# Virus Propagation in a Graph

Tomás Feder
268 Waverley St., Palo Alto, CA 94301, USA
tomas@theory.stanford.edu,
Amin Saberi
Management Science and Engineering Department, Stanford University,
Stanford, CA 94305, USA
saberi@stanford.edu

June 22, 2016

## 1 Introduction

Let $G$ be an $n$-vertex graph with a starting vertex $v$ that has a virus. We consider a process that propagates the virus by growing a set $V$, initially $V = \{v\}$, until $V = V(G)$. At each step, each edge incident to both $V$ and $V(G) - V$ fires according to a Poisson process of rate 1, thus with expected time 1 of firing for each edge. When the first edge $xy$ fires with $x \in V$ and $y \notin V$, the set $V$ is replaced by $V \cup \{y\}$.

Let $V_i$ by the set of size $i$ reached at the $i$th step, and let $E_i$ be the set of edges joining $V_i$ to $V - V_i$.

**Theorem 1.1** *The virus reaches all vertices in expected time* $E(\sum_{1 \leq i < n} \frac{1}{|E_i|})$.

**Proof.** When set $V_i$ is reached, the $|E_i|$ Poisson random variables for the edges in $E_i$ add up to a Poisson with rate $|E_i|$ and thus expected firing time $\frac{1}{|E_i|}$. Adding over all $n-1$ sizes $i$ gives the result. □

**Theorem 1.2** *Suppose each connected set $V$ of vertices in $G$ with $|V| \leq \frac{|V(G)|}{2}$ has number of edges joining $V$ to $V(G) - V$ between $e|V|$ and $e'|V|$. Then the virus reaches all vertices in expected time $\Omega(\frac{\log n}{e'})$ and $O(\frac{\log n}{e})$. In particular, if $G$ is a clique on $n$ vertices $K_n$, the virus reaches all vertices in expected time $\Theta(\frac{\log n}{n})$.*

**Proof.** By Theorem 1.1, we obtain a lower bound $\sum_{1 \leq i < n/2} \frac{1}{e'i} + \sum_{n/2 \leq i < n} \frac{1}{e'(n-i)} = \Omega(\frac{\log n}{e'})$ and an upper bound $\sum_{1 \leq i < n/2} \frac{1}{ei} + \sum_{n/2 \leq i < n} \frac{1}{e(n-i)} = O(\frac{\log n}{e})$. □

## 2 Detecting Viruses in a Layered Graph

A *layered graph* $G$ has the vertices divided into layers $S_1, S_2, \ldots, S_D$ so that all edges are contained in a layer $S_i$ or join consecutive layers $S_i, S_{i+1}$. We shall assume that if $d$ is the maximum degree of $G$, then the number of edges joining a vertex in $S_i$ to each of $S_{i-1}, S_{i+1}$ is $\Omega(d)$. Let $s$ be the maximum size of an $S_i$.

**Theorem 2.1** *Let $G$ be a layered graph, and suppose initially the virus occupies all of $S_1$. Then the expected time to fully occupy the first $r$ layers is $O(\frac{r \log s}{d})$, and with high probability in time $\frac{t}{d}$ the virus will not reach more than $O(\log s + t \log t)$ layers. Thus the virus will be detected after $r$ layers before it reaches more than $O(r \log s \log(r \log s))$ layers.*

**Proof.** As in Theorem 1.2, if layer $S_i$ is fully occupied, the expected time to fully occupy layer $S_{i+1}$ is at most $O(\frac{\log |S_{i+1}|}{d})$, proving the $O(\frac{r \log s}{d})$ bound to fully advance $r$ layers.

To prove the upper bound on the number of layers reached, select a small constant $\epsilon > 0$, and consider a $Q_v$ containing all $S_v$ with $w \leq v$, and to contain at most $|S_v|\epsilon^i$ elements in the union of the $S_w$ for $w \geq v + i$ for all $i \geq 1$ such that $|S_v|\epsilon^i \geq c \log t$ for some large constant $c$. Suppose the virus has reached precisely $Q_v$. Fire $\delta|S_v|$ edges to spread the virus for some $\delta < \epsilon$. Consider the process for as long as there are at most $2|S_v|\epsilon^i$ elements in the union of the $S_w$ for $w \geq v + i$ for all $i \geq 1$ such that $|S_v|\epsilon i \geq c \log t$. The degree of $S_v$ into $S_{v+1}$ is at least $d/c'$ for some constant $c'$, giving at least $(d/c')|S_v|$ edges joining $S_v$ to $S_{v+1}$. The at most $(\epsilon + \delta)|S_v|$ vertices that will be reached in $S_{v+1}$ amount to at most $d(\epsilon + \delta)|S_v|$ edges joining them to $S_v$. This leaves $((d/c') - d(\epsilon + \delta))|S_v| \geq (d/c'')|S_v|$ edges not fired from $S_v$ to vertices not reached in $S_{v+1}$. If there are at most $2|S_v|\epsilon^i$ elements in the union of the $S_w$ for $w \geq v + i$ for some $i \geq 1$ such that $|S_v|\epsilon i \geq c \log t$, these can fire at most $2d\epsilon^i|S_v|$ into the corresponding union for $i + 1$. The expected number that will actually fire is thus at most $\delta(2d\epsilon^i/(d/c''))|S_v|$, and thus with high probability at most twice this many will fire, that is, the probability that more than $4c''\delta\epsilon^i|S_v|$ will fire is at most $1/t^{c'''}$ for a large constant $c'''$ depending on the large constant $c$, by Chernoff's bound. Setting $\delta < \epsilon/(4c'')$ guarantees that at most $\epsilon^{i+1}|S_v|$ will be fired with high probability, satisfying the bound $2\epsilon^{i+1}|S_v|$ for the union starting at $S_{v+i+1}$. In particular, the last $i$ considered will reach at most $\log s + c \log t$. Shifting $v$ to $v' = v + c \log t$ and adding the virus to some extra vertices satisfies the initial assumptions that held before the $\delta|S_v|$ firings again after these firings for $v'$. The shift from $v$ to $v'$ can be done because $(d/c')|S_w| \leq d|S_{w+1}|$ by counting edges joining $S_w$ to $S_{w+1}$, so $|S_{w+1}| \geq |S_w|/c''$ while $\epsilon$ is small. The number of edges that can fire at any point in this process is $O(d|S_v|)$, so $\delta|S_v|$ firings consume time $\Omega(\delta/d) = \Omega(1/d)$ in expectation since $\delta$ is a constant. Thus we initially advance $\log s + c \log t$, and the next $t$ times we advance at most $c \log t$ each time, each time with high probability $1 - 1/t^{c'''}$, proving the bound $O(\log s + t \log t)$ on the number of layers advanced in $O(t/d)$ time. □

