On Computing Ad-hoc Selective Families

Andrea E.F. Clementi[1], Pilu Crescenzi[2], Angelo Monti[3], Paolo Penna[1], and Riccardo Silvestri[3]

[1] Dipartimento di Matematica, Università di Roma "Tor Vergata"
{clementi,penna}@mat.uniroma2.it
[2] Dipartimento di Sistemi e Informatica, Università di Firenze
piluc@dsi.unifi.it
[3] Dipartimento di Scienze dell'Informazione, Università di Roma "La Sapienza",
{monti,silver}@dsi.uniroma1.it

Abstract. We study the problem of computing ad-hoc selective families: Given a collection $\mathcal{F}$ of subsets of $[n] = \{1, 2, \ldots, n\}$, a *selective family for* $\mathcal{F}$ is a collection $\mathcal{S}$ of subsets of $[n]$ such that for any $F \in \mathcal{F}$ there exists $S \in \mathcal{S}$ such that $|F \cap S| = 1$. We first provide a polynomial-time algorithm that, for any instance $\mathcal{F}$, returns a selective family of size $O((1 + \log(\Delta_{max}/\Delta_{min})) \cdot \log |\mathcal{F}|)$ where Δ_{max} and Δ_{min} denote the maximal and the minimal size of a subset in $\mathcal{F}$, respectively. This result is applied to the problem of broadcasting in radio networks with known topology. We indeed develop a broadcasting protocol which completes any broadcast operation within $O(D \log \Delta \log \frac{n}{D})$ time-slots, where n, D and Δ denote the number of nodes, the maximal eccentricity, and the maximal in-degree of the network, respectively. Finally, we consider the combinatorial optimization problem of computing broadcasting protocols with minimal completion time and we prove some hardness results regarding the approximability of this problem.

1 Introduction

Selective Families. The notion of *selective family* has been introduced in [6]. Given a positive integer n, a family $\mathcal{S}$ of subsets of $[n] = \{1, 2, \ldots, n\}$ is said to be (n, h)-selective if and only if, for any subset $F \subseteq [n]$ with $|F| \le h$, there is a set $S \in \mathcal{S}$ such that $|F \cap S| = 1$. This notion is an essential tool exploited in [6,7,8,9] to develop distributed broadcasting algorithms in radio networks with unknown topology. In particular, [6] provide a polynomial-time algorithm that, for any integer ℓ, computes a $(2^\ell, 2^{\lceil \ell/6 \rceil})$-selective family of size $O(2^{5\ell/6})$: Notice that the size of the selective family is a key parameter since, as we will see later, it determines the completion-time of the corresponding broadcasting protocol. Better constructions of selective families have been introduced in [7]. The best known (non-constructive) upper bound has been proved in [9] where it is shown that there exists an (n, h)-selective family of size $O(h \log n)$. This upper bound is almost tight since, in the same paper, it is shown that, any (n, h)-selective family has size $\Omega(h \log \frac{n}{h})$.

M. Goemans et al. (Eds.): APPROX-RANDOM 2001, LNCS 2129, pp. 211–222, 2001.

Radio Networks and Broadcasting. A *radio network* is a set of radio stations that are able to communicate by transmitting and receiving radio signals. A transmission range is assigned to each station s and any other station t within this range can directly (i.e. by one *hop*) receive messages from s. Communication between two stations that are not within their respective ranges can be achieved by *multi-hop* transmissions. In this paper, we will consider the case in which radio communication is structured into synchronous *time-slots*, a paradigm commonly adopted in the practical design of protocols [3,10,16]. A radio network can be modeled as a directed graph $G(V, E)$ where an edge (u, v) exists if and only if u can send a message to v in one hop. The nodes of a radio network are processing units, each of them able to perform local computations. It is also assumed that every node is able to perform *all* its local computations required for deciding the next send/receive operation during the current time-slot. In every time-slot, each node can be *active* or *non-active*. When it is active, it can decide to be either *transmitter* or *receiver* : in the former case the node transmits a message along all of its outgoing edges while, in the latter case, it tries to recover messages from all its incoming edges. In particular, the node can recover a message from one of its incoming edges if and only if this edge is the only one bringing in a message. When a node is non-active, it does not perform any kind of operation.

