

# Near-perfect token distribution

A. Z. Broder<sup>\*</sup>    A. M. Frieze<sup>†</sup>    E. Shamir<sup>‡</sup>    E. Upfal<sup>§</sup>

May 22, 2006

## Abstract

Suppose that  $n$  tokens are arbitrarily placed on the  $n$  nodes of a graph. At each parallel step one token may be moved from each node to an adjacent node. An algorithm for the *near-perfect token distribution* problem redistributes the tokens in a finite number of steps, so that, at the end, no more than  $O(1)$  tokens reside at each node. (In perfect distribution, at the end, exactly one token resides at each node.)

In this paper we present a simple algorithm that works for all *extrovert* graphs, a new property which we define and study. In terms of connectivity requirements, extrovert graphs are roughly in-between expanders and compressors. Our results lead to an optimal solution for the near-perfect token distribution problem on almost all cubic graphs. The new solution is conceptually simpler than previous algorithms, and applies to graphs of minimum possible degree.

---

<sup>\*</sup>DEC Systems Research Center, 130 Lytton Ave, Palo Alto, CA.

<sup>†</sup>Department of Mathematics, Carnegie-Mellon University. A portion of this work was done while the author was visiting DEC SRC. Supported in part by NSF Grant CCR 0089112.

<sup>‡</sup>Department of Computer Science, Hebrew University, Jerusalem. A portion of this work was done while the author was visiting DEC SRC. Supported in part by grant 438/89 of the Israeli Academy of Sciences.

<sup>§</sup>IBM Almaden Research Center, San Jose, CA, and the Weizmann Institute of Science, Rehovot, Israel.

# 1 Introduction

Suppose that  $n$  tokens are arbitrarily placed on the  $n$  nodes of a graph, with no more than  $M$  tokens at each site. At each parallel step one token may be moved from each node to an adjacent node. An algorithm for the *near-perfect token distribution* problem redistributes the tokens in a finite number of steps, so that, at the end, no more than  $O(1)$  tokens reside at each node. (In perfect distribution, at the end, exactly one token resides at each node.)

Algorithms for the near-perfect token distribution have direct applications to fundamental problems in parallel and distributed systems. We give two examples here: First, consider each node as a processor, each edge as a communication line, and each token as a message. An efficient solution to the near-perfect token distribution problem can be combined with an efficient permutation routing algorithm to give an efficient routing algorithm for an arbitrary communication request. Second, consider each token as an independent process awaiting execution and each node as a processor. Then an efficient algorithm for the near-perfect token distribution problem can be used for workload distribution among the processors.

These, and other applications, require a solution in a fully distributed model, where all information processing are done by processors that reside at the nodes of the graph. When the computation is initiated, each process has no information on the status (e.g. number of tokens) of other processors in the system. Information is transmitted via messages along the edges of the graph. In each step a processor can either send one token through an edge, or one  $O(\log n)$ -bit message. Furthermore, most applications require an efficient solution on graphs that are as sparse as possible.

Peleg and Upfal [7] have shown that any solution to the near-perfect token distribution problem requires  $\Omega(M + \log n)$  parallel steps, and this runtime can be achieved only on expander graphs. They gave an optimal  $O(M + \log n)$  algorithm for *sqrt*  $(\alpha, \beta)$ -*expander* graphs. These are  $n$  node,  $d$ -regular expander graphs with the following additional strong properties: there exist constants  $\alpha > 1$  and  $\beta > 0$  such that any set of vertices  $X$ , with  $|X| \leq \beta n$ , has at least  $\alpha |X|$  neighbors outside  $X$  and there exists an  $i$  with  $0 < i \leq d$  such that  $\alpha > (d + i^2)/(i + 1)$ . These two conditions imply that the expansion is at least  $\sqrt{d}$ .

In this paper we present an optimal solution for almost all cubic graphs, which is the best one can hope for from the point of view of the minimum degree required.

Our solution is based on identifying a novel property, called *extrovertness* which replaces the *sqrt*-expander property of [7]. Extrovertness is stronger than the expansion property but weaker than the compression property of [8], that is, every extrovert graph is an expander, and every compressor graph is an extrovert graph. (The exact relations are discussed in Section 3.) In Section 2 we present an asymptotically optimal solution for any extrovert graph, which immediately leads to a solution for any graph  $G$  such that for some constant  $k$ , the graph  $G^k$  or the graph

$\cup_{i \leq k} G^i$  is extrovert. In particular, we show that for any expander  $G$  there is a (large) constant  $k$  that depends only on its degree and on its expansion properties such that  $G^k$  is extrovert.

The main result of this paper, which is perhaps of independent significance, is given in Section 4: we show directly that for almost all cubic graphs  $G$ , the graph  $G \cup G^2$  (that is, the graph of all paths of length at most two in  $G$ ) is extrovert. (In fact, it can be even shown that for almost all cubic graphs the graph  $G^2$  is extrovert, but the proof is rather intricate and the constants are much worse.)

