

コンピューターウイルスとネットワークの形状

池田 ひでみ

はじめに

- 1 ネットワークとは何か
- 2 ネットワークの形状の性質
- 3 「ウイルス拡散防止型」のネットワーク

はじめに

コンピューターウイルスという言葉を知らない人はいないであろう。インターネットを通してコンピューターに侵入し、その機能を破壊するプログラムで、インターネットに接続する全てのコンピューターはいつこの被害に遭っても不思議ではない。市場には「ウイルス対策ソフト」が軒を連ねており、コストとともに「過防衛」による多少の弊害等を負った上、これをインストールすることを多くのユーザーが選択せざるを得ず、またシステム管理者と呼ばれる者にとってはシステム全体に適切なウイルス対策を施すことは1つの不可避の業務となっているのが現状である。

さて、コンピューターウイルスはインターネット回線を伝って伝染するわけであるが、インターネット回線全体は莫大な数の端末を結ぶ、巨大で複雑なつながりを成す。数学のグラフ理論の分野の言葉を使えば、いわゆる「ネットワーク」を形成していると考えることができる。コンピューターウイルスを防御するためには、(そしてもしかしたらコンピューターウイルスを開発する立場をとるときにも!)、インターネット回線全体やサーバー管理下の回線全体をネットワークであるという側面から考察し、対応することは有益であると考えられる。

ネットワークはそれぞれ多様な形状をもち、その形状の性質は様々である。インターネットのネットワークは、ウイルス感染の速度や加速度を決定する要因としてどのような性質を持っているのであろうか。あるいは、ウイルス感染防御が図られる際に、ネットワークについての知見は何か有用な知恵を提供しうるのであろうか。

本稿は、コンピューターウイルスの伝染をイメージしながらネットワークの形状について議論することを目的とする。第1節でネットワークとはどのような概念であるかを説明する。第2節ではネットワークの形状についていくつかの見方を紹介し、それを実際のインターネット回線に照らしてウイルス感染にまつわる1つの診断を提供する。第3節では理論的に得られるあるネットワークがウイルス感染を広げにくい性格を逐一揃えていることを証明する。

付言すれば、本稿は研究者や教員に限らず、好奇心旺盛な本学学生諸子にも読んでもらいたいと願いながら執筆した。読者にとって極力わかりやすく、面白く、数学が現実寄与するダイナミズムや、一定の興味に沿って方向付けていく研究活動の醍醐味を味わうことができるものであるよう心がけたつもりである。この小さな論稿の読破に挑戦する学生が少数でも現れるなら、幸いに思う。

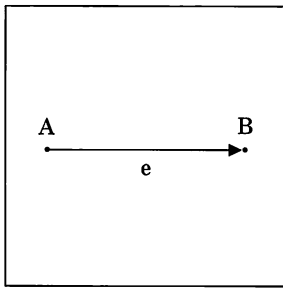
1 ネットワークとは何か

本節では、グラフ理論におけるグラフ並びにネットワークの定義を確認し、それが現実のコンピューターネットワークのモデルとなりえることを理解する。

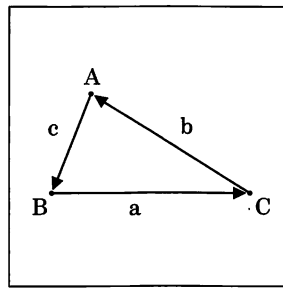
$V = \{v_1, \dots, v_p\}$ を空でない有限集合、 $E = \{e_1, \dots, e_n\}$ を有限集合とし、 ϕ^+ 、 ϕ^- を E から V への写像とする。これらを複合した $G = (V, E, \phi^+, \phi^-)$ をグラフという。 V を G の頂点集合、 E を G の辺集合、 V, E の要素をそれぞれ G の頂点、辺という。 ϕ^+ 、 ϕ^- を G の接続写像といい、 E の元 e に対し、 $\phi^+(e)$ 、 $\phi^-(e)$ をそれぞれ e の始点、終点という。

