Testing Equality in Communication Graphs

Noga Alon, Klim Efremenko, and Benny Sudakov

***Abstract*— Let $G = (V, E)$ be a connected undirected graph with k vertices. Suppose that on each vertex of the graph there is a player having an n-bit string. Each player is allowed to communicate with its neighbors according to a (static) agreed communication protocol, and the players must decide, deterministically, if their inputs are all equal. What is the minimum possible total number of bits transmitted in a protocol solving this problem ? We determine this minimum up to a lower order additive term in many cases. In particular, we show that it is $kn/2 + o(n)$ for any Hamiltonian k-vertex graph, and that for any 2-edge connected graph with m edges containing no two adjacent vertices of degree exceeding 2 it is $mn/2 + o(n)$. The proofs combine graph theoretic ideas with tools from additive number theory.**

***Index Terms*— Communication complexity, equality function, static protocols, 2-connected graphs.**

I. The problem

LET $G = (V, E)$ be a connected undirected graph with k vertices. Suppose that on each vertex of the graph there is a player having an n-bit string. Each player is allowed to communicate with its neighbors according to an agreed communication protocol, and the players must decide, deterministically, whether or not their inputs are all equal. This is a natural, basic communication problem, which may also arise as an initial step of additional computational tasks, when the players want to check that they all have copies of the same input before starting to process it. The protocols we consider here are only those that Liang and Vaidya call *static protocols* in [12]. In these protocols, which player speaks and when is determined in advance, and is independent of the inputs. It is worth noting that in more general message passing protocols, where the inputs are allowed to influence the origin and destination of the messages sent, there are more efficient communication protocols. Indeed, although this fact does not appear explicitly in [10], the authors of that paper observed that in this more flexible model it is possible to solve the equality testing problem in the complete graph on k vertices with total communication $O(nk/\log(k))$, whereas in the static model we consider here $\Omega(nk)$ is a simple lower bound. See also [9], [8] for additional variants of the flexible model.

In a trivial protocol the players fix a rooted spanning tree of the graph, and each of them, besides the one at the root, transmits his bits to his parent, and each one (including the root) checks that his input is equal to those he received from each of his children. This shows that a total communication of roughly $(k-1)n$ bits suffices. Somewhat surprisingly, it turns out that for complete graphs G with at least 3 vertices one can do better. It is shown in [12] that for $G = K_k$ at least $kn/2$ bits of communication are needed, and the authors also obtain a nontrivial upper bound (which is not tight). Brody [6] has used the graphs constructed in [3] to show that for $G = K_3$, $3n/2 + o(n)$ bits suffice, showing that the lower bound is tight in this case up to a low order additive error term. In [3] we mentioned (without giving a detailed proof) that we can use a hypergraph extension of the construction in [3] to show that for $G = K_k$ the minimum possible number of bits in a communication protocol for the above problem on G is $(1 + o(1))kn/2$. Brody and Håstad [6] have independently found a similar protocol, using the k-cliques of the graphs in [3].

Here we consider the case of general graphs G, obtaining upper and lower bounds which are nearly tight in many (but not all) cases. Our upper bounds are based on an extension of the graphs of Ruzsa and Szemerédi [14], similar to the extension given in [1]. We also observe that linear communication protocols cannot improve the trivial upper bound. Finally, we suggest two competing conjectures about the possible answer for every graph.

Let $f(n, G)$ denote the minimum number of bits transmitted in a communication protocol solving the problem on G. It is clear that the function $f(n, G)$ is sub-additive, and hence by Fekete's Lemma (see [11]) the limit of the ratio $f(n, G)/n$ as n tends to infinity exists. Denote this limit by $f(G)$. The parameter $f(G)$ is the main object of study in the present short paper.

II. Results

Recall that a graph is 2-connected if it is connected and stays connected after a removal of any single vertex. It is 2-edge connected if it is connected even after removing any single edge. A *block* of a graph is a maximal two-connected subgraph, where every bridge (an edge not contained in a cycle) is also a block. It is well known that any graph is the edge-disjoint union of its blocks, and the vertices belonging to more than one block are the cut vertices of the graph. An end-block is a block that intersects at most one other block. It is known that any connected graph contains at least one

Manuscript received November 6, 2016; revised June 12, 2017; accepted July 23, 2017. Date of publication August 24, 2017; date of current version October 18, 2017. N. Alon was supported in part by the USA–Israeli BSF under Grant 2012/107, in part by the ISF under Grant 620/13, and in part by GIF under Grant G-1347-304.6. K. Efremenko was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant 257575. B. Sudakov was supported by SNSF under Grant 200021-149111.

N. Alon is with the Sackler School of Mathematics, Tel Aviv University, Tel Aviv 69978, Israel, and also with the Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel (e-mail: nogaa@tau.ac.il).