A preliminary version of this paper has appeared in [5].

## 2 The New Algorithm

Our new approach to the token distribution problem is based on the *extrovertness* property:

**Definition 1** *A vertex  $v \in S \subseteq V$  is called extrovert in  $S$  if  $v$  has strictly more neighbors in  $V \setminus S$  than it has in  $S$ .*

**Definition 2** *A set  $S \subseteq V$  is called  $b$ -extrovert if at least  $b|S|$  of its vertices are extrovert.*

**Definition 3** *A graph  $G = (V, E)$  is called  $(a, b)$ -extrovert if any set  $S \subseteq V$  of size at most  $a|V|$  is  $b$ -extrovert.*

Our algorithm gives an optimal solution for any  $n$ -node  $(a, b)$ -extrovert graph  $G = (V, E)$ .

Let  $\Delta$  be the maximum degree in  $G$  and let  $\gamma = (2\Delta + 1)/(2\Delta + 2)$ . The algorithm works in  $\log_{1/\gamma} M$  phases. (Recall that  $M$  is the initial maximum number of tokens in any node.)

Let  $t_j(v)$  be the number of tokens at node  $v$  at the start of phase  $j$ , and let  $M_j = \max_v \{t_j(v)\}$ . Let  $U_j = \{v : t_j(v) \geq M_j/2\}$  denote the set of “heavily loaded” nodes.

Each phase reduces the maximum number of tokens in any node by a factor of  $\gamma$ , until no vertex has more than  $2/a = O(1)$  tokens. Note that in any phase  $|U_j| \leq a|V|$  because  $M_j > 2/a$  and the total number of tokens is  $|V|$ .

At the beginning of phase  $j$  the algorithm constructs a dag (directed acyclic graph)  $D_j = (V_j, E_j)$  with the following properties:

1.  $U_j \subseteq V_j \subseteq V$ . (All heavily loaded vertices are in the dag.)
2. Each edge in  $E_j$  is the directed version of some edge in  $G$ , and each edge in  $G$  has at most one directed version in  $E_j$ .

3. For every  $v \in U_j$ ,  $\text{outdegree}(v) > \text{indegree}(v)$ .

After the dag is constructed, phase  $j$  consists of  $M_j/(2\Delta + 2)$  iterations; in each iteration each node in  $V_j$  sends one token (if it has one) through each of its outgoing edges in  $D_j$ .

**Lemma 1**  $M_{j+1} \leq \gamma M_j$ .

*Proof:* Let  $v$  be an arbitrary vertex in  $G$ . If  $v$  is not in  $D_j$ , and therefore not in  $U_j$ , then it does not receive (or send) any tokens in phase  $j$ , and thus it has less than  $M_j/2 < \gamma M_j$  tokens at the end of this phase.

If  $v$  is in the dag but not in  $U_j$ , it starts the phase with less than  $M_j/2$  tokens, and it may receive in this phase no more than  $\Delta M_j/(2\Delta + 2)$  tokens. Thus at the end of the phase it has no more than  $(2\Delta + 1)/(2\Delta + 2)M_j = \gamma M_j$  tokens.

Finally, if  $v \in U_j$ , it sends more tokens than it receives in each iteration. Thus, it loses at least  $M_j/(2\Delta + 2)$  tokens in phase  $j$  and has less than  $\gamma M_j$  tokens at the end of the phase.  $\square$

For a set of vertices  $S \subseteq V$  let  $\Phi(S) = \{v \in S : v \text{ is extrovert}\}$ . Given an  $(a, b)$ -extrovert graph we construct the dag  $D_j$ , “bottom-up,” using the following procedure.

1. Set  $W \leftarrow U_j$ .
2. Repeat until  $W = \emptyset$ :
  - (a) Connect all vertices in  $\Phi(W)$  to all of their  $G$  neighbors outside  $W$ .
  - (b)  $W := W \setminus \Phi(W)$ .

**Lemma 2** *The above procedure constructs a dag with the required properties, and terminates in  $O(\log N)$  iterations.*

*Proof:* Clearly when the procedure terminates, all of  $U_j$  is in the dag, and each edge in  $G$  was used no more than once. When a vertex  $v \in U_j$  joins the dag it has more neighbors in the dag than in the remaining  $W$ , thus  $\text{outdegree}(v) > \text{indegree}(v)$  in the final dag. Since the graph  $G$  is  $(a, b)$ -extrovert, and  $|W| \leq |U_j| \leq a|V|$ , in each iteration  $|\Phi(W)| \geq b|W|$ , and the procedure terminates within  $O(\log n)$  iterations.  $\square$

**Corollary 3** *If  $G$  is an  $(a, b)$ -extrovert graph, then the algorithm solves the near-perfect token distribution problem on  $G$  in  $O(\log M)$  phases. When phase  $j - 1$  ends, no vertex has more than  $\gamma^{j-1}M$  tokens and phase  $j$  takes  $O(\gamma^{j-1}M + \log n)$  parallel steps.*

If we execute the  $O(\log M)$  phases sequentially the total run time is  $O(M + \log n \log M)$ . Peleg and Upfal [7] have shown that all the dags for the last  $O(\log \log n)$  phases can be constructed in parallel. Since  $\log n$  becomes dominant only when  $M_j \leq \log n$ , this suffices to reduce the total run-time to  $O(M + \log n)$ . The same idea can be applied to our algorithm. We briefly sketch the construction here, the reader is referred to [7] for a more detailed description.