One of the fundamental tasks in network communication is the *broadcast* operation. It consists in transmitting a message from one source node to all the other nodes of the network. A broadcasting protocol is said to have *completed broadcasting* when all nodes, reachable from the source, have received the source message (notice that when this happens, the nodes not necessarily stop to run the protocol since they might not know that the operation is completed). We also say that a broadcasting protocol *terminates* in time t if, after the time-slot t, all the nodes are in the non-active state (i.e. when all nodes stop to run the protocol). According to the network model described above, a broadcasting protocol operates in time-slots synchronized by a global clock: At every time-slot, each active node decides to either transmit or receive, or turn into the non-active state.

Selective Families and Broadcasting in Unknown Topology. Given a radio network with n nodes, maximal in-degree $\varDelta$, and unknown topology (in the sense that nodes know nothing about the network but their own label), the existence of a $(n, \varDelta)$-selective family of size m implies the existence of a distributed broadcasting protocol in the network, whose completion time is $O(nm)$. The protocol operates in n phases of m time-slots each. During each phase, at time-slot j the nodes (whose labels are) in the j-th set of the selective family which are *informed* (that is, which have already received the source message) transmit the message to their out-neighbors. Even though no node knows the topology of the network, the definition of a selective family implies that, at the end of each phase, at least one new node has received the source message (hence, n phases are sufficient): This is due to the fact that *for any* subset F of $[n]$, with $|F| \leq \varDelta$, there exists at least one subset of the selective family whose intersection with F contains exactly one element. In particular, for any i with $1 \leq i < n$, let R_i be the set of nodes that have received the source message after the first i phases and assume that there are still nodes to be informed. Then, there exists a non-informed node x_i which is an out-neighbor of at least one node in R_i: Let F_i be the set of (labels of the) nodes $r \in R_i$ such that x_i is an out-neighbor of r and let S_j be a subset in the selective family such that $|F_i \cap S_j| = 1$ (notice that S_j *must exist independently of the topology of the network*). Hence, x_i will certainly receive the source message at the latest during the j-th time-slot of the $(i+1)$-th phase since during this time-slot *exactly one* of the informed in-neighbors of x_i transmits the source message to x_i[1].

Ad-hoc Selective Families and Broadcasting in Known Topology. The connection between selective families and broadcasting in radio networks with unknown topology justifies the fact that *each* subset of the domain $[n]$ has to be selected by at least one subset in the family: Indeed, since the topology of the network is not known, in order to be sure that F_i is selected by at least one subset in the family, the family itself has to select every possible subset of the nodes. In this paper, we assume that *each node knows the network topology and the source node* which is a well-studied assumption concerning the communication process in radio networks [4,5,11,1,15]. It is then easy to see that the family to be computed has to select only those sets of nodes which are at distance l from the source node and which are the in-neighbors of a node at distance $l+1$, for any distance l. This observation leads us to the following definition of *ad-hoc selective family*: Given a collection $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ of subsets of $[n]$, a family $\mathcal{S} = \{S_1, S_2, \ldots, S_k\}$ of subsets of $[n]$ is said to be *selective for* $\mathcal{F}$ if, for any F_i, there exists S_j such that $|F_i \cap S_j| = 1$. In Fig.1 it is shown how a radio network determines the sets F_i to be selected: For example, for $l = 2$, there are three sets to be selected, that is, F_2 which is the in-neighborhood of node 4, F_3 which is the in-neighborhood of node 5, and F_4 which is the in-neighborhood of node 6.

Our (and Previous) Results. The main result of this paper is a polynomial-time algorithm that, for any collection $\mathcal{F}$, returns a selective family for $\mathcal{F}$ of size $O((1+\log(\Delta_{max}/\Delta_{min}))\cdot\log|\mathcal{F}|)$ where Δ_{max} and Δ_{min} denote the maximal and the minimal size of a subset in $\mathcal{F}$, respectively. The proof of our result is based on the probabilistic method [2]: We first perform a probabilistic analysis of randomly computing a selective family according to a specific probability distribution and we, subsequently, apply the method of the conditional probabilities in order to de-randomize the probabilistic construction.

The above efficient construction of ad-hoc selective families is then used to derive a broadcasting protocol having completion time $O(D \log \Delta \log(n/D))$, where D denotes the maximal eccentricity of the network (that is, the longest distance from a node to any other node). The protocol is efficiently constructible

[1] Actually, this is not always true since the set of informed nodes might have been changed before the j-th time-slot of the $(i+1)$-th phase: However, this would imply that at least one other new node has been informed during the $(i+1)$-th phase.