# 3 Enforcing Ratios and the Hypercube

The following proof is adapted from [3, 2]. Given a graph $G$ with $V = V(G)$, we shall say that $G$ *enforces ratios* if the following holds inductively. If $|V| = 1$ then $G$ enforces ratios. If there exists a bipartition of $V = V_0 \cup V_1$ into two disjoint nonempty sets such that (1) the subgraphs $G_0$ and $G_1$ induced by $V_0$ and $V_1$ respectively enforce ratios, and (2) for every subset $A_0 \subseteq V_0$, if $A_1$ is the set of neighbors of $A_0$ in $V_1$, then $|A_1|/|V_1| \geq |A_0|/|V_0|$, then $G$ enforces ratios. (The ratios property where we instead choose $A_1 \subseteq V_1$ and neighbors $A_0$ in $V_0$ is equivalent.)

**Theorem 3.1** *Suppose $G$ enforces ratios. Then any subset $A \subseteq V$ with $|A| \leq |V|/2$ has at least $|A| \log_2(|V|/|A|)$ edges coming out of it.*

**Proof.** The argument proceeds inductively. Let $A_0 = A \cap V_0$ and $A_1 = A \cap V_1$. Assume without loss of generality that $|A_0|/|V_0| \geq |A_1|/|V_1|$. Let $C_0$ be the set of edges incident to $A_0$ and $V_0 \setminus A_0$. Let $C_1$ be the set of edges incident to $A_1$ and $V_1 \setminus A_1$. Finally let $C_{01}$ be the set of edges incident to $A_0$ and $V_1 \setminus A_1$.

Suppose $|A_0| \le |V_0|/2$. By inductive hypothesis, $|C_0| \ge |A_0|\log_2(|V_0|/|A_0|)$ and $|C_1| \ge |A_1|\log_2(|V_1|/|A_1|)$. Since $G$ enforces ratios, $A_0$ has at least $|V_1|(|A_0|/|V_0|)$ neighbors in $V_1$, so $|C_{01}| \ge |V_1|(|A_0|/|V_0|) - |A_1|$. The number of edges coming out of $A$ is thus at least

$$|C_0| + |C_1| + |C_{01}| \ge |A_0|\log_2(|V_0|/|A_0|) + |A_1|\log_2(|V_1|/|A_1|) + |V_1|(|A_0|/|V_0|) - |A_1|.$$

Let $b_0 = 2|A_0|$ and $b_1 = |V_1|(b_0/|V_0|)$. Then

$$|C_0| + |C_1| + |C_{01}| \ge |A_0|(\log_2(|V_0|/b_0) + log_2(b_0/|A_0|))$$

$$+|A_1|(\log_2(|V_1|/b_1) + \log_2(b_1/|A_1|)) + |V_1|(|A_0|/|V_0|) - |A_1|$$

$$\ge |A|\log_2(|V_0|/b_0) + |A_0| + |A_1| + (b_1/2) - |A_1|$$

$$\ge |A|\log_2(|V|/(b_0+b_1)) + |(b_0+b_1)/2)$$

$$\ge |A|\log_2(|V|/(b_0+b_1)) + |A|\log_2((b_0+b_1)/|A|)$$

$$\ge |A|\log_2(|V|/|A|)),$$

where the inequality before the last holds with equality at $|A| = (b_0+b_1)/2$.

The final case has $|A_0| \ge |V_0|/2$ and $|A_1| \le |V_1|/2$. The inductive hypothesis gives (i) $|C_0| \ge |V_0| - |A_0|$ and (ii) (i) $|C_1| \ge |A_1|$. Let $D_1$ be the neighbors of $A_0$ in $V_1$. Since $G$ enforces ratios, we have $|D_1|/|V_1| \ge |A_0|/|V_0| \ge 1/2$, hence some $E_1 \subseteq D_1$ satisfies $E_1 \cap A_1 = \emptyset$ and $|E_1| = |V_1|/2 - |A_1|$. Let $C'_{01}$ be he set of edges incident to $A_0$ and $E_1$; clearly (iii) $|C'_{01}| \ge |E_1| = |V_1|/2 - |A_1$. Now let $E'_1 = V_1 \setminus (E_1 \cup A_1)$, and by choice of $E_1$, $|E'_1| = |V_1|/2$. Let $E'_0$ be the set of neighbors of $E'_1$ in $V_0$. Since $G$ enforces ratios, $|E'_0|/|V_0| \ge |E'_1|/|V_1| \ge 1/2$, so $|E'_0 \cap A_0| \ge |A_0| - |V_0|/2$. Let $C''_{01}$ be the set of edges incident to $A_0$ and to $E'_1$; clearly (iv) $|C''_{01}| \ge |E'_0 \cap A_0| \ge |A_0| - |V_0|/2$. Now (i), (ii), (iii), and (iv) yield that the number of edges coming out of $A$ is at least $|C_0| + |C_1| + |C'_{01} + |C''_{01}| \ge |V|/2 \ge |A|\log_2(|V|/|A|)$ where the last inequality holds with equality at $|A| = |V|/2$. □

**Theorem 3.2** *Let $G$ be an $n$-vertex graph that enforces ratios. Then the virus reaches all vertices in expected $O(\log\log n)$ time.*