The difficulty in constructing the dags in parallel is that in order to construct the dag  $D_j$  one needs to know the set  $U_j$ , and the set  $U_j$  is defined in turn, by the dag  $D_{j-1}$ . This difficulty is overcome by replacing the set  $U_j$  in the construction, by an overestimate set  $\tilde{U}_j$  that contains  $U_j$  and is easier to compute. We also compute, for each node, an overestimate of the number of tokens it will have at the end of each phase.

The overestimate sets are constructed *sequentially* as follows: Assume that we have already constructed the set  $\tilde{U}_j$ , and that each node has a bound on the number of tokens it will have at the end of phase  $j - 1$ . We run the first  $c \log \log n$  phases of the construction of  $D_j$ , assuming that  $U_j = \tilde{U}_j$ . Let  $\tilde{W}_j$  denote the set  $W$  at the end of these  $O(\log \log n)$  phases. We fix the number of phase we run so that  $|\tilde{W}_j| = O(n/(\log n)^2)$ .

Now we compute for each node, a bound on the number of tokens it will have at the end of phase  $j$ . If a node  $v$  has a neighbor in  $\tilde{W}_j$ , it assumes that this neighbor will be connected to it in  $D_j$ , and will send to it  $\gamma^{j-1}M$  tokens. Besides this, if  $v$  is in the partially computed dag, it can compute in  $O(\log \log n)$  time, a bound on how many tokens it will receive in phase  $j$ . Thus, every node which potentially is in the dag  $D_j$ , has an estimate on the number of tokens it will receive in phase  $j$ . Vertex  $v$  is in  $\tilde{U}_{j+1}$  if according to this estimate it has more than  $\gamma^j M/2$  tokens, at the end of phase  $j$ .

These estimates add no more than  $\Delta |\tilde{W}_j| M_j$  “dummy” tokens in each phase. We can fix the parameters so that no more than  $n$  dummy tokens are created in the whole process, thus, this procedure can be executed until  $M_j \leq 4/b$ . Once the sets  $\tilde{U}_j$  are constructed, the  $O(\log \log n)$  dags can be constructed in parallel in  $O(\log n)$  parallel steps, with messages of size  $O(\log \log n)$  bits. Thus, we prove:

**Theorem 4** *The near-perfect token distribution problem can be solved on any  $(a, b)$ -extrovert graph  $G$  in  $O(M + \log n)$  time.*

Let  $H(G, k) = \bigcup_{j \leq k} G^j$ , for some constant  $k$  (i.e., the graph  $H(G, k)$  has an edge for any path of length  $k$  or less in  $G$ ). Clearly, a parallel step on the graph  $H(G, k)$  can be simulated on the original graph  $G$  in constant time. Thus, our result applies to any graph  $G$  such that  $H(G, k)$ , or one of its subgraphs, is extrovert for some constant  $k$ , in particular to any expander.

The main result of this paper, which we prove in Section 4, is that if  $G$  is a random cubic graph, then with high probability the graph  $H = G \cup G^2$  is extrovert. Thus we prove:

**Theorem 5** *There is an  $O(M + \log n)$  solution for the near-perfect token distribution problem on almost all cubic graphs.*

### 3 Extrovert graphs, expanders, and compressors

We already noticed that in terms of connectivity requirements, extrovert graphs are roughly in-between expanders and compressors. We now proceed to make this statement more precise.

There are many slightly different definitions of expanders; here we use the following:

**Definition 4** *A graph  $G = (V, E)$  is called an  $(a, b)$ -expander if for every set  $S \subseteq V$  of size at most  $a|V|$*

$$|\Gamma(S) \setminus S| \geq b|S|,$$

where  $\Gamma(S)$  is the set of neighbors of  $S$ , that is

$$\Gamma(S) = \{u \in V \mid \exists v \in S, \{u, v\} \in E\}.$$

When we talk about expanders without further qualifications, we mean an  $(a, b)$ -expander with  $a \geq 1/2$ .

Following [6], we define compressors as follows:

**Definition 5** *A graph  $G = (V, E)$  is called an  $a$ -compressor if for every set  $S \subseteq V$  of size at most  $a|V|$  the set*

$$\{v \in V \mid v \text{ has more neighbors in } S \text{ than outside } S\}$$

has cardinality at most  $|S|/2$ .