PSfrag replacements

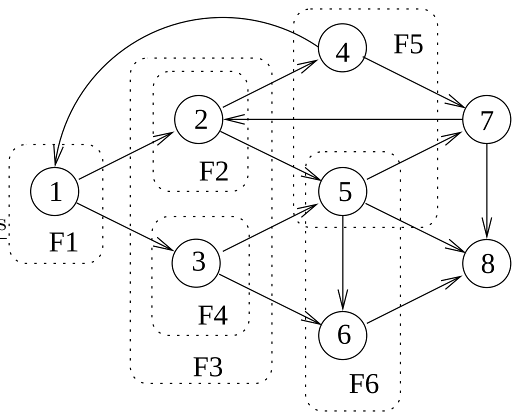

Fig. 1. The collection of sets to be selected corresponding

(i.e. it can be constructed in deterministic polynomial time in the size of the network) and easy-to-implement. Its completion time is better than the $O(D \log^2 n)$ bound obtained in [5] whenever $\Delta = O(1)$ and $D = \Theta(n)$, or $\Delta = o(n^\alpha)$, or $n/D = o(n^\alpha)$ (for every positive constant $\alpha > 0$). Furthermore, our bound implies that the $\Omega(\log^2 n)$ lower bound, shown in [1] in the case in which D is a constant value greater than or equal to 2, only holds when $\Delta = \Omega(n^\alpha)$, for some positive constant $\alpha > 0$. In [11], an $O(D + \log^5 n)$ upper bound is proved. The efficient construction of the protocol with such a completion time relies on a de-randomization of the well-known distributed randomized protocol in [3]: However, it is not clear whether this de-randomization can be done efficiently.

The *communication complexity* of our protocol, that is, the maximum number of messages exchanged during its execution, turns out to be $O(n \log \Delta \log(n/D))$, since each node sends at most $O(\log \Delta \log(n/D))$ messages. Moreover, when the source is unknown our protocol technique works in $O(D \log \Delta \log n)$ time.

The probabilistic argument used in order to efficiently construct small ad-hoc selective families will then be applied in order to develop a polynomial-time approximation algorithm for the Max Pos One-In-k-Sat problem: Given a set of clauses with each clause containing exactly k literals, all positive, find a truth-assignment to the Boolean variables that 1-in-k satisfies the maximal number of clauses, where a clause is *1-in-k satisfied* if exactly one literal in the clause is assigned the value true. According to [17,13], this problem is NP-hard. Furthermore, from the approximation algorithm for the more general maximum constraint satisfaction problem (MAX CSP), it is known that the problem is $\frac{2^k}{k}$-approximable [14]. The performance ratio of our algorithm is bounded by $\frac{4(k-1)}{k} \leq 4$ (notice that the performance ratio is bounded by a constant, that is, 4 which does not depend on the value of k).

Finally, we investigate the combinatorial optimization problem of computing a broadcasting protocol with minimal completion time, called Min Broadcast.

This problem is NP-hard [4]: However, the reduction, which starts from the exact-3 cover problem, yields radio networks with $D = 2$. We instead introduce a new reduction starting from the problem of computing ad-hoc selective families of minimal size, called Min Selective Family. This reduction allows us to show that, for any fixed $D \geq 2$, if Min D-Broadcast (i.e., the problem restricted to radio networks of maximal eccentricity D) is r-approximable, then Min Selective Family is $\frac{rD-1}{D-1}$-approximable. Since the latter is not r-approximable, for any $r < 2$, we obtain that Min Broadcast cannot be approximated within a factor less than $2 - 1/D$ (unless P = NP).

2 Efficient Construction of Ad-hoc Selective Families

This section provides an efficient method to construct selective families of small size. More precisely, we prove the following

Theorem 1 *There exists an algorithm that, given a collection $\mathcal{F}$ of subsets of $[n]$, each of size in the range $[\Delta_{min}, \Delta_{max}]$, computes a selective family $\mathcal{S}$ for $\mathcal{F}$ of size $O((1+\log(\Delta_{max}/\Delta_{min}))\cdot\log|\mathcal{F}|)$. The time complexity of the algorithm is $O(n^2|\mathcal{F}|\log|\mathcal{F}|\cdot(1+\log(\Delta_{max}/\Delta_{min})))$.*

Proof. The proof consists of two main steps. We first show the *existence* of the selective family $\mathcal{S}$ by using a *probabilistic* construction. Then, an efficient algorithm that de-randomizes this construction is presented.

