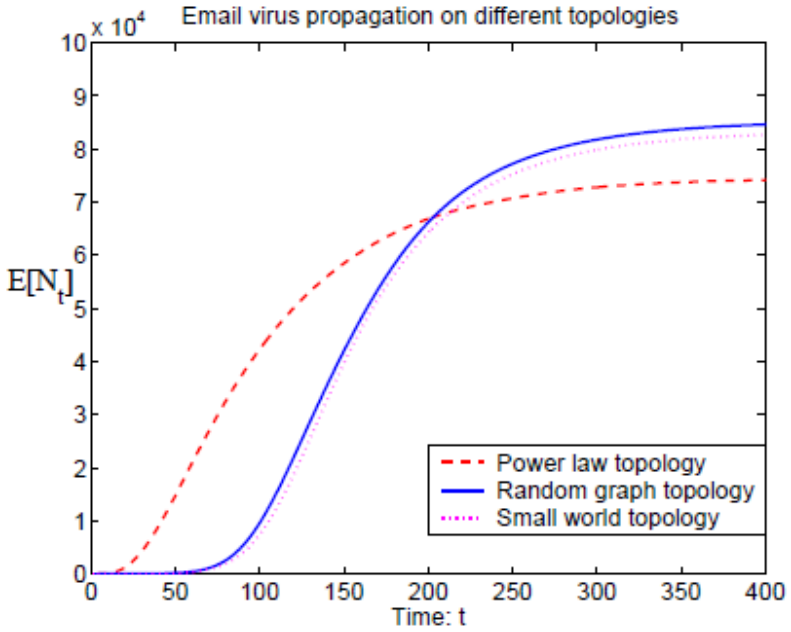
べき乗則ネットワーク　http://marketing.mitsue.co.jp/archives/000122.htmlより

トポロジの影響

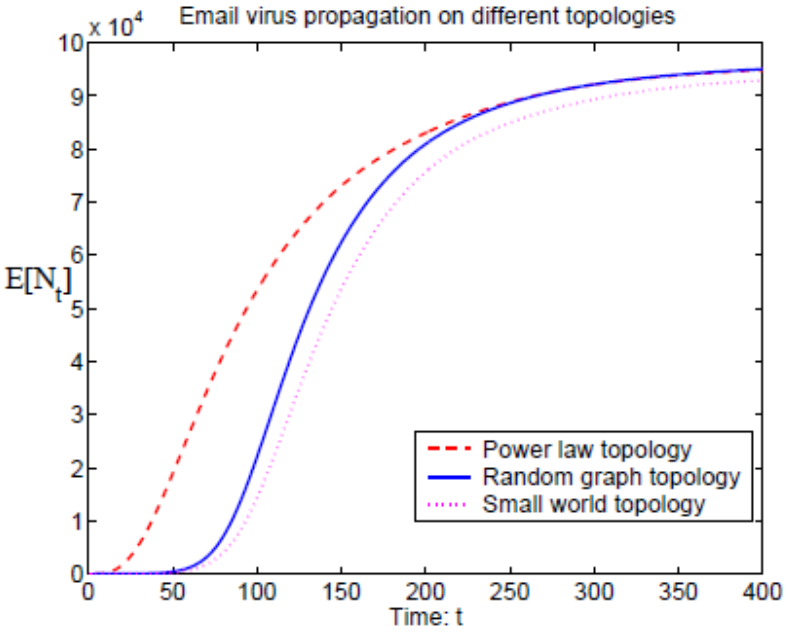


http://www.atmarkit.co.jp/aig/04biz/smallworld.htmlより

トポロジの影響



a. Non-reinfection case



b. Reinfection case

トポロジの影響

- べき乗則ネットワークのほうが感染速度が速い
  - 特徴的な経路長が他の2つに比べて小さいため、ウイルスが早く到達する
    - 特徴的な経路長：2つの頂点間の最小の経路長
  - 「引火する力」がある。異なるノードの度合いがかなり変化