Clearly any  $a$ -compressor graph is  $(a, 1/2)$ -extrovert; and any  $d$ -regular  $(a, b)$ -extrovert graph is a  $(a, b/2)$ -expander since the number of edges going from a set  $S \subseteq V$  of size at most  $a|V|$  to  $V \setminus S$  is at least  $b|S|d/2$ .

In the opposite direction, consider an  $(a, b)$ -extrovert  $d$ -regular graph  $G = (V, E)$  and consider a set  $S \subseteq V$  of size at most  $a|V|/3$ . Let  $S'$  be the set of vertices in  $V$  that have more neighbors in  $S$  than outside  $S$ . Clearly since the total degree of  $S$  is  $d|S|$ , the set  $|S \cup S'|$  has size at most  $3|S|$ , and therefore  $|S \cup S'| \leq a|V|$ . But any vertex in  $S'$  is not extrovert in  $S \cup S'$ . Hence

$$|S'| \leq (1 - b)|S \cup S'| \leq (1 - b)(|S| + |S'|).$$

Therefore  $|S'| \leq \frac{1-b}{b}|S|$ . We conclude that any  $(a, b)$ -extrovert  $d$ -regular graph  $G$ , with  $b > 2/3$  is an  $a/3$ -compressor.

We now turn to the relation between expanders and extrovert graphs. Let  $G = (V, E)$  be a  $d$ -regular  $(1/2, b)$ -expander. Let  $\lambda_1 = d > \lambda_2 \geq \dots \geq \lambda_n$  be the eigenvalues of  $G$ . Let  $S \subseteq V$  be a set of vertices of cardinality  $a|V|$ . Denote  $S : \bar{S}$  the set of edges  $\{\{i, j\} \in E \mid i \in S, j \in V \setminus S\}$ . It is easy to show (see e.g. Alon and Chung [2]) that

$$|S : \bar{S}| \geq |S|(1 - a)(d - \lambda_2).$$

Assume that

$$(1 - a) \left(1 - \frac{\lambda_2}{d}\right) = \frac{1 + b}{2} > \frac{1}{2}$$

Then the average number of edges in  $|S : \bar{S}|$  per vertex in  $S$  is  $(1 + b)d/2$  and the maximum is  $d$ . Hence there are at least  $b|S|$  extrovert vertices in  $S$ . We conclude that any expander with  $\lambda_2 < d/2$  is extrovert. (Since for any infinite family of graphs  $\{G_n\}$ , it is known that  $\liminf \lambda_1(G_n) > 2\sqrt{d-1}$ , the result above is always too weak to prove extrovertness for any infinite family of graphs with  $d < 15$ .)

For a weaker expanders we consider the multigraph  $G^k$ . It has degree  $d^k$  and its second largest eigenvalue (assuming  $G$  is not bipartite) is  $\lambda_2^k$ . For multigraphs we relax the definition of an extrovert vertex as follows: A vertex  $v \in S \subseteq V$  is called *extrovert* in  $S$  if  $v$  has strictly more edges going to vertices in  $V \setminus S$  than edges going to vertices in  $S$ . Then the same computation as above shows that  $G^k$  is extrovert, if

$$1 - \left(\frac{\lambda_2}{d}\right)^k > \frac{1}{2}.$$

Alon [1] has shown that if  $G$  is a  $d$ -regular  $(1/2, b)$ -expander, then

$$\lambda_2 \leq d - \frac{b^2}{4 + 2b^2}.$$

Therefore it suffices to take

$$k \geq \frac{d(4 + 2b^2) \ln 2}{b^2}.$$

Thus, for any expander  $G$ , there exists a constant  $k$  that depends only  $d$  and the second eigenvalue of  $G$  (a measure of expansion), such that  $G^k$  is extrovert. In particular, there exists a constant  $k$  such that if  $G$  is a random cubic graph, then with high probability  $G^k$  is extrovert; however the constant is too large to be practical for any application. (In the next section we show directly that if  $G$  is a random cubic graph, then with high probability  $G \cup G^2$  is extrovert.)

## 4 Extrovert graphs via random cubic graphs

We first state three properties that almost all random cubic graphs possess; then we show that if a cubic graph  $G$  has these properties, then the graph  $H = G \cup G^2$  is extrovert.

In the definition of these properties  $G = (V, E)$  is a cubic graph on  $n$  vertices,  $S$  is a subset of  $V$ , and  $i(S)$  is the number of edges that are contained in  $S$ . All constants were chosen for convenience and no attempt was made to optimize them.