V, E, G, ϕ 等といったアルファベットは、何を使わなければならないというような規則はないが、 V は Vertex (頂点) の頭文字、 E は edge (辺) の頭文字、そして G は graph (「グラフ」と訳す) の頭文字であるから、 V, E, G はこのアルファベットを使うのが通例である。 ϕ は英語の p に当たるギリシャ文字で「ファイ」と読み、写像を表す文字としてよく使われるものであり、他にも慣習的に写像を表す文字はたくさんあるが、ここでは ϕ を採択した。

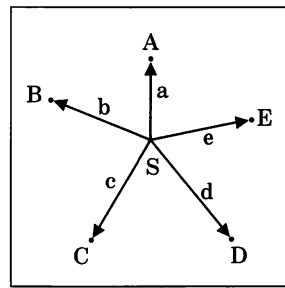
例えば、 $V = \{A, B\}$ 、 $E = \{e\}$ 、 $\phi^+(e) = A$ 、 $\phi^-(e) = B$ のとき、グラフ $G = (V, E, \phi^+, \phi^-)$ は2点 A, B と、 A から B に向かう「矢印」 e から成る図で表すことができる (例1、図1)。また、 $V = \{A, B, C\}$ 、 $E = \{a, b, c\}$ 、 $\phi^+(c) = \phi^-(b) = A$ 、 $\phi^+(a) = \phi^-(c) = B$ 、 $\phi^+(b) = \phi^-(a) = C$ のときは、 $\triangle ABC$ とその「左回りの外周」により、グラフ $G = (V, E, \phi^+, \phi^-)$ は表現される (例2、図2)。あるいは、 $V = \{S, A, B, C, D, E\}$ 、 $E = \{a, b, c, d, e\}$ 、 $\phi^+(a) = \phi^+(b) = \phi^+(c) = \phi^+(d) = \phi^+(e) = S$ 、 $\phi^-(a) = A$ 、 $\phi^-(b) = B$ 、 $\phi^-(c) = C$ 、 $\phi^-(d) = D$ 、 $\phi^-(e) = E$ ならば、 $G = (V, E, \phi^+, \phi^-)$ は1点 S から放射状に伸びる5つの「矢印」と各「矢印」の終着点の図により、イメージが捉えられるであろう (例3、図3)。



(図1)



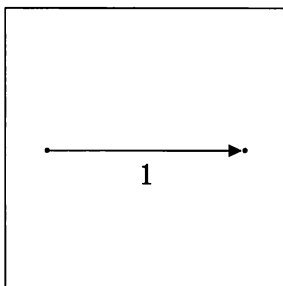
(図2)



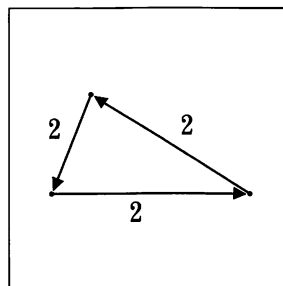
(図3)

次にネットワークであるが、これはグラフにさらにある種の関数を付与したのものとして定義される。しかし、その定義の中身は必ずしも統一されていない。ここでは、最も簡潔な定義を採択しよう。すなわち、グラフ $G=(V, E, \phi^+, \phi^-)$ に、 E 上の非負実数関数 c が付与されているとき、 G をネットワークということにする。

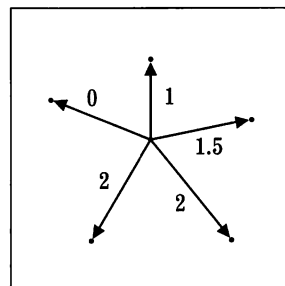
ここでの c は、capacity (容量) の頭文字を使ってみた。上で挙げたグラフの例にこのような関数を付加して、ネットワークの例を作ってみよう。尚、簡便のため、図は V, E の元の名称を省略したものを示す。例1に $c(e)=1$ を付加すると図4ようになる。 c は E の各元に対して個別の実数を与えるが、例えば例2に $c(a)=c(b)=c(c)=2$ と、同じ値が設定されても構わない (図5)。例3で $c(a)=1, c(b)=0, c(c)=c(d)=2, c(e)=1.5$ とすると、図6ようになる。