K. Efremenko is with the Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel (e-mail: klimefrem@gmail.com).

B. Sudakov is with the Department of Mathematics, ETH Zurich, 8092 Zürich, Switzerland (e-mail: benjamin.sudakov@math.ethz.ch).

Communicated by A. Rudra, Associate Editor for Complexity and Cryptography.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2017.2744608

end-block, see, e.g., [5], Section 5.2. Our first observation is the following.

Proposition 1: For any connected graph G with blocks $G_1, G_2, \ldots, G_s$,

$$f(G) = \sum_{i=1}^{s} f(G_i).$$

For a connected graph G let $c_2(G)$ denote the minimum number of edges in a 2-edge connected graph C obtained from G by taking all vertices of G, and some of its edges, where edges are allowed to be taken twice. Thus, for example, for any Hamiltonian graph G on k vertices, $c_2(G) = k$, as shown by a Hamilton cycle C in G. For a tree G on k vertices, $c_2(G) = 2(k-1)$, as shown by the graph C consisting of two copies of every edge of the tree. It is easy to see that for any graph G with k vertices $c_2(G) = k$ if and only if G is Hamiltonian. Our main upper bound for $f(G)$ is the following.

Theorem 2: For any connected graph G, $f(G) \leq 0.5\, c_2(G)$

The proof of this upper bound is the main contribution of the paper.

We proceed with the description of the (simple) lower bound. It is worth noting that this lower bound also follows from the work in [8].

Definition 3: For a connected graph G let $\mathcal{S}$ denote the set of all cuts in G. For any edge e of G let $\mathcal{S}_e$ denote the set of all cuts containing e. A fractional packing of cuts in G is a function $g : \mathcal{S} \mapsto [0,1]$ so that for every edge e of G, $\sum_{(S,\overline{S})\in\mathcal{S}_e} g(S,\overline{S}) \leq 1$. Let $fc(G)$ denote the maximum possible value of $\sum_{(S,\overline{S})\in\mathcal{S}} g(S,\overline{S})$, where the maximum is taken over all fractional packings of cuts g.

Theorem 4: For any connected graph $G = (V,E)$, $f(G) \geq fc(G)$.

Note that this implies that $f(G) \geq k/2$ for any k-vertex graph, as the function assigning the value $1/2$ to all cuts determined by single vertices, that is, all cuts of the form $(\{v\}, V-\{v\})$ for $v \in V$, is a fractional packing of cuts. Note also that clearly if G' is a spanning subgraph of G then $f(G') \geq f(G)$ and hence the above $k/2$ lower bound also follows from the fact that $f(K_k) = k/2$.

By the last theorem $f(G) \geq \alpha(G)$ for every G, where $\alpha(G)$ denotes the maximum size of an independent set in G. Indeed, all the cuts $(\{v\}, V - \{v\})$ as v ranges over all vertices in such an independent set are pairwise disjoint, and hence the function assigning to each of them the value 1 is a fractional packing of cuts.

The two theorems above suffice to determine $f(G)$ in many cases.

Corollary 5:

1) *For any Hamiltonian graph G with k vertices $f(G) = k/2$.*
2) *For any complete bipartite graph $G = K_{s,t}$ with $t \geq s \geq 1$, $f(G) = t$.*
3) *For any 2-edge connected graph G in which no two vertices of degree bigger than 2 are adjacent, $f(G)$ is exactly half the number of edges of G.*

A communication protocol is called *linear* if any bit it transmits is a linear combination of the input bits (and the bits received already). For simplicity we consider only linear combinations over Z_2, but the (simple) result that follows can be easily extended to all finite fields.

Proposition 6: For any connected graph G on k vertices, any linear protocol for solving the equality problem requires communication of at least $(k-1)n$ bits.

III. Proofs

A. Overview of the Main Proof

The main technical contribution of this paper is Theorem 2 that provides an upper bound for $f(G)$ which is tight in many cases. The proof is based on a construction, described in Lemma 9, which may be interesting in its own. In this lemma it is shown that for any fixed 2-connected graph H there is an m-vertex graph F consisting of $m^{2-o(1)}$ pairwise edge disjoint copies of H (called special copies) so that each edge of F is contained in a unique special copy. It is not difficult to see that such a construction does not exist without the 2-connectivity assumption. This extends the construction of Ruzsa and Szemerédi in [14], see also [1].

The proof of the lemma combines the classical theorem of Whitney on the structure of 2-connected graphs with a well known extension of the classical construction of Behrend of dense sets with no 3-term arithmetic progressions.