Probabilistic Construction. Without loss of generality, we can assume that $\Delta_{min} \geq 2$. For each $i \in \{\lceil\log\Delta_{min}\rceil, \ldots, \lceil\log\Delta_{max}\rceil\}$, consider a family $\mathcal{S}_i$ of l sets (the value of l is specified later) in which each set is constructed by randomly picking every element of $[n]$ independently, with probability $\frac{1}{2^i}$.
Fix a set $F \in \mathcal{F}$ and consider a set $S \in \mathcal{S}_i$, where i is the integer such that $\frac{1}{2} \leq \frac{|F|}{2^i} < 1$; then it holds that

$$\Pr[|F \cap S| = 1] = \frac{|F|}{2^i}\left(1 - \frac{1}{2^i}\right)^{|F|-1} > \frac{|F|}{2^i}\left(1 - \frac{1}{2^i}\right)^{2^i} \geq \frac{|F|}{4\cdot 2^i} \geq \frac{1}{8} \qquad (1)$$

where the second inequality is due to the fact that $\left(1 - \frac{1}{t}\right)^t \geq \frac{1}{4}$ for $t \geq 2$. We then define the family $\mathcal{S}$ as the union of the families $\mathcal{S}_i$, for each $i \in \{\lceil\log\Delta_{min}\rceil, \ldots, \lceil\log\Delta_{max}\rceil\}$. Clearly, $\mathcal{S}$ has size $O((1+\log(\Delta_{max}/\Delta_{min}))\cdot l)$. The probability that $\mathcal{S}$ does not select F is upper bounded by the probability that $\mathcal{S}_i$ does not select F. The sets in $\mathcal{S}_i$ have been constructed independently, so, from Eq. 1, this probability is at most $\left(1 - \frac{1}{8}\right)^l \leq e^{-\frac{l}{8}}$. Finally, we have that

$$\Pr[\mathcal{S} \text{ is not selective for } \mathcal{F}] \leq \sum_{F\in\mathcal{F}} \Pr[\mathcal{S} \text{ doesn't select } F] \leq \sum_{F\in\mathcal{F}} e^{-\frac{l}{8}} = |\mathcal{F}|e^{-\frac{l}{8}}.$$

The last value is less than 1 for $l > 8\log|\mathcal{F}|$. Hence, such an $\mathcal{S}$ exists.
De-randomization. The de-randomization is obtained by applying the "greedy" criterium yielded by the method of the *conditional probabilities* [12].

Let us represent any subset $S \subseteq [n]$ as a binary sequence $\langle s_1, \ldots, s_n \rangle$ where, for any $i \in [n]$, $s_i = 1$ if and only if $i \in S$. Let $i \in [n]$, let $F \in \mathcal{F}$ of size Δ, and let $\langle s_1, \ldots, s_{i-1} \rangle$ be any sequence of $i-1$ bits (i.e., any subset of the first $i-1$ elements of $[n]$). Then, define the (conditional) probabilities

$$Y_i(F) = \Pr\left[|F \cap \langle s_1, \ldots, s_{i-1}, 1, x_{i+1}, \ldots, x_n \rangle| = 1\right]$$
$$N_i(F) = \Pr\left[|F \cap \langle s_1, \ldots, s_{i-1}, 0, x_{i+1}, \ldots, x_n \rangle| = 1\right]$$

where, for any $k = i+1, \ldots, n$, x_k is a bit chosen independently at random with

$$\Pr[x_k = 1] = 1/\Delta.$$

The algorithm relies on the following

Lemma 2 *It is possible to compute both $Y_i(F)$ and $N_i(F)$ in $O(n)$ time.*

Proof. Let us define $S_i = \langle s_1, \ldots, s_{i-1}, 0, \ldots, 0 \rangle$, and $I_i = \{i, i+1, \ldots, n\}$. Define also $\delta_i(F) = |F \cap I_i|$. If $\delta_i(F) = 0$, then it is easy to verify that

$$Y_i(F) = N_i(F) = \begin{cases} 1 & \text{if } |F \cap S_i| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

If, instead, $\delta_i(F) > 0$ then two cases may arise

- **Case** $i \in F$. Then, it holds that

$$Y_i(F) = \begin{cases} 0 & \text{if } |F \cap S_i| \geq 1, \\ \left(1 - \frac{1}{\Delta}\right)^{\delta_i(F)-1} & \text{otherwise,} \end{cases}$$

$$N_i(F) = \begin{cases} 0 & \text{if } |F \cap S_i| \geq 2 \\ & \text{or } \delta_i(F) = 1 \ \wedge \ |F \cap S_i| = 0, \\ \left(1 - \frac{1}{\Delta}\right)^{\delta_i(F)-1} & \text{if } |F \cap S_i| = 1, \\ \frac{\delta_i(F)-1}{\Delta}\left(1 - \frac{1}{\Delta}\right)^{\delta_i(F)-2} & \text{otherwise.} \end{cases}$$