**Property 1** *If  $S \subseteq V$  and  $|S| \leq (\ln n)/4$  then  $i(S) \leq |S|$ .*

**Property 2** *There exists a function  $a > 0$  such that for any  $\epsilon > 0$ , and for any  $S \subseteq V$ , if  $|S| \leq a(\epsilon)n$  then  $i(S) \leq (1 + \epsilon)|S|$ .*

**Property 3** *There are fewer than  $\ln \ln n$  vertices on cycles of length less than  $\ln \ln n$*

We will now prove that almost all cubic graphs have these three properties. We will use the model of Bender and Canfield [3] and Bollobás [4]. Specifically, we will adopt the notation of [4].

Let  $[k]$  stand for the set  $\{1, \dots, k\}$ . Let  $W = [3n]$  and let  $W_i = \{3(i-1) + 1, 3(i-1) + 2, 3i\}$ , for  $i = 1, 2, \dots, n$  be a partition of  $W$  into  $n$  sets of size 3, called *blocks*. For  $w \in W$  we define  $\psi(w) = \lceil w/3 \rceil$  so that  $w \in W_{\psi(w)}$ . The elements of  $W$  are called *points* to distinguish them from the vertices of the graph.

A configuration is a partition of  $W$  into  $m = 3n/2$  pairs. The set of configurations is denoted  $\Phi$ . For  $F \in \Phi$  let  $\mu(F)$  be the multigraph with vertex set  $[n]$  and  $m$  edges  $\{\{\psi(x), \psi(y)\} : \{x, y\} \in F\}$ .

We consider  $\Phi$  as a probability space in which each  $F \in \Phi$  is equally likely. In this setting, for large  $n$ , the probability that  $\mu(F)$  is a simple graph is at least  $1/10$  and conditional on this, each cubic graph with vertex set  $[n]$  is equally likely. We need only show therefore that almost all  $\mu(F)$  have the above three properties.

*Proof of Property 1:*

$$\begin{aligned} \Pr(\text{Prop. 1 fails}) &\leq \sum_{k=4}^{(\ln n)/4} \mathbf{E}(\# \text{ of } k\text{-sets with at least } k+1 \text{ edges}) \\ &\leq \sum_{k=4}^{(\ln n)/4} \binom{n}{k} \binom{3k}{k+1} \left(\frac{k}{n}\right)^{k+1} \leq \sum_{k=4}^{(\ln n)/4} \left(\frac{ne}{k}\right)^k 2^{3k} \left(\frac{k}{n}\right)^{k+1} \\ &= \frac{1}{n} \sum_{k=4}^{(\ln n)/4} k(8e)^k = o(1) \end{aligned}$$

Explanation of second inequality:  $\binom{n}{k}$  counts sets of  $k$  blocks. Focus now on the blocks  $W_1, W_2, \dots, W_k$ . If vertex set  $\{1, 2, \dots, k\}$  contains more than  $k$  edges then there are  $k+1$  pairs contained in  $X = W_1 \cup W_2 \cup \dots \cup W_k$ . We can choose the lowest numbered point of each pair in at most  $\binom{3k}{k+1}$  ways. Having fixed these as  $a_1, a_2, \dots, a_k$  say, the probability that they are each paired in  $X$  is at most  $(k/n)^{k+1}$ , for given the pairings of  $a_1, a_2, \dots, a_i$ , the probability that  $a_{i+1}$  is paired in  $X$  is at most  $k/n$ .  $\square$



*Proof of Property 2:* For definiteness let

$$a(\epsilon) = \left( \frac{(1+\epsilon)^{1+\epsilon}}{2e^{2+\epsilon}3^{1+\epsilon}} \right)^{1/\epsilon}.$$

Then

$$\begin{aligned} \Pr(\text{Prop. 2 fails}) &\leq \sum_{k=4}^{a(\epsilon)n} \binom{n}{k} \binom{3k}{(1+\epsilon)k} \left( \frac{k}{n} \right)^{(1+\epsilon)k} \\ &\leq \sum_{k=4}^{a(\epsilon)n} \left( \frac{ne}{k} \left( \frac{3ek}{(1+\epsilon)n} \right)^{1+\epsilon} \right)^k \\ &= \sum_{k=4}^{a(\epsilon)n} \left( \frac{e^{2+\epsilon}3^{1+\epsilon}}{(1+\epsilon)^{1+\epsilon}} \left( \frac{k}{n} \right)^\epsilon \right)^k = o(1). \end{aligned}$$

(Split the range at, say,  $k = \sqrt{n}$ . If  $k \leq \sqrt{n}$  then  $k/n \leq 1/\sqrt{n}$  so the sum up to  $k = \sqrt{n}$  is  $O(1/n^{2\epsilon})$ ; for  $k \geq \sqrt{n}$  we use the fact that the summand is at most  $2^{-k}$ .)  
□