In the Equality protocol for the main case of a 2-connected graph H, we first direct the edges of H such that every vertex has a non-zero indegree. The parameter m is chosen so that $m^{2-o(1)}$ is roughly 2^n. The input of each player can now be mapped to one of the special copies of H in F. The protocol proceeds by sending only $\log m$ bits for each edge of H to the player residing in the vertex to which the edge is incoming. This player checks two properties. First, that the edge is indeed an edge in F, and second, that this edge appears in the copy of H identified by his input. If any of this local check fails, the protocol outputs NOT-equal. If all these local checks go through, this means that H is a special copy in F such that it has at least one edge in common with the special copy of H that each player sees. By the property of F, in this case all the copies of the players must be the same. Hence, the local checks ensure the global consistency. The details require some work, and are described in what follows.

1) Applications of Similar Techniques in Related Contexts: Tools from additive number theory, and in particular the Behrend construction of dense sets of integers with no 3-term arithmetic progressions and its variants, have found several applications in the study of communication problems. An early application appears in [7], where the authors describe an efficient multi-party communication protocol in the "number on the forehead model", for deciding if the sum of inputs of a set of players has a prescribed value. Another upper bound for a communication game obtained by a variant of the Behrend construction appears in [13]. This communication game arises in the study of time-space tradeoffs for oblivious branching programs for element distinctness. Yet another example based on some extensions by Ruzsa of the Behrend construction appear in [2], in the study of parent identifying codes. Finally, the graphs in [3] have been applied in the study of radio networks (besides their application to the problem considered

here for the complete graph). Their construction also relies on a certain Behrend-type construction.

Although the number theoretic constructions in all these works are similar to each other, the proof of our main result here is very different from those of all the other results above in the way it combines the graph theoretic tools, and in particular, Whitney's classical Theorem on the so called ear decomposition of graphs, with the arithmetic construction.

B. Preliminaries

We start with the simple proofs of Propositions 1 and 6

Proof of Proposition 1: We apply induction on the number of blocks s. For $s = 1$ there is nothing to prove. Assuming the result holds for $s-1$ we prove it for s, $s \geq 2$. Let $G, G_1, \ldots, G_s$ be as in the proposition, and assume, without loss of generality, that G_s is an end-block. Let v be the unique cut-vertex in G_s and let G' be the graph obtained from G by removing all vertices of G_s besides v. Thus G' has $s-1$ blocks $G_1, G_2, \ldots, G_{s-1}$.

To show that $f(G) \leq \sum_{i=1}^{s} f(G_i)$ observe that one can first apply the best protocol for solving the problem in G_s. If all vertices of G_s have the same bit string as v, we can now apply the best protocol for G' to complete the required task, thus establishing the upper bound.

To prove the lower bound consider the best protocol for solving the problem for G. By considering its behavior only on inputs of length n in which all vertices of G_s have equal inputs we conclude that the number of bits transmitted by this protocol along edges of G' is at least $f(n, G')$. Similarly, by considering the scenarios in which all vertices of G' have the same strings we conclude that the number of bits transmitted along edges of G_s is at least $f(n, G_s)$. This establishes the lower bound, completing the proof. □

Proof of Proposition 6: Consider a linear protocol for the problem, and suppose it transmits m bits. Each bit is a linear combination of the nk bits representing the inputs of the k vertices. For each such combination, define a linear equation equating it to zero. The set of all these m equations is a homogeneous system of m linear equations in kn variables. If $m < (k-1)n$ then the dimension of the solution space is bigger than n. However, the dimension of the space of all inputs in which all strings are equal is n, hence there is a solution, call it s, in which not all input strings are equal. Note that if each input string is the 0 vector, then all bits transmitted are 0, and the protocol must accept. Therefore, it must also accept the input s, as with this input all bits transmitted are also zero. But this means that the protocol errs on the input s, showing that a total communication of less than $(k-1)n$ is impossible in the linear case, as needed. □

C. The Upper Bound

In this section we prove Theorem 2. We need several lemmas, the first one is a known extension of the construction of Behrend in [4] of dense sets of integers with no 3-term arithmetic progressions.

A linear equation with integer coefficients

$$\sum a_i x_i = 0 \tag{1}$$

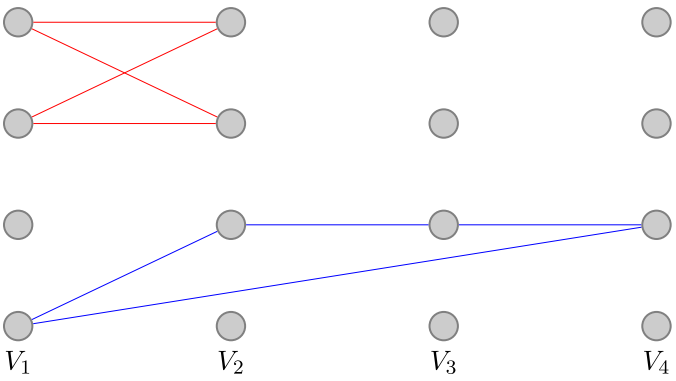

Fig. 1. The blue cycle is special copy of C_4, while the red one is not.