- **Case** $i \notin F$. Then, it holds that

$$Y_i(F) = N_i(F) = \begin{cases} 0 & \text{if } |F \cap S_i| \geq 2, \\ \left(1 - \frac{1}{\Delta}\right)^{\delta_i(F)} & \text{if } |F \cap S_i| = 1, \\ \frac{\delta_i(F)}{\Delta}\left(1 - \frac{1}{\Delta}\right)^{\delta_i(F)-1} & \text{otherwise.} \end{cases}$$

The proof is completed by observing that all the computations required by the above formulas can be easily done in $O(n)$ time. □

Figure 2 shows the algorithm $\texttt{greedy}_{\texttt{MSF}(\Delta)}$ that finds the desired selective family when all subsets in $\mathcal{F}$ have the same size. As for the general case, the algorithm must be combined with the technique in the probabilistic construction that splits $\mathcal{F}$ into a logarithmic number of families, each containing subsets having "almost" the same size. A formal description of this generalization will be given in the full version of the paper. However, we observe here that the time complexity of the general algorithm is $O(1 + \log(\Delta_{max}/\Delta_{min}))$ times the time complexity of $\texttt{greedy}_{\texttt{MSF}(\Delta)}$.

Input $\mathcal{F} = \{F_1, \ldots, F_m\}$
$\mathcal{F}' := \mathcal{F}$
$j := 0$
While $\mathcal{F}' \neq \emptyset$ **Do** /* Construct the jth selector S_j */
  For each $i = 1, \ldots, n$ **Do**
    For each $F \in \mathcal{F}'$ **Do** compute $Y_i(F)$ and $N_i(F)$ (using Lemma 2)
    $Y_i := \sum_{F \in \mathcal{F}'} Y_i(F)$
    $N_i := \sum_{F \in \mathcal{F}'} N_i(F)$
    If $Y_i \geq N_i$ **Then** $s_i := 1$ **Else** $s_i := 0$.
  End (For)
  $j := j + 1$
  $S_j := \langle s_1, \ldots, s_n \rangle$
  $\mathcal{F}' := \mathcal{F}' - \{F \in \mathcal{F}' : |F \cap S_j| = 1\}$
End (While)
Return $\mathcal{S} = \{S_1, \ldots, S_j\}$.

Fig. 2. Algorithm $\texttt{greedy}_{\texttt{MSF}(\Delta)}$.

Lemma 3 *Let $\mathcal{F}$ be a family of subsets of $[n]$, each of size Δ. Then, Algorithm $\texttt{greedy}_{\texttt{MSF}(\Delta)}$ (with input $\mathcal{F}$) computes a selective family $\mathcal{S}$ for $\mathcal{F}$ of size $O((1 + \log(\Delta_{max}/\Delta_{min})) \log |\mathcal{F}|)$ in time $O(n^2 |\mathcal{F}| \log |\mathcal{F}|)$.*

Proof. We first prove that, at each iteration of the **While** loop, the computed subset S_j selects at least 1/8 of the remaining subsets of $\mathcal{F}$, i.e., $\mathcal{F}'$.
Let B be a subset of $[n]$ randomly chosen according to the following probability function: For each $i \in [n]$, $i \in B$ with probability $1/\Delta$. Let $E(B)$ denote the expected number of subsets F in $\mathcal{F}'$ such that $|F \cap B| = 1$.
For any $i \in [n]$ and for any bit sequence $b_1, \ldots, b_i$, let $E(B|b_1, \ldots, b_i)$ be the expected number of subsets F in $\mathcal{F}'$ such that $|F \cap B| = 1$, where $B = \langle b_1, \ldots, b_n \rangle$ is the random completion of the sequence $b_1, \ldots, b_i$ such that, for any $i + 1 \leq l \leq n$, $b_l = 1$ with probability $1/\Delta$.

Let $s_1, \ldots, s_n$ be the choices made by the **For** loop.