トポロジの影響

- 非再感染のケースでウイルスの増殖が終わった後、べき乗則ネットワークのほうが健全なノードが多い
  - 平均度以下のノードが他のトポロジより多いため健全なノードが残る

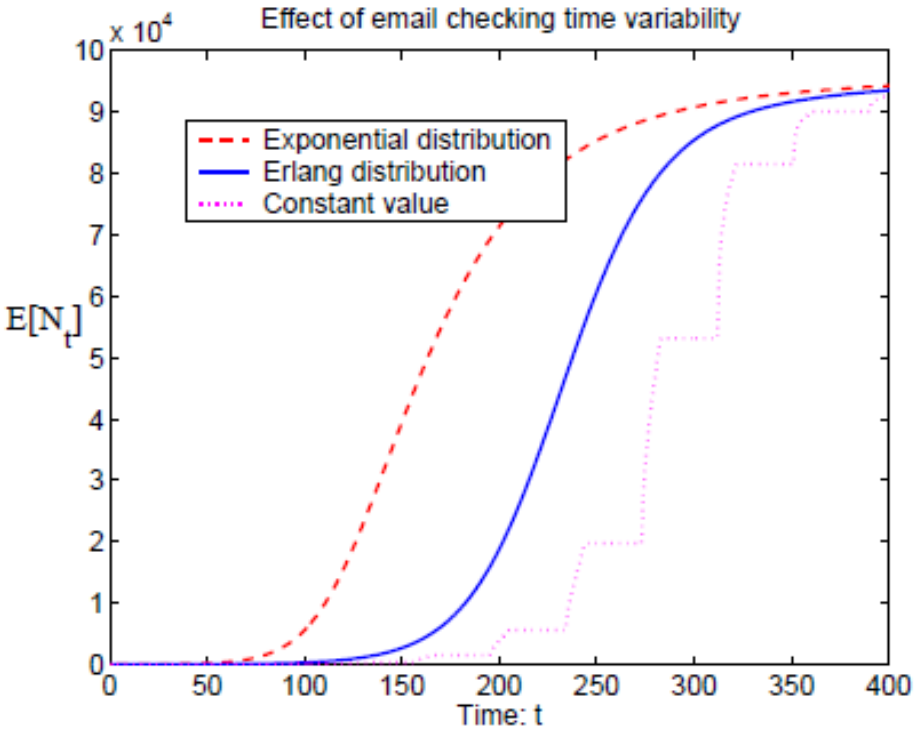
電子メールのチェック時間変化の影響

- 指数分布の平均値:L
- 3次のアーラン分布の平均値:L
- 最初に感染するノードとして無作為に10個のノードを選ぶ

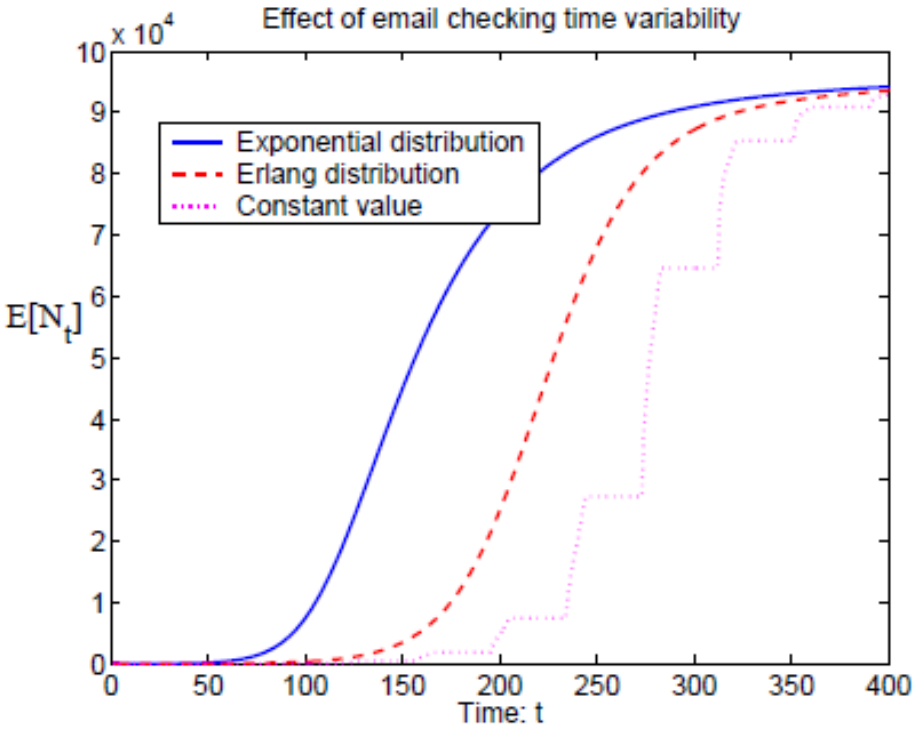
  平均値$L = 40$

  $N_0 = 10$

電子メールのチェック時間変化の影響

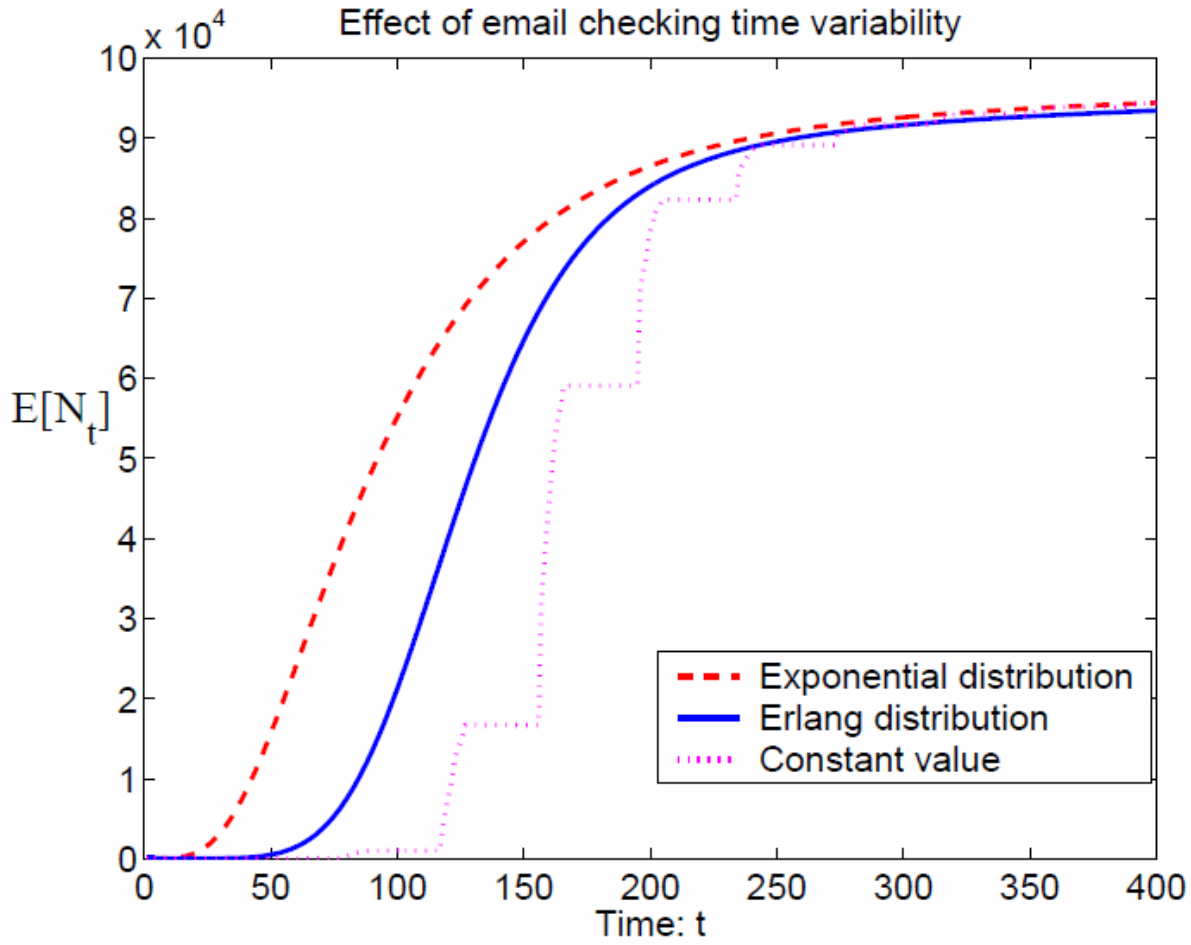