**Proof.** From Theorems 1.1 and 3.1 we obtain the bound

$$\sum_{1\le i<n/2} \frac{1}{i\log_2(n/i)} + \sum_{n/2\le i<n} \frac{1}{(n-i)\log_2(n/(n-i))}$$

$$\le 2\sum_{1\le i\le n/2} \frac{1}{i\log_2(n/i)}$$

$$\le \frac{2}{\log_2 n} + 2\sum_{1\le j\le\lceil\log_2 n\rceil-1} 2^{j-1}\frac{1}{2^{j-1}((\log_2 n) - j)}$$

$$\le 2\sum_{1\le j\le\log_2 n} \frac{1}{j}.$$

This proves the bound $O(\log\log n)$. □

If a graph $G$ enforces ratios, it can be viewed as a graph whose vertices are vertices of a hypercube $H$, by dividing $H$ into two subcubes $H_0$ and $H_1$ and then representing $V_0$ and $V_1$ in $H_0$

and $H_1$ respectively. Let $V$ be a set of vertices in a hypercube $H$. Then we may define a graph $G$ on the vertices $V$ by taking the convex hull of $V$, and the edges of this polytope define the edges $E$ of $G$. It was conjectured in [3] that all such graphs $G$ represented in a hypercube $H$ have cutset expansion 1, that is, if $|A| \leq |V|/2$ then the number of edges coming out of $A$ is at least $|A|$. By Theorem 1.2, if this conjecture is true, the virus will take $O(\log n)$ time to propagate; we shall later verify this bound under slighly more general circumstances, independently of the conjecture. The notion of enforcing ratios was introduced in [3] as a means to verifying this conjecture. It was defined in more general form by requiring the ratios property in every subcube $H'$ with respect to every division of $H$ into two subcubes $H'_0$ and $H'_1$. Notice that when ratios are enforced we obtain the stronger bound $O(\log\log n)$ on the propagation of the virus. The graphs $G$ for which the above geometric construction gives edges having Hamming distance in the cube at most 2 are precisely the bases exchange graph of delta-matroids (see e.g. [1] for a definition of delta-matroids), which include matroids as a subcase. The matroids that enforce ratios in this stronger sense are precisely the balanced matroids [2]. Other examples where this geometric construction has cutset expansion 1 are the matching polytope and the partial order polytope [3]. An example of balanced matroids where stronger bounds for the propagation of a virus is the uniform matroid, consisting of all vertices of the hypercube having the same number of 1s in the bit representation. For instance, at the level with just a single 1 in the bit representation, the uniform matroid is a clique, giving bound $O((\log n)/n)$. The hypercube itself can be viewed as a balanced matroid, giving bound $O(\log\log n)$ for the hypercube Applying Theorem 1.2 with $n = 2^d$ and $e' = d$ gives only the bound $\frac{\log n}{e'} = \Omega(1)$ for the hypercube.

We summarize the results as follows.

**Theorem 3.3** *If $G$ is the bases exchange graph of a balanced matroid, then the virus propagates in $O(\log\log n)$ time. In particular, if $G$ is the Cartesian product of $d$ cliques of size $k$ (the hypercube is the case $k = 2$), then $n = k^d$ and the virus propagates in expected time $O(\log d + \log\log k)$ and $\Omega(\frac{\log k}{k})$. Similarly, if $G$ is the bases exchange graph of a uniform matroid of size $k$ and rank $d$, we have $n = \binom{k}{d}$ and the virus propagates in time $O(\log d + \log\log k)$.*

These results will be improved in the next section.

## 4 Product of Cliques and the Uniform Matroid

**Theorem 4.1** *Let $G$ be the Cartesian product of $d$ cliques of size $k$, for a total of $n = k^d$ vertices. Then any set $V$ of size $s \leq n/2$ has at least $\epsilon sk(d + 1 - \frac{\log s}{\log k})$ edges coming out of $V$ for $\epsilon = 1/6$.*

**Proof.** We first prove at least $\delta ks$ edges come out of $V$ for some $\delta > 0$. If $d = 1$, the number of edges coming out of $V$ is $s(k - s) \geq sk/2$, proving the bound for $\delta \leq 1/2$. Suppose $d \geq 2$, and proceed by induction on $d$. Decompose $G$ into $k$ copies $G_1, \ldots, G_k$ of Cartesian products of $d - 1$ cliques of size $k$. Let $s_i$ be the number of vertices of $V$ in $G_i$, and suppose $s_i \leq s_j$ for $i \leq j$. Let $\ell$ be such that $s_\ell \leq n/(2k) \leq s_{\ell+1}$, or $\ell = k$ if no such $\ell$ exists. Then the number of edges coming out of $V$ is at least $(s_k - s_1)\ell(k - \ell) + \delta k(s_1 + \ldots + s_\ell + ((k - \ell)n/k) - s_{\ell+1} - \ldots - s_k)$. This quantity is only reduced if $s_1, \ldots, s_\ell$ are replaced by their average and $s_{\ell+1}, \ldots, s_k$ are replaced by their average. The number of edges coming out of $V$ is then at least $E = (s_k - s_1)\ell(k - \ell) + \delta k(\ell s_1 + (k - \ell)((n/k) - s_k))$. Suppose $s_1$ is increased by $\alpha/\ell$ and $s_k$ is decreased by $\alpha/(k - \ell)$. Then $E$ is decreased by $\alpha(k - \ell) + \alpha\ell - 2\delta\alpha = (1 - 2\delta)\alpha k \geq 0$ for $\delta \leq 1/2$. We may choose $\alpha$ so that $s_k \leq n/(2k)$, so $\ell = k$. Then the number of edges coming out of $V$ is at least $\delta k\ell s_1 = \delta ks$ for $\delta = 1/2$.

We show that any set $V$ of size $s \le n/2$ has at least $F = \epsilon sk(d+1-\frac{\log s}{\log k})$ edges coming out of $V$. by induction on $d$. If $s \ge n/k^2 = k^{d-2}$ then $F \le 3\epsilon ks \le \delta ks$ for $\epsilon \le 1/6$ and the bound follows. We may thus assume $s \le n/k^2$ and thus $s_i \le s \le n/2k$.