Claim 1 *For any $i = 1, \ldots, n$, $E(B|s_1, \ldots, s_i) \geq E(B)$.*

Proof. The proof is by induction on i. For $i = 1$, by definition, we have that

$$E(B) = \frac{1}{\Delta} E(B|1) + \left(1 - \frac{1}{\Delta}\right) E(B|0)$$

So, $E(B) \leq \max\{E(B|1), E(B|0)\} = \max\{Y_1, N_1\}$, and s_1 is chosen so that $E(B|s_1) = \max\{Y_1, N_1\}$. We now assume that the claim is true for $i - 1$. Then, s_i is chosen so that

$$\begin{aligned} E(B|s_1, \ldots, s_i) &= \max\{Y_i, N_i\} \\ &= \max\{E(B|s_1, \ldots, s_{i-1}, 1), E(B|s_1, \ldots, s_{i-1}, 0)\}. \end{aligned}$$

It also holds that

$$E(B|s_1,\ldots,s_{i-1}) = \frac{1}{\Delta}E(B|s_1,\ldots,s_{i-1},1) + \left(1-\frac{1}{\Delta}\right)E(B|s_1,\ldots,s_{i-1},0)$$
$$\leq \max\{E(B|s_1,\ldots,s_{i-1},1), E(B|s_1,\ldots,s_{i-1},0)\}.$$

By combining the above inequalities with the inductive hypothesis, we get

$$E(B|s_1,\ldots,s_i) \geq E(B|s_1,\ldots,s_{i-1}) \geq E(B).$$

□

Let us observe that, $E(B|s_1,\ldots,s_n)$ is equal to the number of subsets in $\mathcal{F}'$ that are selected by $S = \langle s_1,\ldots,s_n\rangle$. From Claim 1, this number is at least $E(B)$. Moreover, from Eq. 1, it holds that $E(B) \geq |\mathcal{F}'|/8$. Finally, from Lemma 2, it follows that the time complexity of $\texttt{greedy}_{\texttt{MSF}(\Delta)}$ is $O(n^2|\mathcal{F}|\log|\mathcal{F}|)$. □

3 Two Applications of Theorem 1

The Broadcast Protocol. For any possible source node $s \in V$, let $L_i(s)$ be the set of nodes whose distance from s is i. For each node in $L_{i+1}(s)$ let us consider the set of its in-neighbors belonging to $L_i(s)$; let $\mathcal{F}_i(s)$ be the family of all such sets. Then, let $\mathcal{S}_i$ be an arbitrarily ordered selective family for $\mathcal{F}_i(s)$.

Description of Protocol BROAD**.** The protocol consists of D phases. The goal of phase i is to inform nodes at distance i from the source.
– In the first phase the source sends its message.
– The i-th phase, with $i \geq 2$, consists of $|\mathcal{S}_{i-1}|$ time-slots. At time-slot j of the i-th phase a node v sends the source message if and only if the following two conditions are satisfied:

- v belongs to the j-th set of $\mathcal{S}_{i-1}$;
- v has been informed for the first time during phase $i-1$.

All the remaining nodes work as receivers.

Theorem 4 *Protocol* BROAD *completes (and terminates) a broadcast operation on an n-node graph of maximum eccentricity D and maximum in-degree Δ within $O(D\log\Delta\log\frac{n}{D})$ time-slots. Moreover, the cost of the protocol is $O(n\log\Delta\log\frac{n}{D})$.*

Proof. To show the correctness (and the performances) of the protocol we prove the following

Claim 2 *all the nodes at distance i from the source s are informed, for the first time, during phase i.*

Sketch of the proof. The proof is by induction on the distance i. For $i = 1$ the claim is obvious. We thus assume that all nodes at distance i have received the source message, for the first time, during phase $i - 1$. Let consider a node v at distance $i + 1$ and let F_v be the set of all its in-neighbors at distance i from the source. Since F_v belongs to $\mathcal{F}_i(s)$ and $\mathcal{S}_i$ is selective for $\mathcal{F}_i(s)$, there will be a time-slot in phase $i + 1$ in which only one of the nodes in F_v transmits the source message so that v will correctly receive it. Notice that, by the inductive hypothesis, any in-neighbor of v that is not in F_v has not been informed in phase i, so it does not transmit during phase $i + 1$. □