*Proof of Property 3:* Say that a cycle is small if it has at most  $\ell = \ln \ln \ln n$  vertices. Then

$$\begin{aligned} \mathbf{E}(\# \text{ of vertices on small cycles}) &\leq \sum_{k=3}^{\ell} k \binom{n}{k} \frac{1}{2} (k-1)! 6^k \left( \frac{1}{3n-2k} \right)^k \\ &\leq \sum_{k=3}^{\ell} 2^k \leq 2^{\ln \ln \ln n + 1} = o(\log \log n). \end{aligned}$$

and so we can use the Markov inequality. (Explanation of first inequality: having chosen  $k$  vertices and a cyclic ordering in  $\binom{n}{k} \frac{1}{2} (k-1)!$  ways, there are  $3^k$  choices for the points in pairs not in edges of the cycle and  $2^{k-1}$  ways of pairing up the points in the blocks to make the edges of the cycle. The probability that these  $k$  pairs exist is at most  $(3n-2k)^{-k}$ .) □

**Theorem 6** *If  $G$  is a cubic graph that has properties 1, 2, and 3, then the graph  $H = G \cup G^2$  is extrovert.*

*Proof:* For a set  $S \subseteq V$  let

$$\Phi(S) = \{v \in S : v \text{ is extrovert in } S \text{ with respect to } H\}$$

and let  $g(S) = |\Phi(S)|$ .

From now until the end of the proof all adjacency and neighborhood relations, when not specified otherwise, are with respect to  $G$ .

Let  $\epsilon$  be a sufficiently small constant and let  $S$  be a set of vertices with  $|S| \leq \frac{1}{4}a(\epsilon)n$ . Let  $T = \Gamma(S)$  be the set of neighbors of  $S$  in  $G$ . (Thus  $|S \cup T| \leq a(\epsilon)n$ .) Let  $i = i(S)$ , the number of edges in  $S$ . Now partition  $S$  and  $T$  as follows: Let

$$\begin{aligned} S_j &= \{v \in S : v \text{ has } j \text{ neighbors in } S\}, & j = 0, 1, 2, 3. \\ T_j &= \{v \notin S : v \text{ has } j \text{ neighbors in } S\}, & j = 1, 2, 3. \end{aligned}$$

Finally let  $s = |S|$ ,  $s_j = |S_j|$ , and  $t_j = |T_j|$ .

We first prove that  $S$  is extrovert in  $H$  under two assumptions:

1. No vertex of  $S$  lies on a cycle of length  $\leq 4$  in  $G$ .
2.  $|S| \geq s^* > 0$  where  $s^*$  is an absolute constant.

These assumptions are removed in the last part of the proof.

Under Assumption 1, each vertex in  $S$  has 9 distinct neighbors in  $H$ . For  $v$  to be extrovert with respect to  $S$  in  $H$ , it must have at least 5 neighbors in  $H$  outside  $S$ . Having two  $G$ -neighbors in  $T_1$  is sufficient, but if  $v$  has no neighbor in  $S_3$ , then one  $G$ -neighbor in  $T_1$  is sufficient.

There are two cases in the proof. In the first case,  $i \geq (1 - \alpha)s$  for a suitably chosen  $\alpha > 0$ . In this case  $T_2 \cup T_3$  is quite small, and the proof works readily; In the second case,  $i < (1 - \alpha)s$ , and the graph induced by  $S$  breaks into  $k \geq \alpha s$  components out of which  $k' \geq (1 - 2\alpha)k$  are trees of size at most  $\alpha^{-2}$ . Here we estimate the number of extrovert vertices in terms of  $\alpha k$ .

We start by deriving a series of relations among the  $s_j$ 's and the  $t_j$ 's. From Property 2 applied to  $S \cup T_2 \cup T_3$  we deduce that

$$i + 2t_2 + 3t_3 \leq (1 + \epsilon)(s + t_2 + t_3),$$

or

$$(1 - \epsilon)t_2 + (2 - \epsilon)t_3 \leq (1 + \epsilon)s - i. \quad (1)$$

Let

$$S_2^g = \{v \in S_2 : v \text{ has no neighbor in } S_3\}$$

and let  $s_2^g = |S_2^g|$ . Observe that if  $v \in S_2^g$  is adjacent to  $w \in T_1$  then  $v \in \Phi(S)$ . Clearly

$$s_2^g \geq s_2 - 3s_3 \quad (2)$$

### Claim 7

$$s_1 \geq s_3 - 2\epsilon s$$

*Proof:* Consider the graph  $G_S$  induced by  $S$ . Let  $C$  be a component of  $G_S$  with at least one edge. Let  $C$  have  $\sigma_i$  vertices of degree  $i$ , for  $i = 1, 2, 3$ , and let  $\sigma = |C|$ . Due

to Property 2, by removing at most  $\epsilon\sigma + 1$  edges from  $C$  we can obtain a spanning tree  $C'$ . Observe that in  $C'$

$$\sigma'_1 = \sigma'_3 + 2.$$

Now add back the removed edges. Each additions reduces the difference between the number of degree 3 vertices and degree 1 vertices by at most two. Hence

$$\sigma_1 - \sigma_3 \geq \sigma'_1 - \sigma'_3 - 2(\epsilon\sigma + 1) = -2\epsilon\sigma.$$