(図4)



(図5)



(図6)

以上が、グラフ理論におけるネットワークの定義であるが、例えば V の各元はコンピューターを表すとし、 $\phi^+(e)=A, \phi^-(e)=B$ なる E の元 e をサーバー A とそれが管理するクライアント B を結ぶ回線、あるいはプロバイダ A とユーザー B を結ぶ回線とみなし、 $c(e)$ は回線 e の容量と考えてはどうだろうか。ここに導入したネットワークというアイデアは、いわゆるコンピューターネットワークのモデルになりうるということが理解できるであろう。実は、ネットワークは交通網、人口の入出流、伝染病感染経路網、脳科学における化学物質伝達経路網など、実世界の現象や問題について広範囲に亘る応用分野をもつ、重要な数学的概念であることが知られている。

2 ネットワークの形状の性質

前節を受け、本節では、コンピューターネットワークにおいてコンピューターウイルスが拡散するのに好都合な条件・不都合な条件を、ネットワークのモデルに照らして考察する。

本節はネットワークの形状に注目していく。従って、グラフの容量関数の問題は無視し、つまりネットワークと言いつつ、実質的にグラフの形状について論じることとする。

例に即して考えていこう。以下、3つの例を挙げる。

例4 (図7) $V=\{A, B, C, D, E\}$, $E=\{p, q, r, s\}$,

$$\phi^+(p)=A, \phi^-(p)=\phi^+(q)=B, \phi^-(q)=\phi^+(r)=C, \phi^-(r)=\phi^+(s)=D, \phi^-(s)=E.$$

これは、5つの頂点を4つの辺で1列に並べた例である。列中央部に位置するBとDを結ぶ新たな1辺を加え、適当に始点と終点、アルファベットを入れ替えることにより、次の例5を得る。

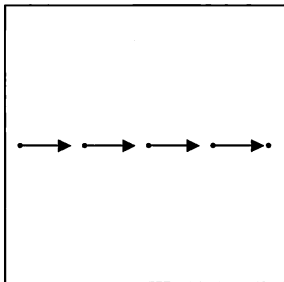
例5 (図8) $V=\{A, B, C, D, E\}$, $E=\{p, q, r, s, t\}$,

$$\phi^+(p)=A, \phi^-(p)=\phi^+(q)=B, \phi^-(q)=\phi^+(r)=C, \phi^-(r)=D, \phi^+(s)=\phi^-(t)=E, \phi^-(s)=B, \phi^+(t)=C.$$

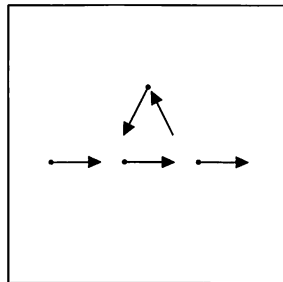
もう1つ、頂点の個数と辺の個数は例4, 5と同じで、辺の配置を変えた例を設ける。

例6 (図9) $V=\{A, B, C, D, E\}$, $E=\{p, q, r, s\}$,

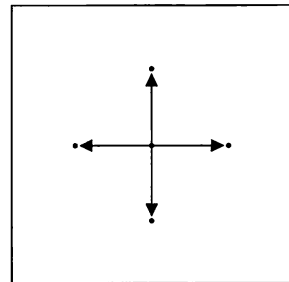
$$\phi^+(p)=\phi^+(q)=\phi^+(r)=\phi^+(s)=A, \phi^-(p)=B, \phi^-(q)=C, \phi^-(r)=D, \phi^-(s)=E.$$



(図7)



(図8)



(図9)