Since the graph has maximum in-degree Δ, the size of any subset in $\mathcal{F}_i(s)$ is at most Δ. Hence, from Theorem 1, we have that $|\mathcal{S}_i| \leq c \log \Delta \log |\mathcal{F}_i(s)|$, for some constant $c > 0$. The total number of time-slots required by the protocol is thus

$$1 + \sum_{i=1}^{D-1} c \log \Delta \log |\mathcal{F}_i(s)| = 1 + c \log \Delta \log \prod_{i=1}^{D-1} |\mathcal{F}_i(s)| \leq 1 + c \log \Delta \log \prod_{i=1}^{D-1} \frac{n}{D}$$

where the last inequality is due to the facts that $\sum_{i=1}^{D-1} |\mathcal{F}_i(s)| \leq n$ and that $\prod_{i=1}^{D-1} |\mathcal{F}_i(s)|$ is maximized when all the $|\mathcal{F}_i(s)|$ are equal. It thus follows that BROAD has $O(D \log \Delta \log \frac{n}{D})$ completion time.
As for the cost of the protocol, it suffices to observe that once a node has acted as transmitter during a phase, after that phase it can turn into the inactive state forever. □

Remark. If we require a protocol that works for any source, we need to select a bigger set of families, i.e., the families $\mathcal{F}_i = \cup_{s \in V} \mathcal{F}_i(s)$, $i = 1 \ldots n - 1$. By applying the same arguments of the above proof, we can easily obtain a broadcast protocol having $O(D \log \Delta \log n)$ completion time.

Approximation of the MAX POS ONE-IN-k-SAT *Problem.* An (even) simplified version of the algorithm $\texttt{greedy}_{\texttt{MSF}(\Delta)}$ can be successfully used to obtain a constant factor approximation for MAX POS ONE-IN-k-SAT.

Corollary 5 *There exists a polynomial-time $\frac{4(k-1)}{k}$-approximation for* MAX POS ONE-IN-k-SAT.

Sketch of the proof. Given a set $\mathcal{C}$ of clauses with each clause containing exactly k positive literals, for any clause $C = \{x_{i(1)}, x_{i(2)}, \ldots, x_{i(k)}\}$ in $\mathcal{C}$, we consider the subset $F(C) = \{i(1), i(2), \ldots, i(k)\} \subseteq [n]$. Then, we apply Algorithm $\texttt{greedy}_{\texttt{MSF}(\Delta)}$ on the instance $\mathcal{F}(\mathcal{C}) = \{F(C) \ : \ C \in \mathcal{C}\}$ (notice that $\Delta = k$). The same probabilistic argument adopted in the proof of Theorem 1 guarantees that the first selector S_1, computed by the algorithm, satisfies at least $\frac{k}{4(k-1)}$ clauses of $\mathcal{C}$. The corollary, hence, follows. □

4 Hardness Results

By adopting the definitions in the proof of Corollary 5, and from the fact that Max Pos One-In-k-Sat is NP-hard (see [17,13]), it easily follows

Theorem 6 *It is* NP*-hard to approximate* Min Selective Family *within a factor smaller than 2.*

The following result reverts the connection between selective families and broadcast protocols. Indeed, it shows that any non-approximability result for the Min Selective Family directly translates into an equivalent negative result for Min Broadcast, when restricted to networks of constant eccentricity.
Let Min D-Broadcast denote the restriction of Min Broadcast to networks of eccentricity D.

Theorem 7 *For any fixed positive integer* $D \geq 2$*, if* Min D-Broadcast *is* r*-approximable, then* Min Selective Family *is* $\frac{rD-1}{D-1}$*-approximable.*

Sketch of the proof. Let $\mathcal{F}$ be an instance of Min Selective Family, where $\mathcal{F} = \{F_1, \ldots, F_m\}$ is a collection of subsets of $[n]$. We construct (in polynomial time) an instance $\langle G_D^{\mathcal{F}}, s\rangle$ of Min D-Broadcast such that $\mathcal{F}$ has a selective family of size k if and only if $\langle G_D^{\mathcal{F}}, s\rangle$ has a broadcast protocol with completion time equal to $1+k(D-1)$. The network $G_D^{\mathcal{F}}$ is a $D+1$ layered graph with layers $L_0, \ldots, L_D$, with $L_0 = \{s\}$ and the number of nodes in $G_D^{\mathcal{F}}$ is at most $n|\mathcal{F}|^D$. The graph $G_D^{\mathcal{F}}$ is defined by induction on D:
Base Step ($D = 2$). The network $G_2^{\mathcal{F}}$ consists of three levels: $L_0 = \{s\}$, $L_1 = \{x_1, \ldots, x_n\}$ and $L_2 = \{y_1, \ldots, y_m\}$, where s is connected to every $x_i \in L_1$, and the edge (x_i, y_j) exists iff $x_i \in F_j$.
Inductive Step. The graph $G_{D+1}^{\mathcal{F}}$ can be obtained from $G_D^{\mathcal{F}}$ as follows: The layer L_{D+1} of $G_{D+1}^{\mathcal{F}}$ is obtained by replacing every node in the layer L_D of $G_D^{\mathcal{F}}$ by a copy of the graph $G_2^{\mathcal{F}} \setminus \{s\}$. More formally,