in the unknowns x_i is *homogeneous* if $\sum a_i = 0$. If $X \subseteq M = \{1, 2, \ldots, m\}$, we say that X *has no non-trivial solution to* (1), if whenever $x_i \in X$ and $\sum a_i x_i = 0$, it follows that all x_i are equal. Thus, for example, X has no nontrivial solution to the equation $x_1 - 2x_2 + x_3 = 0$ iff it contains no three-term arithmetic progression.

Lemma 7 (see, e.g., [1], Lemma 3.1): For every fixed integer $k \geq 2$ and every positive integer m, there exists a subset $X \subset M = \{1, 2, \ldots, m\}$ of size at least

$$|X| \geq \frac{m}{e^{10\sqrt{\log m \log k}}}$$

with no non-trivial solution to the equation

$$x_1 + x_2 + \cdots + x_k = k x_{k+1}. \tag{2}$$

Note that if there is no nontrivial solution for the above equation there is also no non-trivial solution for each of the equations $x_1 + x_2 + \cdots + x_r = r x_{r+1}$ for $r \leq k$, since a non-trivial solution of that together with $x_{r+1} = x_{r+2} = \cdots = x_k = x_{k+1}$ yields a non-trivial solution of (2).

We also need a basic result on 2-connected graphs, first proved by Whitney [16]. An *ear* of an undirected graph G is a path P where the two endpoints of the path may coincide, but where otherwise no repetition of edges or vertices is allowed. A *proper ear decomposition* of G is a partition of its set of edges into a sequence of ears, such that the first ear is a cycle, the two endpoints of any other ear are distinct and belong to earlier ears in the sequence and the internal vertices of each ear (if any) do not belong to any earlier ear. The following result was first proved by Whitney (it is also an easy consequence of Menger's Theorem.)

Lemma 8 (Whitney [16]): A graph G is 2-connected if and only if it has a proper ear decomposition.

Let H be a graph with k vertices $\{v_1, v_2, \ldots, v_k\}$. Let F be a k-partite graph with classes of vertices $V_1, V_2, \ldots, V_k$. A copy of H in F is called a *special copy* if for each $1 \leq i \leq k$ the vertex playing the role of v_i belongs to V_i, see Figure 1. Call F a *faithful host* for H if the set of its edges is the edge-disjoint union of special copies of H, and F contains no other special copy of H besides the $|E(F)|/|E(H)|$ copies defining its set of edges. See for example Figure 2, it contains three cycles: red, blue and green, but it is not a faithful host since the set of nodes V_{11}, V_{22}, V_{33} is another special copy of C_3. The following lemma is a crucial ingredient in the proof of

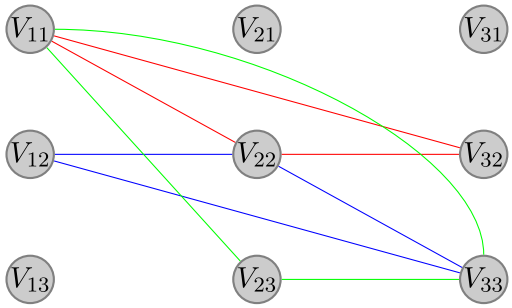

Fig. 2. The above is not a *faithful host* for C_3.

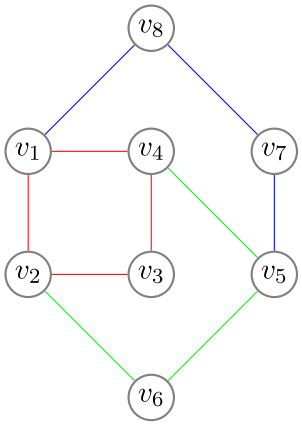

Fig. 3. An ear decomposition of H and its numbering. The first ear is red, the second is green and the third is blue.

Theorem 2. The special case when H is a cycle is proved in [1].

Lemma 9: Let H be a 2-connected graph with k vertices, and let m be a positive integer. Then there is a faithful host F for H with classes of vertices $V_1, \ldots, V_k$, each of size km, containing at least

$$\frac{m^2}{e^{10\sqrt{\log m \log k}}}$$

special copies of H.