Suppose $s_i < s_{i+1} = \ldots = s_{i+a} < s_{i+a+1} = \ldots = s_{i+a+b} < s_{i+a+b+1}$, where $s_0 = 0$ and $s_{k+1} = n$. Then the number of edges coming out of $V$ has a term $G = (s_{i+a+1} - s_{i+1})ab + \epsilon s_{i+1}ak(d - \frac{\log s_{i+1}}{\log k}) + \epsilon s_{i+a+1}bk(d - \frac{\log s_{i+a+1}}{\log k})$. Letting $x = s_{i+1}a$, $y = s_{i+a+1}b$, and $s' = x + y$, we may fix $s', a, b$ and minimize over $x$. We have $G = (s'-x)a - xb + \epsilon skd + \epsilon xk(\frac{\log a - \log x}{\log k}) + \epsilon(s'-x)k(\frac{\log b - \log(s'-x)}{\log k})$. Differentiating with respect to $x$, we get $G' = -a - b + \epsilon k\frac{\log a}{\log k} - \epsilon k\frac{\log b}{\log k} + \frac{\epsilon k}{\log k}(\log(s'-x) - \log x)$. We note that $G'$ is a decreasing function of $x$, so the minimum of $G$ is achieved at the boundary values of $x$, which happen either when $s_{i+1} = s_{i+a+1}$ or when $s_{i+1} = s_i$ and $s_{i+a+1} = s_{i+a+b+1}$. This reduces the number of distinct values of $s_i$, and we may in the end assume $0 = s_1 = \ldots = s_a < s_{a+1} = \ldots = s_k$.

Letting $x = k - a$, we have $s_{a+1} = s/x$, and the number of edges coming out of $V$ is at least $H = (k-x)s + \epsilon sk(d - \frac{\log s - \log x}{\log k})$. Differentiating with respect to $x$, we get $H' = -s + \frac{\epsilon sk}{x \log k}$. We note that $H'$ is a decreasing function of $x$, so the minimum of $H$ is achieved at the boundary values of $x$, which happen at $x = k$ and $x = 1$. For $x = k$ we get $H = \epsilon sk(d - \frac{\log s - \log k}{\log k}) = F$ as required, and for $x = 1$ we get $H = s(k-1) + \epsilon sk(d - \frac{\log s}{\log k}) \ge F$ as required since $\epsilon \le 1/6$. □

**Theorem 4.2** *Let $G$ be the Cartesian product of $d$ cliques of size $k$, for a total of $n = k^d$ vertices. Then the virus reaches all vertices in expected time $O(\frac{(1+\log d)(\log k)}{k})$ and $\Omega(\frac{\log k}{k})$.*

**Proof.** Applying Theorem 1.1 and 4.1, we get for $s \le n/2$ that if $k^i \le s < k^{i+1}$ for $0 \le i \le d-1$ then $|E_s| \ge sk(d-i)/6$. For $0 \le j < \log k$, if $k^i e^j \le s < k^i e^{j+1}$, we get at most $k^i e^{j+1}$ terms where $|E_s| \ge k^i e^j k(d-i)/6$, so the sum of $1/|E_i|$ in this range of $j$ is at most $\frac{6e}{k(d-i)}$, and the sum in the range of $i$ is at most $\frac{6e(1+\log k)}{k(d-i)}$. Summing over all $i$ gives $O(\frac{(1+\log d)(\log k)}{k})$. Applying Theorem 1.2 with $n = k^d$ and $e' = kd$ gives $\frac{\log n}{e'} = \Omega(\frac{\log k}{k})$. □

The bases exchange graph $G(k,d)$ of the uniform matroid of size $k$ and rank $0 \le d \le k$ has $\binom{k}{d}$ vertices consisting of all $k$-bit strings with exactly $d$ ones, and with two such strings adjacent if and only if they differ in exactly two bits.

**Theorem 4.3** *If $1 \le d \le \sqrt{k/3}$ then the virus reaches all vertices of $G(k,d)$ in expected $O(\frac{(1+\log d)(\log k)}{k})$ time. If $1 \le d \le k^{1-\delta}$ for some constant $\delta > 0$ then the virus reaches all vertices of $G(k,d)$ in expected $\Omega(\frac{\log k}{k})$ time.*

**Proof.** Embed $G(k,d)$ in the product of $d$ cliques of size $k$ by representing each $k$-bit string with exactly $d$ ones with the $d!$ vertices $(i_1, \ldots, i_d)$ such that the set of positions in the $k$-bit string that are ones is $\{i_1, \ldots, i_d\}$. If there were originally $|E_s|$ edges coming out of a set of size $s$, then the number $|F_s|$ of edges coming out of the corresponding set of size $d!s$ satisfies $|F_s| = d!|E_s| + d!sd(d-1)$. Since $|F_s| \ge d!sk/2$ from the proof of Theorem 4.1 and $d(d-1) \le k/3$, we have that $|E_s| \ge sk(d+1-\frac{\log(d!s)}{\log k})/36$. Proceeding as in Theorem 4.2 with $k^i e^j \le d!s < k^i e^{j+1}$, we get at most $k^i e^{j+1}/d!$ terms where $|E_s| \ge k^i e^j k(d-i)/(36d!)$, so the sum of $1/|E_i|$ in this range of $j$ is at most $\frac{36e}{k(d-1)}$, and the sum in the range of $i$ is at most $\frac{36e(1+\log k)}{k(d-i)}$. Summing over all $i$ gives $O(\frac{(1+\log d)(\log k)}{k})$. Applying Theorem 1.2 with $n = \binom{k}{d}$ and $e' = (k-d)d$ gives $\frac{\log n}{e'} = \Omega(\frac{\log k}{k})$ for $1 \le d \le k^{1-\delta}$ given a constant $\delta > 0$. □

## 5 The Grid

**Theorem 5.1** *Let $G$ be the 2-dimensional grid with sides having $k$ vertices, for a total of $n = k^2$ vertices. The number of edges coming out of a set of size $s$ is at least $f_2(s) = \min(2\sqrt{s}, k, 2\sqrt{k^2 - s})$. Then the virus reaches all vertices in expected time $O(k) = O(\sqrt{n})$.*

**Proof.** Let $V$ be a set of vertices of $G$ having $s$ vertices. We lower bound the number of edges coming out of $V$. Divide $G$ into $k$ copies of one dimension, and divide $V$ correspondingly into sets $V_1, V_2, \ldots, V_k$. The number of edges coming out of $V_i$ in the corresponding dimension is at least 1 unless $|V_i| = 0$ or $|V_i| = k$, in which case it is 0. In the other direction, we may reorder the $V_i$ in decreasing order $|V_1| \geq |V_2| \geq \ldots \geq |V_k|$, and the number of edges coming out of $V$ in that dimension is at least $|V_1| - |V_k|$. By possibly exchanging dimensions and possibly replacing $V$ with $V(G) - V$, we may assume $|V_1| < k$. Consider the largest $x$ such that $|V_x| > 0$. The number of edges coming out is only reduced if we replace each $V - i$ for $1 \leq i \leq x$ with a set of size the average $y$ of the sizes of $V_1, V_2, \ldots, V_x$. We thus have $V$ a rectangle with $xy = s$. The number of edges coming out of $V$ is $k$ if $x = k$ and $y < k$, and is $x + y$ if $x, y < k$. The quantity $x + y$ over $xy = s$ is minimized at $x = y = \sqrt{s}$.