一般に、グラフの2頂点 A, B に対し、頂点の例 $A=P_0, P_1, \dots, P_r=B$ と辺の例 e_1, e_2, \dots, e_r が存在して、 $i=1, 2, \dots, r$ について

$$\phi^+(e_i)=P_{i-1} \quad \text{かつ} \quad \phi^-(e_i)=P_i$$

または

$$\phi^-(e_i)=P_{i-1} \quad \text{かつ} \quad \phi^+(e_i)=P_i$$

が成り立つとき、A と B は連結であるといい、 e_1, e_2, \dots, e_r を A と B の連結鎖、 r をこの連結鎖の長さという。A と B の連結鎖の中で長さが最小のものを A と B の極小連結鎖といい、極小連結鎖の長さを A と B の距離といって $d(A, B)$ で表す。グラフ G の任意の2頂点 A, B が連結であるとき、G を連結であるという。さらに、グラフ G が連結であるとき、頂点間の距離の平均値を、G の平均距離といい、 $d(G)$ で表す。

本節で挙げた3つのグラフの例は、いずれも頂点を5個もつ連結なグラフであるから、2頂点の組の選び方は ${}_5C_2=10$ 通り存在する。ここに生じた10組の頂点間の距離の平均値が、各グラフの平均距離となる。計算すると、例4, 5, 6の順に、 $d(G)=2.0, 1.6, 1.6$ を得る。

さて、これらのグラフの例を、コンピューターネットワークに見立て、それぞれ1つの頂点すなわちコンピューターに、ウィルスが感染したとしよう。平均距離が大きいということはどのようなことを意味するであろうか。グラフの辺はコンピューター回線、つまり感染経路に当たるのであったから、A と B の距離が1ならば、A と B とは直接感染する可能性があり、距離が2ならば、A と B の間のもう1つの頂点を介して感染が起こりうる。距離が3ならば、2つの頂点を介して3つの辺をウィルスが通らなければ感染が生じないということになる。つまり、平均距離が大きいということは、ウィルスの蔓延が遅れるための要因になると考えられる。従って、例4のように直列に並んだ形状を多く含むネットワークよりも、例5や例6のような、より平均距離の小さい形状のネットワークの方が、ウィルス防御は難しいということができる。

次に、グラフ G の頂点 A について、 $\phi^+(e)=A$ となる辺 e の個数と $\phi^-(e)=A$ となる辺 e の個数の和を A の次数といい、 $\delta(A)$ で表す。 $\delta(A)$ は、A を始点または終点とする辺の個数であり、始点も終点も A である辺も含む。グラフ G の頂点の次数の平均値 $\delta(G)$ を G の平均次数といい、頂点の次数の分散 $v(G)$ を G の次数分散と呼ぶことにする。

例4では、 $\delta(A)=\delta(E)=1, \delta(B)=\delta(C)=\delta(D)=2$ であり、この5個の数値の平均値と分散を求めると、 $\delta(G)=1.6, v(G)=0.24$ となる。例5では、頂点の次数はより散らばりを見せ、 $\delta(A)=\delta(D)=1, \delta(E)=2, \delta(B)=\delta(C)=3$ となっており、 $\delta(G)=2.0, v(G)=0.80$ である。例6は、 $\delta(A)=4, \delta(B)=\delta(C)=\delta(D)=\delta(E)=1$ と、ほとんどの頂点は小さな次数をもつものの、突出して大きな次数をもつ頂点が存在する形状を成すグラフであり、 $\delta(G)=1.6, v(G)=1.44$ となっている。例4と例6が4個の辺をもっているのにたいし、例5は5個の辺をもつものだから、平均次数 $\delta(G)$ が大きくなるのは当然であると感ぜられるが、ここでは特に $v(G)$ に着目したい。例4, 5, 6の順に $v(G)=0.24, 0.80, 1.44$ と、大きな違いが見られる。このことは、例4に比べて例5, 6は頂点の次数に