Proof: By Lemma 8 there is a proper ear decomposition of H. Fix such a decomposition, and denote the ears in it by $P_1, P_2, \ldots, P_s$, in order, where P_1 is a cycle and each P_j for $j > 1$ is a path whose endpoints lie on vertices of earlier ears. Define a numbering of the vertices of H as follows. The vertices of the cycle P_1 are numbered $v_1, v_2, \ldots, v_t$, according to their order on the cycle. Assuming we have already numbered all vertices in the first p ears by $v_1, v_2, v_3 \ldots, v_\ell$, consider the next ear P_{p+1}. If it contains no internal vertices there is no new vertex in it that should be numbered. Otherwise, suppose the endpoints of this ear are v_i and v_j, where $i < j$, and suppose it has q internal vertices. Then this ear is a path of length $q+1$ from v_i to v_j and its vertices are numbered so that the vertices of the path are $v_j, v_{\ell+1}, v_{\ell+2}, \ldots, v_{\ell+q}, v_i$ in this order. See for example Figure 3

Let $X \subset \{1, 2, \ldots, m\}$ be as in Lemma 7. The host graph F is defined as follows. Its vertex classes are the classes $V_1, V_2, \ldots, V_k$, where each V_i is of size km (the first classes can be smaller, but this is not essential for our purpose here, hence we prefer the more symmetric description as above).

With slight abuse of notation denote the vertices of each set V_i by $\{1, 2, \ldots, km\}$ but recall that these sets are pairwise disjoint. The graph F contains $m|X|$ special copies of H defined as follows. For each integer y, $1 \le y \le m$ and each $x \in X$, there is a special copy of H in F, which we denote by $H_{x,y}$, in which $y + (i-1)x \in V_i$ is the vertex playing the role of v_i (for all $1 \le i \le k$). It is easy to see that all these special copies are pairwise edge disjoint. In fact, these copies satisfy a stronger property: no two of them share two vertices, since the values of $y + (i-1)x$ for two distinct indices i determine uniquely x and y.

It remains to prove that the only special copies of H in F are the copies $H_{x,y}$ used in its definition. Let H' be such a special copy. Then it contains an edge between V_1 and V_2 which connects $y \in V_1$ to $y + x \in V_2$, where $1 \le y \le m$ and $x \in X$. Let $u_1, u_2, \ldots, u_k$ be the vertices of H', where $u_i \in V_i$ for all i. Note that we denote the vertices of H' by u_i, whereas the vertices v_i denote those of H. The special copy H' is isomorphic to H, where the isomorphism maps u_i to v_i for each i. Our objective is to prove that $u_i = y + (i-1)x$ for all i. To do so we show, by induction on p, that this holds for each of the vertices $u_i \in V(H')$ where u_i plays the role of $v_i \in V(H)$ and v_i belongs to the union of the vertices in the first p ears in the ear decomposition of H. The first ear, P_1, is a cycle on the vertices $v_1, v_2, \ldots, v_t$. By the construction of F there are $x_1 = x, x_2, \ldots, x_t \in X$ so that $u_{i+1} - u_i = x_i$ for all $1 \le i \le t-1$ and $u_t - u_1 = (t-1)x_t$. Indeed the construction of F ensures that for every edge $u'u''$ connecting a vertex $u' \in V_i$ and a vertex $u'' \in V_j$, the difference $u'' - u'$ is $(j-i)x$ for some $x \in X$. Therefore $x_1 + x_2 + \ldots + x_{t-1} = (t-1)x_t$. Since $t \le k$, the property of the set X implies that $x_i = x_1 = x$ for all $1 \le i \le t$, establishing the required beginning of the induction. Assuming the induction claim holds for the vertices in the first p ears, consider the next ear P_{p+1}. If it contains no internal vertices there is nothing to prove, hence assume it contains q internal vertices. Let the ear P_{p+1} be $v_j, v_{\ell+1}, v_{\ell+2}, \ldots, v_{\ell+q}, v_i$, where $i < j$. By the induction hypothesis $u_i = y + (i-1)x$ and $u_j = y + (j-1)x$. By the construction of F there are $x_1, x_2, \ldots, x_{q+1} \in X$ so that $u_{\ell+1} - u_j = (\ell + 1 - j)x_1$, $u_{\ell+i+1} - u_{\ell+i} = x_{i+1}$ for $1 \le i \le q-1$, and $u_{\ell+q} - u_i = (\ell + q - i)x_{q+1}$. Since

$$(u_j - u_i) + (u_{\ell+1} - u_j) + (u_{\ell+2} - u_{\ell+1}) + \cdots + (u_{\ell+q} - u_{\ell+q-1}) = u_{\ell+q} - u_i$$

we conclude that

$$(j-i)x + (\ell + 1 - j)x_1 + x_2 + \cdots + x_q = (\ell + q - i)x_{q+1}.$$

As $\ell + q - i \le k$ the property of X implies that $x = x_1 = x_2 = \cdots = x_{q+1}$ completing the proof of the induction and implying the assertion of the lemma. □

Proof of Theorem 2: Let G' be a two edge-connected graph with $c_2(G)$ edges obtained from G as in the definition of $c_2(G)$. Thus, the set of vertices of G' is equal to that of G, and each of its edges is an edge of G, where some edges may be taken twice. In addition, G' is 2-edge connected and has the minimum possible number of edges among all graphs as above. By the minimality, the only edges of G' that appear

twice are the ones not contained in any cycle of length at least 3 of G', that is, these are bridges of (the underlying subgraph of) G'. We have to show that $f(G')$ is at most half the number of edges of G'.