b. Small world topology



c. Random graph topology

電子メールのチェック時間変化の影響



a. Power law topology

電子メールのチェック時間変化の影響

- ウイルスの増殖速度は電子メールチェック時間がより変化するにつれて速くなる

ウイルス防御に関する免疫化とパーコレーション

- 静的な免疫化防御
- 電子メールウイルスが増殖する前に、ネットワーク内の少数のノードはすでに免疫化
- 一部の電子メールユーザがよく教育され、そのユーザが疑わしい電子メールの添付ファイルを決して開けないなら、それは電子メールネットワーク内の免疫化されたノード

電子メールウイルス防御に関する免疫化とパーコレーション

- 選択的な免疫化の影響
- 選択的パーコレーションと電子メールウイルスの予防

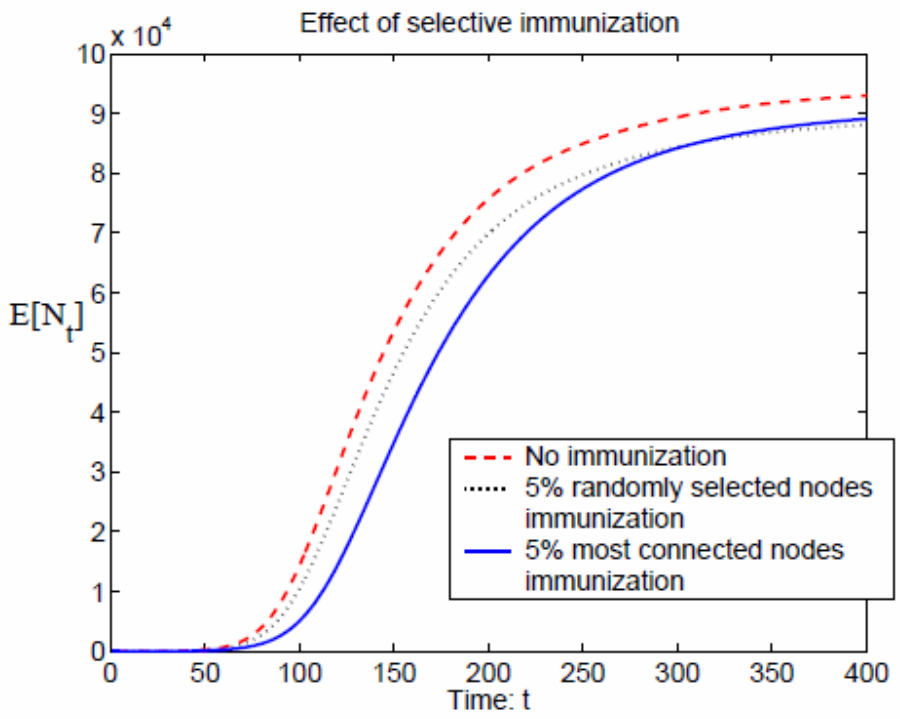
選択的な免疫化の影響

- 全ての電子メールネットワーク内のノードを免疫化するのは不可能
- 現実的なアプローチはノードの小集団を免疫化
- 電子メールウイルスの拡散を失速させるため、この小集団の適当な大きさを選ぶ方法が必要
- 選択的な免疫化は階層的トポロジに関してウイルス増殖をかなり失速させる
- べき乗則ネットワークでは、免疫化するために非常に関係のあるノードを選択することがウイルス増殖に対してとても効果的