散らばりがあることを示し、特に、例 4 に 1 辺を付け加えて全体的に次数の大きな頂点をもつようになった例 5 よりも、辺の個数は同じでも抜本的に形状を変えて 1 個であっても非常に大きな次数をもつ頂点が存在する例 6 の方が、さらに大きな次数分散をもつことを例示している。

グラフの次数分散という概念を、実際のコンピューターネットワークに照らして解釈してみよう。コンピューター A の次数が大きいということは、A が多くのコンピューターと回線で直接つながっていることを表現している。従って、次数分散が大きいということは、かなり多くの他のコンピューターに接続しているコンピューターが少数存在するという状況に相当する。インターネットのネットワークならば、突出して多いアクセス数を誇るサイトが存在するということである。もしもこのような形状のネットワークにおいて、次数の大きな頂点が狙い打ちされてウイルスに感染したらどうなるであろうか。次数分散の大きなグラフをモデルとしてもつ形状のコンピューターネットワークにおいては、ウイルス拡散のリスクが高いと言わざるをえない。

グラフの平均距離と次数分散という概念を導入し、コンピューターネットワークのウイルス感染への応用を考えてきた。実際のコンピューターネットワークをグラフでモデル化するならば、平均距離も次数分散も相当大きくなることは、その機能と実情を考慮するだけで、容易に想像できよう。つまり、厄介なことであるが、コンピューターネットワークは、本質的にウイルス感染を防ぎにくい形状を有するということになる。

3 「ウイルス拡散防止型」のネットワーク

本節では、いわゆるブール束から想起されるネットワークの具体例を構築し、前節で提供したネットワークの諸性質について考察する。

自然数 n に対し、グラフ $G(n) = (V(n), E(n), \phi_n^+, \phi_n^-)$ を次により定める。

$$\begin{aligned} V(n) &= \{v(\mu_1, \dots, \mu_n) \mid \mu_i = 0, 1\}, \\ E(n) &= \{e(\mu_1, \dots, \mu_n) \mid \mu_i = 0, 1 \text{ for } i \neq p; \mu_p = 1/2, p = 1, \dots, n\}, \\ \phi_n^+(e(\mu_1, \dots, \mu_n)) &= v(\lambda_1, \dots, \lambda_n), \\ \phi_n^-(e(\mu_1, \dots, \mu_n)) &= v(\nu_1, \dots, \nu_n), \\ \mu_i \neq 1/2 \text{ ならば } \lambda_i = \nu_i = \mu_i; \mu_i = 1/2 \text{ ならば } \lambda_i = 0, \nu_i = 1. \end{aligned}$$

ここに構築したグラフ $G(n)$ は、いわゆるブール束から着想を得たグラフである。この $G(n)$ について、まず、前節で議論した平均距離と次数分散を求めよう。

$$\begin{aligned} \text{定理 } d(G(n)) &= \frac{\sum_{r=1}^n r_n C_r}{2^n - 1}, \\ \delta(G(n)) &= n, \\ v(G(n)) &= 0. \end{aligned}$$

証明 一般に $V(n)$ の元 $A=v(\lambda_1, \dots, \lambda_n)$ に対し、 $\lambda_i=1$ である i の個数を $\kappa(A)$ とおく。 $A=v(\lambda_1, \dots, \lambda_n)$ を固定し、 $V(n)$ における A 以外の元 $X=v(\nu_1, \dots, \nu_n)$ を任意にとる。このとき、 $d(A, X)=|\kappa(A)-\kappa(X)|$ である。これは、 $r=|\kappa(A)-\kappa(X)|$ とおくと、 r は $\lambda_i \neq \nu_i$ を満たす添字 i の個数であることに注意すれば、 r についての帰納法を用いて次のようにして確かめられる。実際、 $r=1$ のとき、 $\lambda_p \neq \nu_p$ なる唯一の添字 p を取り、