By Proposition 1 it suffices to prove it for all blocks of G', where for blocks consisting of a single edge (taken twice) this is trivial, as obviously $f(K_2) = 1$. Every nontrivial block of G' is 2-connected, and it thus suffices to show that for any 2-connected graph $H = (V, E)$, $f(H) \leq 0.5|E|$.

Let k denote the number of vertices of H. For a given (large) integer n, let m be the smallest integer so that

$$\frac{m^2}{e^{10\sqrt{\log m \log k}}} \geq 2^n.$$

Thus

$$\log_2\ m = 0.5\ n + O(\sqrt{n \log k})$$

and

$$\lceil \log_2(km) \rceil = 0.5\ n + O(\sqrt{n \log k}) + O(\log k) = (0.5 + o(1))n.$$

Fix a numbering $v_1, v_2, \ldots, v_k$ of the vertices of H according to the proof of Lemma 9, and let F be a faithful host for H, with classes of vertices $V_1, V_2, \ldots, V_k$, containing at least 2^n special copies of H. Fix 2^n special copies. The input strings are now represented by special copies of H in F. Orient the edges of H so that the indegree of every vertex is positive. This is possible, since H is 2-connected. Indeed, using an ear decomposition of H we can orient the initial cycle cyclically and then orient each ear as a directed path.

For each special copy H' of H, let $u_i \in V(H')$ denote the vertex playing the role of $v_i \in V(H)$. The player P_i residing at the vertex v_i of H transmits the identity of the vertex u_i in the special copy of H representing his input to all players P_j so that there is an edge of H oriented from v_i to v_j. Note that this amounts to a total transmission of

$$\lceil \log_2(km) \rceil |E(H)| = (0.5 + o(1))|E|n$$

bits. In addition, each player observes if the identities of the vertices he received from his in-neighbors are indeed consistent with the ones in his copy, and reports about this to his out-neighbors (this amounts to another single bit per edge). If there is some inconsistency, this information reaches some player who reports that the inputs are not all equal. If everything is consistent, the players report that all inputs are equal.

It is clear that if all inputs are equal then the players report so. To complete the proof we show that if they report that the inputs are all equal, this is indeed the case. For every i let u_i be the identity of the vertex in V_i reported by i to his out-neighbors. Let the special copies of the players be $H_1, H_2, \ldots, H_k$, where H_i is the copy of the player P_i. If (v_j, v_i) is an edge of H oriented from v_j to v_i, and v_i who gets the identity of the vertex $u_j \in V_j$ from the player P_j, finds it consistent with his copy, then the edge $u_j u_i$ belongs to the special copy H_i of P_i. Therefore, if no player reports an inconsistency, then the subgraph of F on the vertices $u_1, u_2, \ldots, u_k$ is a special copy of H in F. However, since F is a faithful host for H this copy must be one of the original special copies of H in F, and as it contains an edge of each H_i (as the indegree of each vertex is positive) this special copy must be equal to H_i for all i, showing that indeed all these copies are equal. This completes the proof. □

D. The Lower Bound

As mentioned in Section 2 the assertion of Theorem 4 follows from the results of [8]. For completeness we include a short proof.

Proof of Theorem 4: Consider a deterministic communication protocol that solves the equality problem for inputs with n bits on $G = (V, E)$. For each edge $e \in E$, let $b(e)$ denote the number of bits transmitted during the protocol along e. We claim that for every cut $(S, \overline{S})$ in G $\sum_{e \in (S,\overline{S})} b(e) \geq n$. Indeed, otherwise there are two distinct strings of length n, x and y, so that the communication along the edges of the cut is identical when all inputs are x and when all inputs are y. But in that case it is easy to see that the protocol behaves identically when all inputs are x, when all inputs are y, and also when all vertices of S have input x and all those in $\overline{S}$ have input y (and vice versa). Thus the protocol cannot behave correctly, proving the claim.