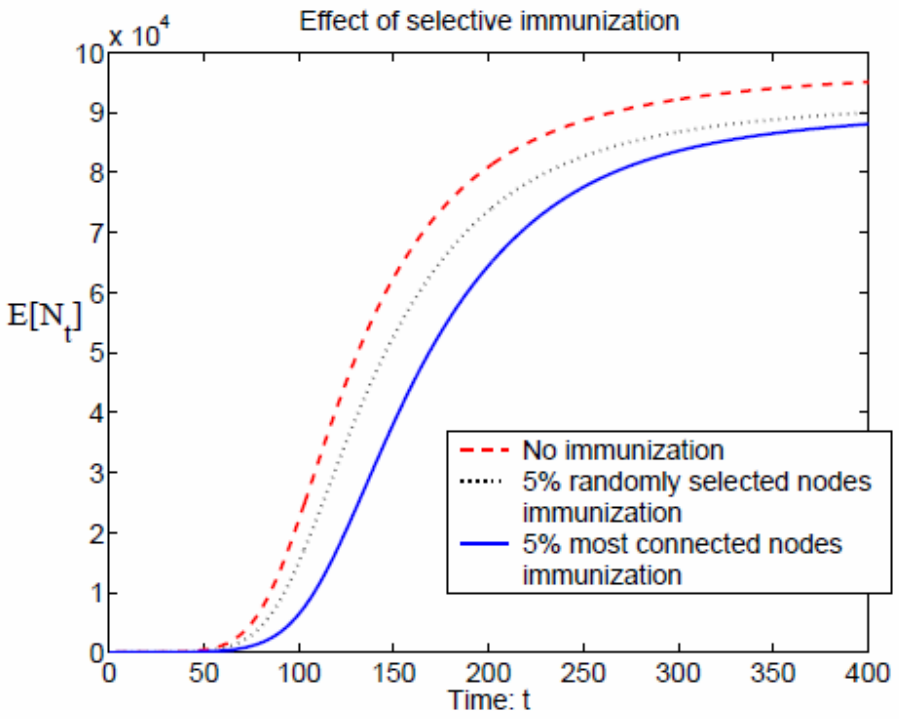
選択的な免疫化の影響

- 二つの異なる免疫化防御の方法の下でウイルス増殖をシミュレーション
  - 無作為に5%のノードを選ぶ方法
  - 5%の最も関係のあるノードを選ぶ方法

選択的な免疫化の影響

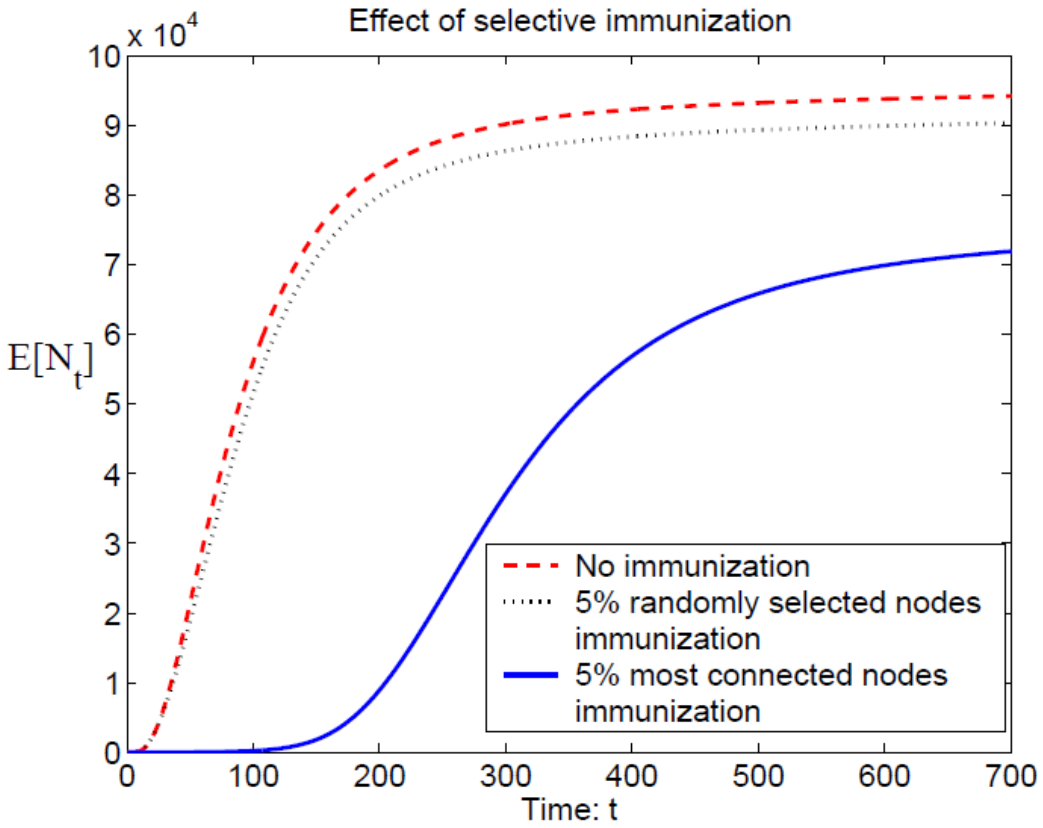


b. Small world topology



c. Random graph topology

選択的な免疫化の影響



a. Power law topology

選択的な免疫化の影響

- 電子メールウイルスは広がるために電子メールネットワークの連結性に依存するので、最も関係のあるノードを免疫化することは急速にネットワークの直径を増やす影響がある

選択的なパーコレーションと電子メールウイルスの予防

- 電子メールウイルスの観点から、部分的に免疫化された電子メールネットワークの連結性はパーコレーション問題
- 「パーコレーション」は、一様にネットワークから若干のノードを取り除くことを意味する
- ネットワークからトップのp%の大部分の関係のあるノードを取り除いた後、まだ関係のあるノードがどれくらい残っているかという割合C(p)
- ネットワークからトップのp%の大部分の関係のあるノードを取り除いた後、残っている関連の割合L(p)

選択的なパーコレーションと電子メールウイルスの予防

$G =< V, E >$: 電子メールグラフ

$|V|$: ノードの数

$|E|$: エッジの数

p : 選択的なパーコレーションの値

$C(p)$: 接続の割合

$L(p)$: 残りの関係の割合

選択的なパーコレーションと電子メールウイルスの予防

$$\begin{cases} C(p) & = c_p/(|V| - |V|p) \\ L(p) & = (|E| - e_p)/|E| \end{cases} \quad 0 < p < 1$$