The number of edges coming out of $V$ is thus at least $f_2(s) = \min(2\sqrt{s}, k, 2\sqrt{k^2 - s})$. This function can be rewritten as $f_2(s) = 2\sqrt{s}$ for $s \leq k^2/4$, $f_2(s) = k$ for $k^2/4 \leq s \leq 3k^2/4$, and $f_2(s) = 2\sqrt{k^2 - s}$ for $s \geq 3k^2/4$.

Applying Theorem 1.1, we obtain the upper bound $\sum_{1 \leq s < n = k^2} \frac{1}{f_2(s)}$ which is $O(k) = O(\sqrt{n})$. $\square$

**Theorem 5.2** *Let $G$ be a $d$-dimensional grid with $k$ vertices per side, for a total of $n = k^d$ vertices. Let $f(s) = ds^{1-\frac{1}{d}}$ for $s \leq \frac{n}{e^d}$, $f(s) = \frac{es}{k} \log \frac{n}{s}$ for $\frac{n}{e^d} \leq s \leq \frac{n}{2}$, $f(s) = f(n - s)$ for $s \geq \frac{n}{2}$. Then there exists a constant $\epsilon > 0$ such that the number of edges coming out of a set $V$ of size $s$ is at least $\epsilon f(s)$ for $d \geq 2$.*

**Proof.** We note that the function $f$ is concave down. We also note that the first expression for $f(s)$ is always at least as large as the second expression. To see this, write $s = n/e^{d(x+1)}$, and the first expression gives $f(s) = \frac{eds}{k} e^x$, while the second expression gives $f(s) = \frac{eds}{k}(x + 1)$, withe equality at $x = 0$.

The case $d = 2$ was shown in Theorem 5.1. Suppose the result holds inductively for $d$. We show the result for a $(d + 1)$-dimensional product with $nk = k^{d+1}$. Divide $G$ into $k$ $d$-dimensional grids $G_1, \ldots, G_k$. Let $V$ be a set of size $s \leq n/2$. Let $V_i$ be the set of elements of $V$ in $G_i$, with $|V_i| = s_i$. Let $M = \max_i |V_i|$ and $Q = \min_i |V_i|$. Then the number of edges coming out of $V$ is at least $M - Q + \epsilon(f(s_1) + \ldots + f(s_k))$. For $a < b$, suppose there are $A$ values of $i$ for which $s_i = a$ and $B$ values of $j$ for which $s_j = b$, and let $s' = Aa + Bb$. If we increase $b$ and decrease $a$ so as to leave $A$, $B$, $s'$ fixed, say by replacing $b$ with $b + \alpha/B$ and replacing $a$ with $a - \alpha/A$, then the quantity $Af(a) + Bf(b)$ does not increase, since $f$ is concave down. We may thus modify the $s_i$ until all $s_i$ equal either $M$ or $Q$, thus obtaining a decomposition $k = A+B$, where $A, B \geq 0$ are not necessarily integer, so that the number of edges coming out of $V$ is at least $M - Q + \epsilon(Af(Q) + Bf(M))$.

Suppose the induction can be carried through with some $\epsilon'$ in the case where either $Q = 0$ or $M = n$. If $M - Q \geq \delta M$, then we may replace two $V_i$ of size $Q$ and $M$ by $V_i$ of size 0 and $M + Q$ respectively, or if $M + Q > n$, by $V_i$ of size and $Q - n + M$ and $n$ respectively. This replaces the term $M - Q \geq \delta M$ by $M + Q \leq 2M$, so the induction carries through with $\epsilon = \delta\epsilon'/2$.

Suppose instead $M - Q < \delta M$ with $Q = t$, $M = t(1 + \mu)$, $\mu/(1 + \mu) < \delta$. If $f(t) = dt^{1-1/d}$ and $f(t(1 + \mu)) = d(t(1 + \mu))^{1-1/d}$, then $f(t(1 + \mu)) = dt^{1-1/d}(1 + (1 - 1/d)\mu - (1/d)O(\mu^2))$.

We may thus replace $Q$ and $M$ by a common value $t(1+\nu)$ with $0<\nu<\mu$, with a decrease $t\mu-\epsilon t^{1-1/d}O(\mu^2)$. If $f(t)=(et/k)\log(n/t)$ and $f(t(1+\mu))=(et(1+\mu)/k)\log(n/(t(1+\mu)))$, then $f(t(1+\mu))=(et/k)\log(n/t)+(et/k)(\log(n/t)-1)\mu-(t/k)O(\mu^2)$, this time with a decrease $t\mu-\epsilon(t/k)O(\mu^2)$. In the mixed case with $t<n/e^d$ and $t(1+\mu)>n/e^d$, the two expressions above have identical coefficients at $t=n/e^d$ for $1,\mu$, and $O(\mu^2)$ so the decrease goes through as well. Finally at $t=n/2$ there is a decrease of $t\mu-\epsilon(t/k)O(\mu)$ for $\epsilon$ small enough.