Now add up over all the components of  $S$ .  $\square$

Now from equations (2) and Claim 7 we obtain that

$$\begin{aligned} s &= s_0 + s_1 + s_2 + s_3 \\ &\leq s_0 + s_1 + s_2^g + 4s_3 \\ &\leq s_0 + s_1 + s_2^g + 4s_1 + 8\epsilon s \end{aligned}$$

Hence

$$s_0 + 5s_1 + s_2^g \geq (1 - 8\epsilon)s. \quad (3)$$

Now choose  $\alpha$  small but large compared with  $\epsilon$ . There are two cases.

**Case 1:**  $i \geq (1 - \alpha)s$ .

We deduce from (1) that

$$2t_2 + 3t_3 \leq \frac{2}{1 - \epsilon}(\alpha + \epsilon)s \leq 3\alpha s. \quad (4)$$

The number of edges coming from  $T_2 \cup T_3$  into  $S$  is  $2t_2 + 3t_3$ . Delete from  $S$  all the vertices hit by such edges. A remaining vertex  $v$  has out-of- $S$  neighbors only in  $T_1$ , thus  $v$  has at least two such neighbors if  $v \in S_0 \cup S_1$  and at least one such neighbor if  $v \in S_2^g$ . Hence

$$\begin{aligned} g(S) &\geq s_0 + s_1 + s_2^g - (2t_2 + 3t_3) \\ &\geq \frac{1}{5}(s_0 + 5s_1 + s_2^g) - 3\alpha s \\ &\geq \left(\frac{1 - 8\epsilon}{5} - 3\alpha\right)s \end{aligned} \quad (5)$$

**Case 2:**  $i < (1 - \alpha)s$ .

In this case the number of components  $k$  of  $G_S$  satisfies

$$k \geq \alpha s. \quad (6)$$

The number of components of  $G_S$  of size greater than  $s/(\alpha k)$  is at most  $\alpha k$ . Hence there are at least  $(1 - \alpha)k$  components of size  $< s/(\alpha k) \leq \alpha^{-2}$ .

**Claim 8** *If  $i < (1 - \alpha)s$  then the number of components of  $S$  of size  $\leq \alpha^{-2}$  that are trees is at least  $(1 - 2\alpha)k$ .*

*Proof:* We can assume without loss of generality that  $G_{S \cup T}$  is connected.

If  $|S \cup T| \geq \ln n / \ln \ln n$  then we use Property 3 to assert that at most  $o(k)$  of the small components are not trees.

If  $|S \cup T| < \ln n / \ln \ln n$  then at most one of the small components is not a tree; otherwise two non-tree components can be connected via  $T$  and contradict Property 1. By Assumption 2

$$s \geq s^* > \alpha^{-2} \tag{7}$$

so  $\alpha k \geq \alpha^2 s \geq 1$ .

Hence in both cases the number of small components that are not trees is less than  $\alpha k$ .  $\square$

Suppose now that the  $k$  components have sizes  $c_1, c_2, \dots, c_k$ , and that these components span  $i_1, i_2, \dots, i_k$  edges. We re-write equation (1) as

$$\begin{aligned} (1 - \epsilon)t_2 + (2 - \epsilon)t_3 &\leq (1 + \epsilon) \sum_{1 \leq t \leq k} c_t - \sum_{1 \leq t \leq k} i_t \\ &= \sum_{1 \leq t \leq k} (c_t - i_t) + \epsilon s \leq k + \epsilon s. \end{aligned}$$

Thus

$$t_2 + 2t_3 \leq \frac{k + \epsilon s}{1 - \epsilon}. \tag{8}$$

Now we divide the components of  $s$  of size  $\leq \alpha^{-2}$  that are trees into three classes:

- $k_1$  trees that are isolated vertices.
- $k_2$  trees that are paths of length  $\geq 1$ .
- $k_3$  trees that have at least one degree 3 vertex.

We now classify the leaves of these trees. Within these trees there are

- $l_1$  isolated vertices adjacent to  $T_1$ , and thus  $k_1 - l_1$  other vertices triply adjacent to  $T_2 \cup T_3$ .
- $l_2$  end-of-path leaves adjacent to  $T_1$ , and thus  $2k_2 - l_2$  other vertices doubly adjacent to  $T_2 \cup T_3$ .
- $l_3$  other leaves doubly adjacent to  $T_1$ , and thus at least  $3k_3 - l_3$  other vertices adjacent to  $T_2 \cup T_3$ .