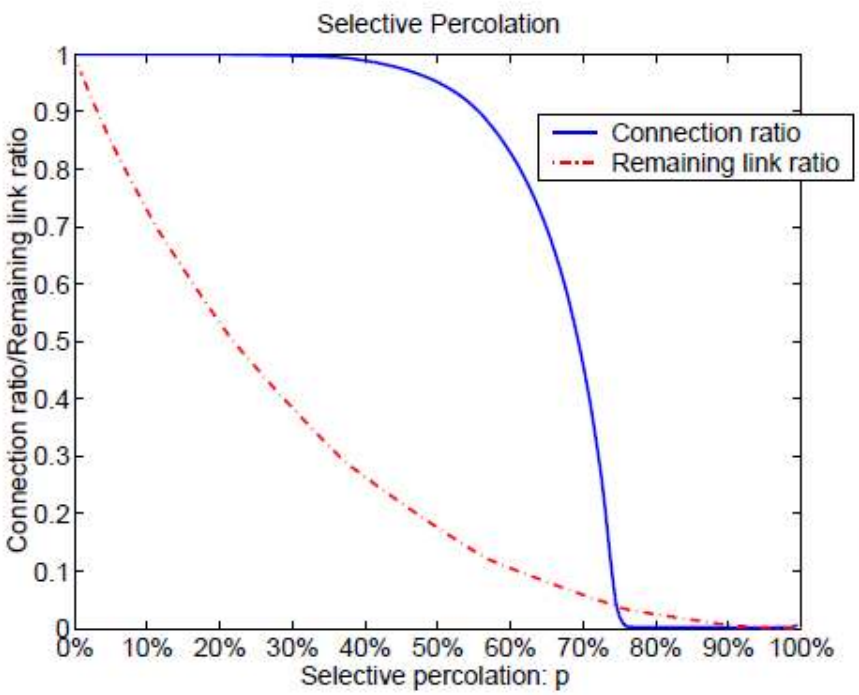
e_p: 取り除かれたエッジの数

c_p: ネットワークに残っている最も大きな集団の大きさ

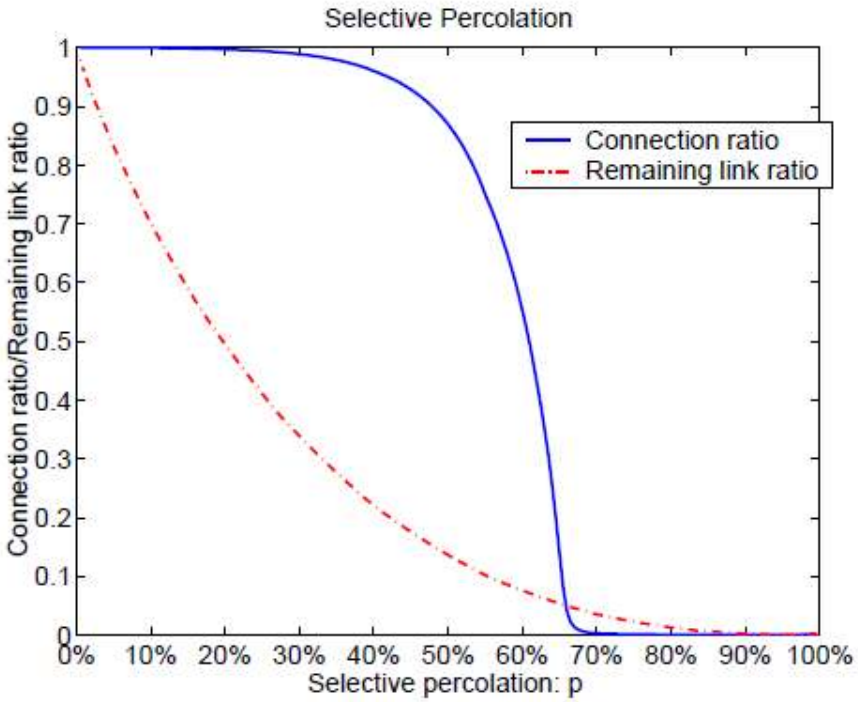
選択的なパーコレーションと電子メールウイルスの予防

- 3つのネットワークをそれぞれ100個生成
- それぞれのネットワーク
  - 平均度：8
  - ノード数：100000
- あらゆる選択的なパーコレーションの値pのために、100のネットワークの各々から導かれるそれらの100個を平均することにより、C(p)とL(p)を計算
- 従って、C(p)とL(p)は対応するトポロジ（1つの単独ネットワークでない）のもの