By the claim it follows that a lower bound for $f(n, G)$ is the solution of the following linear program:

$$\text{Minimize } \sum_e b(e) \quad \text{subject to the constraints}$$
$$b(e) \geq 0 \quad \text{for all} \quad e \in E \quad \text{and}$$
$$\sum_{e \in (S,\overline{S})} b(e) \geq n \quad \text{for every cut} \quad (S, \overline{S}) \in \mathcal{S}, \qquad (3)$$

where $\mathcal{S}$ is the set of all cuts of G.

The dual of this program is:
Maximize $n \cdot \sum_{(S,\overline{S}) \in \mathcal{S}} g(S, \overline{S})$ subject to the constraints
$g(S, \overline{S}) \geq 0$ for all $(S, \overline{S}) \in \mathcal{S}$ and $\sum_{(S,\overline{S}), e \in (S,\overline{S})} g(S, \overline{S}) \leq 1$ for every edge $e \in E$.
This last maximum is exactly $n \cdot fc(G)$, completing the proof. □

Proof of Corollary 5:

1) By Theorem 4 and the paragraph following its statement $f(G) \geq k/2$ for any k-vertex graph G. By Theorem 2, for the cycle C_k on k vertices $f(C_k) \leq k/2$. The desired result follows since if G' is a spanning subgraph of G then clearly $f(G) \leq f(G')$.
2) For any tree T on k vertices $f(T) = k - 1$ (for example, by Proposition 1). This implies the result for $s = 1$. For larger s the lower bound follows from Theorem 4 by the fact that for any graph G with independence number $\alpha = \alpha(G)$, $fc(G) \geq \alpha$ as the α cuts $(v, V(G) - \{v\})$ for v in a maximum independent set are pairwise edge disjoint. The upper bound follows from Theorem 2 by considering a spanning subgraph of $K_{s,t}$ consisting of a cycle of length $2s$ together with two of the edges incident with any vertex of $K_{s,t}$ uncovered by the cycle.
3) The upper bound follows from Theorem 2. To prove the lower bound note that G is the edge disjoint union of induced paths, each of length at least 2. For each such path $v_1, v_2, \ldots, v_s$ in which all internal vertices are of

degree 2 in G, consider the cuts $(v_i, V - \{v_i\})$ for all $1 < i < s$, and the cut

$$(\{v_2, v_3 \ldots, v_{s-1}\}, V - \{v_2, v_3 \ldots, v_{s-1}\})$$

(if $s = 3$ we take the same cut twice). This is a collection of $|E(G)|$ cuts covering each edge exactly twice, hence $fc(G) \geq |E(G)|/2$, as shown by giving each of these cuts weight $1/2$. This completes the proof. □

IV. Open Problems

- Is $f(G) = 0.5\ c_2(G)$ for any connected graph G ?
- If not, is $f(G) = fc(G)$ for any connected graph G ? As pointed out by L. Esperet, it is known that the ratio between $0.5\ c_2(G)$ and $fc(G)$ is at most $4/3$ for any connected graph G, see [15], Theorem 11.
- Is it true that for a graph G on k vertices $f(G) = k/2$ if and only if G is Hamiltonian? (Note that if this is the case, then the computational problem of computing $f(G)$ for a given input graph G is NP-hard.)
- It is not difficult to show that for any d-regular graph G on k vertices which is also d-edge connected, $fc(G) = k/2$. Indeed, as mentioned in the paragraph following the statement of Theorem 4, $fc(G) \geq k/2$ for any k vertex graph. To prove the upper bound note that for any d-regular d edge-connected graph $G = (V, E)$, the function $b(e) = n/d$ for every edge $e \in E$ is a solution of the linear program (3).
  Thus, for any such G the lower bound for $f(G)$ provided by Theorem 4 is $k/2$ whereas if it is not Hamiltonian the upper bound provided by Theorem 2 is strictly larger.
  A specific interesting example is the Petersen graph P which is 3-regular, 3-connected and non-Hamiltonian. Indeed $c_2(P) = 11$ and $fc(G) = 5$, implying that

$$5 \leq f(P) \leq 5.5$$

  What is $f(P)$?

Acknowledgments

We thank Louis Esperet for telling us about [15]. The second author thanks Benny Applebaum and Pavel Raykov for meaningful discussions.