$$i \neq p \text{ ならば } \mu_i = \lambda_i ; \mu_p = 1/2$$

により定まる $E(n)$ の元 $e=e(\mu_1, \dots, \mu_n)$ により、「 $\phi_n^+(e)=A$ かつ $\phi_n^-(e)=B$ 」または「 $\phi_n^+(e)=B$ かつ $\phi_n^-(e)=A$ 」となるから、 $d(A, X)=1$ である。 $r>1$ とし、 $\lambda_p \neq \nu_p$ なる添字 p を取り、

$$i \neq p \text{ ならば } \lambda'_i = \lambda_i ; \lambda'_p = \nu_p$$

により $V(n)$ の元 $A'=v(\lambda'_1, \dots, \lambda'_n)$ を考える。 $|\kappa(A')-\kappa(X)|=r-1$ であるから、帰納法の仮定により、 $d(A', X)=|\kappa(A')-\kappa(X)|=r-1$ である。よって、

$$d(A, X) \leq d(A, A') + d(A', X) = 1 + (r-1) = r$$

である。一方、 ϕ_n^+ 、 ϕ_n^- の定義によれば、 $E(n)$ の任意の元 e について、

$$|\kappa(\phi_n^+(e)) - \kappa(\phi_n^-(e))| = 1$$

であることに注意せよ。しからば、 A と X を結ぶ任意の連結鎖の長さは r より小さくないことがわかり、これより $d(A, X) \geq r$ であることが知れる。ここに、 $d(A, X)=r$ を得る。

$A=v(\lambda_1, \dots, \lambda_n)$ を固定したとき、 $d(A, X)=r$ となる $V(n)$ の元 $X=v(\nu_1, \dots, \nu_n)$ の個数は、 $\lambda_i \neq \nu_i$ を満たす r 個の添字 i の組合せ ${}_n C_r$ に他ならない。従って、

$$\sum_{\substack{X \neq A \\ X \in V(n)}} d(A, X) = \sum_{r=1}^n r {}_n C_r$$

である。この式の右辺は A に寄らないから、 $|V(n)|=2^n$ に注意しながら平均値を求めると、

$$\begin{aligned} d(G(n)) &= \frac{\sum_{\substack{X \neq A \\ X \in V(n)}} d(A, X) \times 2^n \div 2}{2 C_2} \\ &= \frac{\sum_{r=1}^n r {}_n C_r \times 2^n \div 2}{2^n (2^n - 1) \div 2} \\ &= \frac{\sum_{r=1}^n r {}_n C_r}{2^n - 1} \end{aligned}$$

となる。

次に、 $V(n)$ の任意の元 $A=v(\lambda_1, \dots, \lambda_n)$ に対し、 n 個の $E(n)$ の元

$$e_i = e(\mu_{i1}, \dots, \mu_{in}), (i=1, \dots, n)$$

を

$$j \neq i \text{ ならば } \mu_{ij} = \lambda_j ; \mu_{ii} = 1/2$$

により定める。このとき、 $i=1, \dots, n$ について

$$\lambda_i=0 \text{ ならば } \phi_n^+(e_i)=A,$$

$$\lambda_i=1 \text{ ならば } \phi_n^-(e_i)=A$$

が成り立つ。しかも、 e_1, \dots, e_n 以外に $\phi_n^+(e)=A$ または $\phi_n^-(e)=A$ を満たす $E(n)$ の元 e は存在しない。ゆえに、 $\delta(A)=n$ である。任意の A について $\delta(A)=n$ であるから、次数 $\delta(A)$ の平均と分散を取って、 $\delta(G(n))=n$ および $v(G(n))=0$ を得る。(証明終わり)

平均距離について具体的に数値を計算すると、 $d(G(1))=1$, $d(G(2))=4/3 \approx 1.33$, $d(G(3))=12/7 \approx 1.71$, $d(G(4))=32/15 \approx 2.13$, $d(G(5))=80/31 \approx 2.58$ となっている。この値は「小さい」のであろうか、それとも「大きい」のであろうか。