選択的なパーコレーションと電子メールウイルスの予防

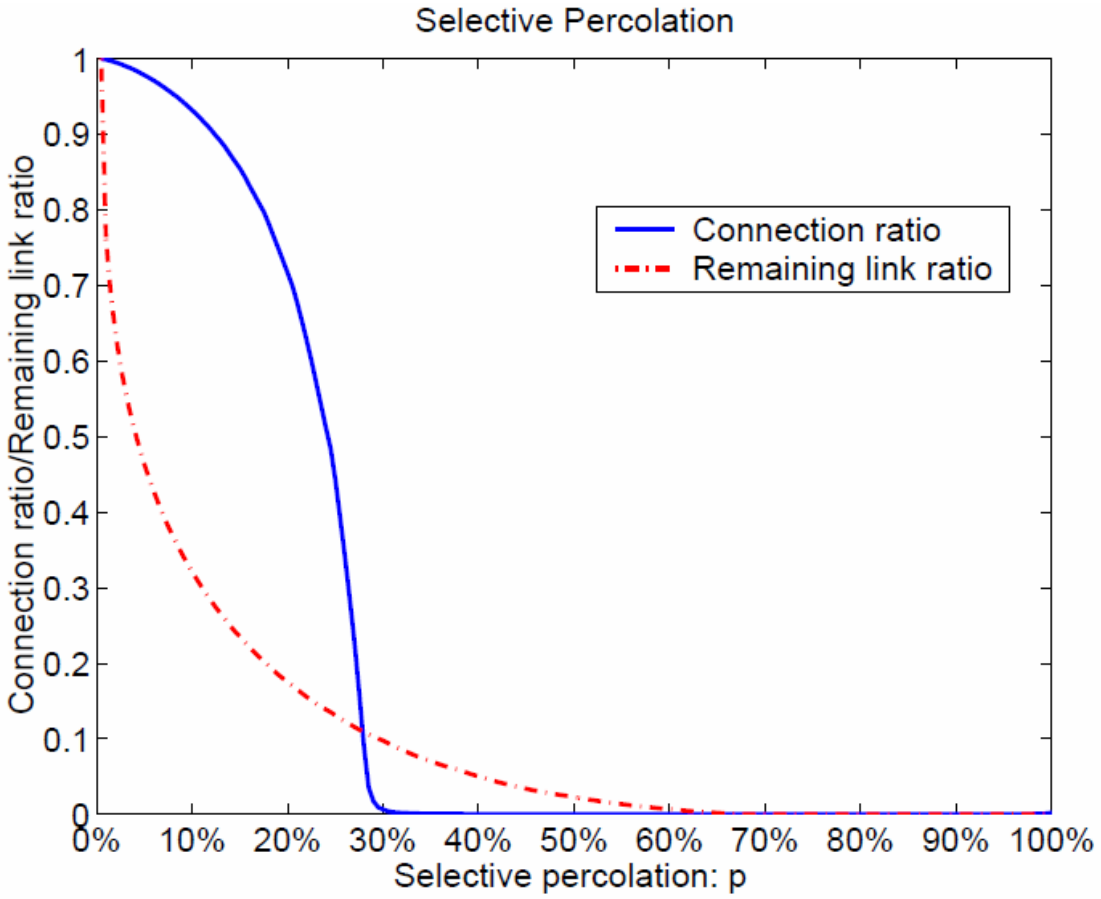


b. Small world topology



b. Random graph topology

選択的なパーコレーションと電子メールウイルスの予防



a. Power law topology

選択的なパーコレーションと電子メールウイルスの予防

- べき乗則トポロジは、選択的なパーコレーション閾値（閾値はここでは、およそ0.29）を持つ
- 選択的に免疫化されたわずかのユーザがこの閾値を上回ると、電子メールネットワークは切り離された断片になるまで壊され、ウイルスの暴動は起こらない
- べき乗則ネットワークで大部分の関係のあるノードのトップの5%を免疫化すると、ネットワークの97.5%の残りのノードがまだ繋がっているが、エッジの55.5%は取り除かれている

選択的なパーコレーションと電子メールウイルスの予防

- 電子メールウイルスは、残りのネットワークでノードに到達し、感染させるためにより少なく長い経路をもつ

結論

- 3つのトポロジについて比較し、べき乗則ネットワークの影響は混合していた
  - ウイルスが急速に広がること
  - 一方で、選択的な免疫化が効果的であること
- 次の段階は電子メールウイルスの拡散を数学的に分析すること
- 現実世界では動的な免疫化も考える必要がある
- 電子メールアドレスの関係は双方向であると仮定したが、ある電子メールユーザにとっては本当ではない可能性がある
- 電子メールウイルスのより良い実態を得るためには有向グラフを考えることが必要

ご清聴ありがとうございました。