References

[1] N. Alon, "Testing subgraphs in large graphs," *Random Struct. Algorithms*, vol. 21, nos. 3–4, pp. 359–370, 2002.
[2] N. Alon, E. Fischer, and M. Szegedy, "Parent-identifying codes," *J. Combinat. Theory, Ser. A*, vol. 95, no. 2, pp. 349–359, 2001.
[3] N. Alon, A. Moitra, and B. Sudakov, "Nearly complete graphs decomposable into large induced matchings and their applications," in *Proc. 44th ACM STOC*, 2012, pp. 1079–1089.
[4] F. A. Behrend, "On sets of integers which contain no three terms in arithmetical progression," *Proc. Nat. Acad. Sci. USA*, vol. 32, no. 12, pp. 331–332, 1946.
[5] J. A. Bondy and U. S. R. Murty, *Graph Theory, Graduate Texts in Mathematics*, vol. 244. New York, NY, USA: Springer, 2008, p. 651.
[6] J. Brody, *Private Communication*. 2012.
[7] A. K. Chandra, M. L. Furst, and R. J. Lipton, "Multi-party protocols," in *Proc. 15th Annu. ACM STOC*, 1983, pp. 94–99.
[8] A. Chattopadhyay and A. Rudra, "The range of topological effects on communication," in *Proc. ICALP*, 2015, pp. 540–551.
[9] A. Chattopadhyay, J. Radhakrishnan, and A. Rudra, "Topology matters in communication," in *Proc. FOCS*, Oct. 2014, 631–640.
[10] F. Ellen, R. Oshman, T. Pitassi, and V. Vaikuntanathan, "Brief announcement: Private channel models in multi-party communication complexity," in *Proc. DISC*, 2013, pp. 575–576.
[11] M. Fekete, "Über die verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten," *Math. Zeitschrift*, vol. 17, no. 1, pp. 228–249, 1923.
[12] G. Liang and N. Vaidya, "Multiparty equality function computation in networks with point-to-point links," in *Proc. SIROCCO*, 2011, pp. 258–269.
[13] P. Pudlák and J. Sgall, "An upper bound for a communication game related to time-space tradeoffs," in *The Mathematics of Paul Erdös, I* (Algorithms and Combinatorics), vol. 13, Berlin, Germany: Springer, 1997, pp. 393–399.
[14] I. Z. Ruzsa and E. Szemerédi, "Triple systems with no six points carrying three triangles," in *Proc. 5th Hungarian Colloq.*, vol. 2. Keszthely, Hungary, 1976, pp. 939–945.
[15] A. Sebö and J. Vygen, "Shorter tours by nicer ears: 7/5-Approximation for the graph-TSP, 3/2 for the path version, and 4/3 for two-edge-connected subgraphs," *Combinatorica*, vol. 34, pp. 1–34, Jul. 2014.
[16] H. Whitney, "Non-separable and planar graphs," *Trans. Amer. Math. Soc.*, vol. 34, no. 2, pp. 339–362, 1932.

Noga Alon is a Baumritter Professor of Mathematics and Computer Science at Tel Aviv University, Israel. He received his Ph. D. in Mathematics at the Hebrew University of Jerusalem in 1983 and had visiting positions in various research institutes including MIT, The Institute for Advanced Study in Princeton, IBM Almaden Research Center, Bell Laboratories, Bellcore and Microsoft Research. He serves on the editorial boards of more than a dozen international technical journals and has given invited lectures in many conferences, including plenary addresses in the 1996 European Congress of Mathematics and in the 2002 International Congress of Mathematicians. He published one book and more than five hundred research papers. His research interests are mainly in Combinatorics, Graph Theory and their applications in Theoretical Computer Science. His main contributions include the study of expander graphs and their applications, the investigation of derandomization techniques, the foundation of streaming algorithms, the development and applications of algebraic and probabilistic methods in Discrete Mathematics and the study of problems in Information Theory, Combinatorial Geometry and Combinatorial Number Theory. He is an ACM Fellow and an AMS Fellow, a member of the Israel Academy of Sciences and Humanities and of the Academia Europaea, and received the Erdős Prize, the Feher Prize, the Polya Prize, the Bruno Memorial Award, the Landau Prize, the Gödel Prize, the Israel Prize, the EMET Prize, the Dijkstra Prize, and Honorary Doctorate from ETH Zürich and from the University of Waterloo.

Klim Efremenko received his Ph.D. from Tel Aviv University in 2012. He made a post-doc at the Institute for Advanced Studies University of Chicago and Berkeley. Currently, he is in Tel-Aviv Universty and starting a position in Ben-Gurion University. His main research interests are error correction codes and interactive communication.

Benny Sudakov is a Professor of Mathematics at ETH, Zurich. He received his PhD from Tel Aviv University in 1999 and had appointments in Princeton University, the Institute for Advanced Studies and in UCLA. He serves on the editorial boards of more than a dozen international technical journals and has given invited lectures in many conferences, including invited address at the 2010 International Congress of Mathematicians. He published more than two hundred research papers. His main research interests are Algebraic and Probabilistic Methods in Combinatorics, Extremal Graph and Hypergraph Theory, Ramsey Theory, Random Structures and Application of Combinatorics to Theoretical Computer Science. He is an AMS Felow and is recipeient of a Sloan Fellowship, NSF CAREER Award and Humboldt Research Award.