We may thus suppose either $Q=0$ or $M=n$. If $M\geq n/2$ then $M-Q\geq n/2\geq\epsilon f(s)$ for $\epsilon$ small enough. Finally if $M<n/2$ then $Q=0$ and the number of edges coming out of $V$ is at least $L=M+\epsilon Bf(M)$. Setting $x=M$ and $B=s/x$, this expression becomes $L=x+\epsilon(s/x)f(x)$. We may consider the two cases of the definition of $f$. In the first case $L\geq\epsilon(x+(s/x)dx^{1-1/d})=\epsilon(x+dsx^{-1/d})$, and the derivative with respect to $x$ is $\epsilon(1-sx^{-1-1/d})$, showing that the minimum occurs at $x=s^{d/(d+1)}$, giving $L\geq\epsilon(d+1)s^{1-1/(d+1)}\geq\epsilon f_{d+1}(s)$. and completing the induction. In the second case $L\geq\epsilon e(x+(s/x)(x/k)\log(n/x))=\epsilon e(x+(s/k)(\log n-\log x))$, and the derivative with respect to $x$ is $\epsilon e(1-s/(kx))$, showing that the minimum occurs at $x=s/k$, giving $L\geq\epsilon e(s/k)(1+\log n-\log s+\log k)\geq\epsilon((se/k)\log((nk)/s)$. This quantity is at least $\epsilon f_{d+1}(s)$ if $s$ corresponds still to the second case of $f(s)$, otherwise we may switch to the first case at $x=n/e^d$. It remains to show that $f_{d+1}(s)\leq kf(x)$ for $x=s/k$. If $f_{d+1}(s)$ is given by the second case in the definition of $f_{d+1}$, then we have $(es/k)\log(nk/s)=k(ex/s)\log(n/x)\leq kdx^{1-1/d}$ as observed in the beginning of the proof, so the result follows for either case of the definition of $f(x)$. If $f_{d+1}(s)$ is given by the first case in the definition of $f_{d+1}$, then $s\leq nk/e^{d+1}$, so $x\leq n/e^{d+1}$ and $f(x)$ is given by the first case in the definition of $f$. We must show $(d+1)s^{1-1/(d+1)}\leq kd(s/k)^{1-1/d}$, which is equivalent to $s\leq(1-1/(d+1)^{d(d+1)}nk$. This follows from $s\leq nk/e^{d+1}$, since $1/e\leq(1-1/(d+1))^d$ because $e^{1/d}\geq 1+1/d$. □

**Theorem 5.3** *Let $G$ be a $d$-dimensional grid with $k$ vertices per side, for a total of $n=k^d$ vertices. Then a virus propagates to the whole grid in expected time $O(k(1+\log d))$ and $\Omega(\frac{k}{\log k+\log d}+\log k)$.*

**Proof.** We use Theorems 1.1 and 5.2. We first consider the range where $f(s)=ds^{1-\frac{1}{d}}$. We divide this range into $e^{di+j}\leq s<e^{di+j+1}$ for $0\leq j<d$ and $0\leq i<\log k$. In this range $f(s)\geq de^{di+j}/e^{i+1}$ and the sum of $1/f(s)$ is at most $e^{i+2}/d$. Summing over all $j$ gives $e^{i+2}$. Summing over all $i$ gives $2e^3k=O(k)$.

We next consider the range where $f(s)=\frac{es}{k}\log\frac{n}{s}$. We divide this range into $n/e^{i+1}\leq s\leq n/e^i$ for $0\leq i<d$. In this range $f(s)\geq(n/(ke^i))i$ and the sum of $1/f(s)$ is at most $k/i$ for $i>0$. The case $i=0$ also assumes $s\leq n/2$ and gives $O(k)$ like the case $i=1$ as well. Summing $k/i$ over all $1\leq i<d$ gives $O(k\log d)$, proving the upper bound.

Using Theorem 1.2 with $n=k^d$ and $e'=2d$ gives $(\log n)/e'=\Omega(\log k)$. We show the bound $\Omega(\frac{k}{\log k+\log d})$.

The $d$-dimensional grid is a product of paths whose elements can be labeled $1,2,\ldots,k$. For any $d\leq w\leq kd$, the plane $S_w$ consists of those vertices such that the sum of the elements labeled in each path is $w$. Select a small constant for $kd/4\leq w\leq 3kd/4$. Applying Theorem 2.1 with $D=kd/2$, and $t\log t=D$ gives $t/d=D/(d\log D)=k/(\log k+\log d)$ as claimed. □

# 6 Ratios Revisited

We define a family of $k$-vertex graphs $H_k$ inductively. The graph $H_1$ is just a single vertex. If $k\geq 2$, the graph $H_k$ consists of a path $v_1,v_2,\ldots,v_k$, a copy of some inductively constructed $H_{\lfloor k/2\rfloor}$

on the vertices $v_{2i}$, and a copy of some inductively constructed $H_{\rceil k/2\rceil}$ on the vertices $v_{2i+1}$. Clearly $H_k$ enforces ratios with the obvious bipartition of the vertices.

Suppose we are given a set $U_k$ of $k$ vertices and a subset $V$ of $U$ of size $s \leq k/2$. We shall describe a randomized process that assigns a graph $H_k$ to the vertices $U$. Choose a random bipartition of $U$ into two subsets $U_{\lfloor k/2\rfloor}$ to be assigned $H_{\lfloor k/2\rfloor}$ and $U_{\lceil k/2\rceil}$ to be assigned $H_{\lceil k/2\rceil}$. Choose the path $v_1, v_2, \ldots, v_k$ joining these two subsets by assiging to the vertices of $V$ the vertices $v_i$ in each subset of least possible values of $i$.

**Lemma 6.1** *The expected number of edges coming out of the set $V$ of size $s \leq k/2$ in $H_k$ is $O(s\log(k/s))$.*

**Proof.** We may assume $s \geq k^{0.9}$ since otherwise $s\log(k/s) \geq 0.1s\log k$ and the degree of $H_k$ is $O(\log k)$. We expect to divide $s$ into two parts $s_1, s_2$ of about $s/2$ in $H_{\lfloor k/2\rfloor}$ and $H_{\lceil k/2\rceil}$. We may assume $|s_1 - s_2| \leq k^{.8}$, since by Chernoff's bound the probability of this not being the case is at most $e^{-k^{2(0.8)-1}} = e^{-k^{0.6}}$, which can only affect the expectation by a negligible $(s\log k)e^{-k^{0.6}}$.