そこでまず、前節で取り上げた例について考察を重ねよう。 $G(n)$ と同じ個数の頂点を持ち、例 4 のように 1 列に並んだグラフの平均距離は

$$\frac{2^n+1}{3}$$

であることは容易に確認でき(読者各自で確かめられよ)、この値は $n=1, 2, 3, 4, 5$ のときそれぞれ $1, 5/3 \approx 1.67, 3, 17/3 \approx 5.67, 11$ と、 n の増加に伴って「指数関数的に」増大し、上限はない。また、 $G(n)$ と同じ個数の頂点をもつグラフで 1 頂点を中心にして残りの頂点へ辺が伸びている、例 6 に代表される型のグラフと考えれば、その平均距離は

$$\frac{2^n-1}{2^{n-1}} = 2 - \frac{1}{2^{n-1}}$$

であり(この確認も読者に委ねる)、 $n=1, 2, 3, 4, 5$ のときそれぞれ平均距離 $1, 3/2=1.5, 7/4=1.75, 15/8 \approx 1.88, 31/16 \approx 1.94$ となり、 n が大きくなるに従って上限 2 に近づく。この 2 例は辺の個数の少ないグラフとして、平均距離に関する両極端に相当する例であるが、辺の個数を増やして、任意の頂点 A, B に対して辺 e が存在して $\phi_n^+(e)=A, \phi_n^-(e)=B$ となるようにすれば、平均距離は最小値 1 をとる。

こうした観点から、 $G(n)$ の平均距離について考える。 $t = \frac{n}{2}$ とおく。このとき、

$$t = \frac{\sum_{r=1}^n r_n C_r}{2^n}$$

である。実際、

$$\begin{aligned} & \sum_{r=1}^n r_n C_r - t 2^n \\ &= \sum_{r=0}^n r_n C_r - t \sum_{r=0}^n n C_r \\ &= \sum_{r=0}^n (r-t)_n C_r \\ &= \sum_{0 \leq r < t} (r-t)_n C_r + \sum_{t < r \leq n} (r-t)_n C_r \\ &= \sum_{0 \leq r < t} (r-t)_n C_r + \sum_{0 \leq r < t} ((n-r)-t)_n C_r \end{aligned}$$

$$= \sum_{0 \leq r < t} (n-2t)_n C_r$$

$$= 0$$

となっている。ゆえに、

$$\frac{\sum_{r=1}^n r_n C_r}{2^n - 1} > \frac{\sum_{r=1}^n r_n C_r}{2^n} = t$$

であり、定理の結果より

$$d(G(n)) > \frac{n}{2}$$

を得る。この式は、 $d(G(n))$ が有界でなく、 n を十分大きくとればいくらでも大きくなることを示唆している。

以上より、 $G(n)$ は、「大きな」平均距離 $d(G(n))$ と「最低の」次数分散 $v(G(n))=0$ をもつグラフであると考えられることができる。前節の議論に従えば、 $G(n)$ は、まさにウィルス拡散防止に寄与する形状を有しているということになる。

現実のコンピューターネットワークに $G(n)$ の理論をそのまま適用できる場面は、あるいは少ないことが予想される。しかし、平均距離や次数分散といった概念を踏まえつつ、モデルとして念頭におくことは可能ではないであろうか。

最後に、 $G(n)$ の着想源であるブール束に対応する代数的概念として、ある種のスクエア・フリーな完全交差環が考えられるが、必ずしもスクエア・フリーでない完全交差環、さらにはゴーレンシュタイン環をヒントにして、グラフを構築することができる。しかし、それらのグラフは $G(n)$ に比べて格段に複雑な構造をもつことが予想される。今後の課題として残そうと思う。

参考文献

- [1] L. R. Ford and D. R. Fulkerson [1962], Flows in Networks, Princeton University Press.
- [2] F. Harary [1969], Graph Theory, Addison-Wesley.
- [3] R. Pastor-Satorras and A. Vespignani [2004], Evolution and Structure of the Internet, Cambridge University Press.