Note that  $l_1 + l_2 + l_3$  counts extrovert vertices with respect to  $H$ . Counting edges out of  $T_2 \cup T_3$  we get the inequality

$$2t_2 + 3t_3 \geq 3(k_1 - l_1) + 2(2k_2 - l_2) + (3k_3 - l_3)$$

or

$$3l_1 + 2l_2 + l_3 \geq (3k_1 + 4k_2 + 3k_3) - (2t_2 + 3t_3)$$

and hence

$$l_1 + l_2 + l_3 \geq (k_1 + k_2 + k_3) - \frac{2}{3}(t_2 + 2t_3).$$

Now using Claim 8 and equation (8) and assuming  $\epsilon < 1/4$ , we obtain that

$$\begin{aligned} l_1 + l_2 + l_3 &\geq (1 - 2\alpha)k - \frac{2}{3} \frac{k + \epsilon s}{1 - \epsilon} \\ &\geq \left(\frac{1}{9} - 2\alpha\right)k - \frac{8\epsilon}{9}s \end{aligned}$$

Hence (via (6))

$$g(S) \geq ((1 - 18\alpha)\alpha - 8\epsilon)s/9 \tag{9}$$

Equations (5) and (9) conclude the proof under the two initial assumptions, since they show that for any set  $S \subseteq V$ , if  $|S| \leq a(\epsilon)n$ , then  $g(S)/|S|$  is greater than a certain constant. We now proceed to remove the assumptions that we made.

Let  $b$  be the lower bound on  $g(S)/|S|$  derived above. Suppose first that  $|S| \geq (\ln n)/4$ . Remove the  $k \leq \ln \ln n$  vertices in  $S$  that lie on cycles of length  $\leq \ln \ln \ln n$ , thus obtaining  $S'$ . All extrovert vertices in  $S'$  are extrovert in  $S$  with the possible exceptions of vertices at distance 1 or 2 from the removed vertices. (Because the removed vertices are on cycles, there are only  $3k$  neighbors.) Therefore

$$g(S) \geq g(S') - 3k \geq (b - o(1))|S|.$$

Suppose now that

$$\left(8 + \frac{24}{b}\right) \leq |S| \leq \frac{1}{4} \ln n$$

By Property 1, there is at most one cycle in  $S \cup \Gamma(S) \cup \Gamma^2(S)$ . If this cycle has length less than 4 and contains vertices of  $S$ , remove them to obtain  $S'$ . Then, as before

$$g(S) \geq g(S') - 12 \geq b(s - 4) - 12 \geq \frac{bs}{2}.$$

It remains to consider the case  $2 \leq |S| \leq s^*$ , where  $s^* = \max\{8 + 24/b, \alpha^{-2}\}$ , to accommodate also equation (7). We shall argue that  $S$  contains at least one extrovert vertex.

Consider the graph  $K$  that consists of all edges incident to  $S$  in  $G$ . (In other words,  $K$  is the graph induced by  $S \cup T$  except for the edges whose both endpoints are in  $T$ .) Without loss of generality we can assume that  $K$  is connected.

Clearly  $K$  has the following properties:

- Any  $v \in S$  has degree 3 in  $K$ .
- Any leaf in  $K$  belongs to  $T_1$ .
- Any neighbor of a leaf in  $K$  belongs to  $S$ .
- There is at most one cycle in  $K$ .

If  $K$  is a tree then there must be a  $v \in S$  that is the neighbor of two leaves in  $K$ , and therefore is extrovert in  $S$ . Similarly if  $K$  contains a cycle, then if there is any  $v \in S$  that does not belong to the cycle, then there must be a  $v' \in S$  that is the neighbor of two leaves in  $K$ . (Consider the tree that “hangs” at  $v$  away from the cycle.) It remains to consider the case when all the vertices in  $S$  belong to a cycle in  $K$ . But then all vertices in  $S$  are extrovert.  $\square$

## References

- [1] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6:83–96, 1986.
- [2] N. Alon and F. R. K. Chung. Explicit construction of linear sized tolerant networks. *Discrete Mathematics*, 72:15–19, 1989.
- [3] E. A. Bender and E. R. Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory B*, 24:296–307, 1978.
- [4] B. Bollobás. *Random Graphs*. Academic Press, 1985.
- [5] A. Z. Broder, A. M. Frieze, E. Shamir, and E. Upfal. Near-perfect token distribution. In *Proceedings of the 19th International Colloquium on Automata, Languages, and Programming*, volume 623 of *Lecture Notes in Computer Science*, pages 308–317. Springer Verlag, 1992.
- [6] C. Dwork, D. Peleg, N. Pippenger, and E. Upfal. Fault tolerance in networks of bounded degree. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, pages 370–379, 1986.
- [7] D. Peleg and E. Upfal. The token distribution problem. *SIAM Journal of Computing*, 18:229–243, 1989.
- [8] N. Pippenger. On networks of noisy gates. In *Proceedings of the 26th Annual Symposium on Foundations of Computer Science*, pages 30–38, 1985.