We proceed by halving $k_i = k/2^i$ and $s_i = s/2^i$ until we reach $\hat{\imath}$ such that $s_i \leq k_i^{0.9}$ or $s_i \leq c$ for some large constant $c$. The error introduced in $s_i = s_{\hat{\imath}}$ is at most

$$\frac{k^{.8}}{2^i} + \frac{(k/2)^{.}8}{2^{i-1}} + \frac{(k/2^2)^{.}8}{2^{i-2}} + \ldots + \frac{(k/2^i)^{.}8}{2^0}$$

$$\leq (s_i)^{0.8/0.9}(1 + 2^{-0.2} + 2^{-0.4} + \cdots + 2^{-(0.2)i})$$

$$\leq s_i c^{-0.1}/(1 - 2^{-0.2})$$

$$\leq s_i \epsilon$$

for some arbitrarily small $\epsilon > 0$ dependent on the large $c$.

Therefore the number of edges leaving $V$ because of the paths $v_1, v_2, \ldots, v_k$ at all iterations of the induction is at most $O(\epsilon s)$. At the bottom of the recurrence if $s_i \leq c$ then $k_i \leq s_i^{1/0.9} \leq c^{0.2} = O(1)$, so $O(s)$ is incurred in total; and if $s_i \leq k_i^{0.9}$ is reached while $2s_i \geq (2k_i)^{.9}$, then for some constant $c' > 0$ the number of edges coming out of $V$ is at most $s\log k_i \leq s\log(c'(k_i/s_i)^1 0 = O(s\log(k/s))$. □

We say that an n-vertex graph $G$ has logarithmic expansion $p > 0$ if for every $s$-vertex subset $V$ of the vertices of $G$ with $s \leq n/2$, the number of edges coming out of $V$ in $G$ is at least $ps\log_2(n/s)$. Clearly if $G$ enforces ratios then $G$ has logarithmic expansion $p = 1$ by Theorem 3.1. By scaling by $p$ the argument of Theorem 3.2 it follows that if $G$ has logarithmic expansion $p > 0$ then the virus reaches all vertices in expected time $O(\frac{\log\log n}{p})$.

We say that an $n$-vertex graph $G$ has recursive logarithmic expansion $p > 0$ if it satisfies the following inductive definition. If $n = 1$ then $G$ has logarithmic expansion $p$. If (1) there exists $k \geq 2$ such that $V(G)$ can be partitioned into $k$ sets $V_1, V_2, \ldots, V_k$, with each $V_i$ of size $n/k$, (2) each $V_i$ induces a graph $G_i$ that has recursive logarithmic expansion $p$, (3) the set $V(G)$ can also be partitioned into $n/k$ sets $W_1, \ldots, W_{n/k}$, with each $W_j$ of size $k$, such that $|V_i \cap W_j| = 1$ for each $1 \leq i \leq k$ and $1 \leq j \leq n/k$, and (4) each $V_j$ induces a graph $R_j$ that has logarithmic expansion $p$, then $G$ has recursive logarithmic expansion $p$.

**Theorem 6.2** *If an $n$-vertex graph $G$ has recursive logarithmic expansion $p$, then $G$ has logarithmic expansion $\Omega(p)$, and thus a virus propagates to the entire graph $G$ in expected time $O(\frac{\log\log n}{p})$. Two examples of this are Cartesian products of $d$ cliques of size at least $k$ and at most $K$, giving*

$p = \frac{k}{2\log_2 k}$ *and* $\log\log n = O(\log d + \log\log K)$*, for virus propagation in* $O(\frac{(\log k)(\log d + \log\log K)}{k})$ *expected time; and Cartesian products of* $d$ *paths on at most* $K$ *vertices each, giving* $p = \frac{1}{K}$ *and* $\log\log n = O(\log d + \log\log K)$*, for virus propagation in* $O(K(\log d + \log\log K))$*.*

**Proof.** Let $V$ be a set of size $s \le n/2$ We replace each $R_j$ with a random graph $H_k$ as in Lemma 6.1. The partition into $H_{\lfloor k/2 \rfloor}$ and $H_{\lceil k/2 \rceil}$ of the indices $i = 1, 2, \ldots, k$ is uniformly the same for all $R_j$, but the paths $v_1, v_2, \ldots, v_k$ for $H_k$ depend on $V \cap V(R_j)$. Since $R_j$ has logarithmic expansion $p$, we obtain from Lemma 6.1 that after substitution of each $R_j$ with $H_k$ the expected number of edges coming out of $V$ in $H_k$ is at most $O(1/p)$ the number of edges coming out of $V$ in $R_j$. The partition of $G$ into two groups of $G_j$ of sizes $\lfloor k/2 \rfloor n/k$ and $\lceil k/2 \rceil n/k$ given by the partition of indices in $H_k$ satisfies the definition of enforcing ratios. Therefore by Theorem 3.1 after substitution the number of edges coming out of $V$ is at least $s \log_2(s/n)$; thus before substitution this number of edges is at least $\Omega(ps\log(s/n))$, that is, $G$ has logarithmic expansion $\Omega(p)$. The argument in Theorem 3.2 then shows that the virus reaches all vertices in $G$ in expected $O(\frac{\log\log n}{p})$ time. □

# 7 Trees and Diameter

**Theorem 7.1** *Let* $G$ *be an* $n$*-vertex tree of diameter* $d$*. Then the expected time to propagate a virus has matching upper and lower bounds* $\Theta(d + \log n)$*.*

**Proof.** Consider a path $P$ of length $c(d + \log n)$ and fire in $P$ from one end along side with the tree $T$. If $P'$ is a path from the starting virus vertex of length at most $d$, the process chooses to add a vertex from $P$ or $P'$ with equal probability, giving a binomial experiment which guarantees that the end of $P'$ will be reached with probability at least $n^{-b}$ before the end of $P$ is reached, by letting $b$ be any constant sufficiently smaller than $c$, by the Chernoff bound. Thus with probability at least $n^{1-b}$ all possible paths $P'$ will be covered, and the probability that a firing will take more than $n^{2-b}$ is exponentially small and decreases exponentially fast. This proves the $O(d + \log n)$ time.

A path $P'$ from the virus of length at least $d/2$ gives the expected time $\Omega(d)$. The number of edges coming out of a set of size $i$ is at most $n - i$, and the bound $\sum_{1 \le i < n} \frac{1}{n-i}$ gives a lower bound $\Omega(\log n)$, for a combined $\Omega(d + \log n)$ matching lower bound. □

We have mentioned a polytope $P$ defined as the convex hull of a set of vertices in the hypercube. The edges of this polytope define a graph $G(P)$ for which cutset expansion 1 has been conjectured [3]. If this conjecture holds, it follows that $O(\log n)$ time suffices for the virus to reach all vertices in $G(P)$. We shall define a combinatorial subgraph $G$ of $G(P)$ and show that $O(\log n)$ time suffices for the virus to reach all vertices in $G$, and thus in $G(P)$. For delta-matroids $M$, the polytope $G(P)$ and the combinatorial graph $G$ give the bases-exchange graph of $M$, and those $G(P)$ whose edges have Hamming distance at most 2 are precisely the bases-exchange graphs of delta-matroids $M$.

Given a set $V$ of $n$ vertices in the hypercube, if $x, y \in V$ are at Hamming distance $h \ge 1$, and the $h$-cube $H$ containing both $x, y$ contains a $h-1$-subcube $H'$ of size half the size of $H$, and the only vertex of $V$ in $H'$ is either $x$ or $y$, then include in the combinatorial graph $G$ the edge $(x, y)$. It is easy to verify that $(x, y)$ is also an edge of $G(P)$, but there are examples where edges of $G(P)$ are not in $G$. In the case of delta-matroids $G$ and $G(P)$ coincide.

Let $x \neq y$ be two vertices of $V$. Flip some bits of $x, y$ so that $x \leq y$, that is, $x_i \leq y_i$ in each bit position $i$. A special path from $x$ to $y$ is a sequence $x = x^0 \leq x^1 \leq \cdots \leq x^d = y$ of vertices in $V$ such that each $(x^i, x^{i+1})$ is an edge of $G$. Let the special distance $\hat{d}(x, y)$ be the least $d$ such that such a special path of length $d$ exists in $G$.

**Lemma 7.2** *If $d(x, y) = d$ in the combinatorial graph $G$, then $G$ has at least $n \geq 2^d$ vertices.*

**Proof.** Consider the path from $x$ to $y$ of length $d$ just described. We proceed by induction. The path from $x$ to $x^{d-1}$ of length $d-1$ is a shortest special path, so the smallest subcube $C$ containing both $x$ and $x^{d-1}$ has at least $2^{d-1}$ vertices of $V$ by inductive assumption. Let $i$ be a bit position such that $0 = x_i^{d-1} \neq y_i = 1$. The smallest subcube containing both $x$ and $y$ must contain a $z$ such that $z_i = 1$ and $(x, z)$ is an edge in $G$. The shortest special path from $z$ to $y$ has at length at least $d - 1$, so by inductive assumption (as $z$ is closer than $x$ to $y$) the smallest subcube $C'$ containing both $z$ and $y$ has at least $2^{d-1}$ vertices of $V$. Furthermore $C$ and $C'$ are disjoint, as $t$ in $C$ satisfies $t_i = 0$ for the chosen $i$, while $t'$ in $C'$ satisfies $t'_i = 1$. Combining $C$ and $C'$ gives $2^{d-1} + 2^{d-1} = 2^d$ vertices of $G$ in the smallest subcube containing $x$ and $y$. □

**Theorem 7.3** *In the combinatorial polytope $G$, and thus also in $G(P)$, the virus reaches all vertices in $O(\log n)$ time. This holds in particular for the bases-exchange graph of delta-matroids.*

**Proof.** If $d$ is the diameter of $G$, then a breadth-first search tree of $G$ has diameter at most $2d$, and thus by Theorem 7.1 the virus reaches all vertices in $O(d + \log n)$ time. The graph $G$ must have two vertices at special distance at least $d$, so by Lemma 7.2 we have $n^{\geq} 2^d$, so $d \leq \log_2 n$ and the time for the viruses is $O(d + \log n) = O(\log n)$. □

## 8 Viruses and Antidotes

In this section we consider a modification of the basic problem where two vertices in the graph $G$ are initially chosen, one vertex $v$ to start a virus and one vertex $w$ to start an antidote. The sets of vertices $V$ and $W$ that have the virus and the antidote grow by selecting a random edge coming out of $V \cup W$, where each such edge is chosen according to a Poisson process as before. The question is what fraction of the graph will be taken over by the virus versus the antidote.

If $G$ is a path with $n$ vertices, and initially one end of the path is the virus $v$ and the other end of the path is the antidote $w$, then the process corresponds to tossing $n - 2$ fair coins, where heads adds one virus vertex and tails adds one antidote vertex. This gives a binomial distribution for the number $k$ of vertices that end up contracting the virus, with probability $\binom{n-2}{k} 2^{-n+2}$ for each $0 \leq k \leq n - 2$.

The next case we consider is that of a clique $K_n$ of size $n$. Suppose at some point in time $x$ vertices have the virus and $y$ vertices have the antidote. Then the probability that exactly $k$ out of the remaining $n - x - y$ vertices will end up with the virus is given by the recurrence

$$f(n, x, y, k) = \frac{x}{x+y} f(n, x+1, y, k-1) + \frac{y}{x+y} f(n, x, y+1, k),$$

$$f(n, x, n-x, 0) = 1,$$

$$f(n, x, y, -1) = 0,$$

$$f(n, x, y, n-x-y+1) = 0.$$

This recurrence resolves to

$$f(n,x,y,k) = \frac{(x+y-1)!(x+k-1)!(n-x-y)!(n-x-k-1)!}{(x-1)!(y-1)!(n-x-y-k)!k!(n-1)!}.$$

In particular, $f(n,1,1,k) = \frac{1}{n-1}$ for each $0 \leq k \leq n-2$, giving a uniform distribution for the number $k$ of vertices that endup with the virus starting with a single virus vertex and a single antidote vertex.

**Theorem 8.1** *The problem with a single initial virus vertex and a single initial antidote vertex gives to the number of vertices that contract the virus a binomial distribution in the case of a path and a uniform distribution in the case of a clique.*

# References


[1] T. Feder and D. Ford, Classification of bipartite Boolean constraint satisfaction through delta-matroid intersection, Electronic Colloquium on Computational Complexity, ECCC TR05-016, 2005. SIAM J. Discrete Math., to appear.

[2] T. Feder and M. Mihail, Balanced Matroids, Proc. 24th Ann. ACM Symp. on Theory of Computing (1992) 26–38.

[3] M. Mihail and M. Sudan, Connectivity properties of matroids, TR-89, U.C. Berkeley.