1. Replace every $z_i \in L_D$ by the set $X_{D+1}^i = \{x_1^i(D+1), \ldots, x_n^i(D+1)\}$; each of such new vertices has the same in-neighborhood of z_i.
2. Add a set $Y_{D+1}^i = \{y_1^i(D+1), \ldots, y_m^i(D+1)\}$ of m new nodes. Then, add the edge $(x_k^i(D+1), y_l^i(D+1))$ if and only if (x_k, y_l) is an edge in $G_2^{\mathcal{F}}$.

So, the layer L_D of $G_{D+1}^{\mathcal{F}}$ is the union of all X_{D+1}^i's determined by the last level of $G_D^{\mathcal{F}}$ and the layer L_{D+1} of $G_{D+1}^{\mathcal{F}}$ is the union of all Y_{D+1}^i's.

Claim 3 *$\mathcal{F}$ has a selective family of size k if and only if $\langle G_D^{\mathcal{F}}, s\rangle$ has a broadcast protocol with completion time equal to* $1+k(D-1)$.

Sketch of the proof. The proof is by induction on D.
Base Step ($D = 2$). Consider the family $\mathcal{F}^{L_2}$ of in-neighborhoods of the nodes in L_2. Then, $G_2^{\mathcal{F}}$ admits a broadcast protocol of completion time $k+1$ iff $\mathcal{F}^{L_2}$ has a selective family of size k. Since $\mathcal{F} = \mathcal{F}^{L_2}$, then the theorem follows.

Inductive Step. ($\Rightarrow$). It is easy to show that by a suitable iteration of the broadcast protocol for $G_2^{\mathcal{F}}$ (yielded by the selective family $\mathcal{S}$ for $\mathcal{F}$) on $G_D^{\mathcal{F}}$, we obtain a completion time $1 + k \cdot (D-1)$, for any D.
($\Leftarrow$). Consider any broadcast protocol P for $\langle G_{D+1}^{\mathcal{F}}, s\rangle$ with completion time $1 + kD$. Also, let t be the number of time slots required by P to inform all the nodes in the second-last layer L_D of $G_{D+1}^{\mathcal{F}}$. It is easy to see that P completes broadcasting on $\langle G_D^{\mathcal{F}}, s\rangle$ within time-slot t.

If $t \leq 1 + k(D-1)$, then by inductive hypothesis, $\mathcal{F}$ has a selective family of size k. Otherwise, we first observe that for any i, all the nodes in X_{D+1}^i have the same in-neighborhood, thus implying that they are informed at the same time slot. Hence, there must exist a set X_{D+1}^{last} that is informed (according to P) at time slot t. Let $t' = t + \Delta t$ be the number of time slots necessary to P to inform Y_{D+1}^{last}, that is, the set of out-neighbors of X_{D+1}^{last}. From the fact that $t > 1 + k(D-1)$ and $t' = t + \Delta t \leq 1 + kD$, we obtain $\Delta t < k$. By construction, the subgraph induced by $X_{D+1}^{last} \cup Y_{D+1}^{last}$ is isomorphic to $G_2^{\mathcal{F}} \setminus \{s\}$. Hence, there exists a protocol for $\langle G_2^{\mathcal{F}}, s\rangle$ with broadcasting time $1 + \Delta t < 1 + k$. By inductive hypothesis, $\mathcal{F}$ has a selective family of size k. □

The proof of Claim 3 easily implies the following

Claim 4 *Given any broadcast protocol P with completion time on $G_D^{\mathcal{F}}$ equal to t, it is possible to construct (in time polynomial in $|G_D^{\mathcal{F}}|$) a protocol P' with completion time on $G_D^{\mathcal{F}}$ equal to $t' = 1 + k(D-1) \leq t$, for some integer $k \geq 1$.*

Consider any r-approximation algorithm for MIN D-BROADCAST. From Claim 4, we can assume that such an algorithm returns a broadcast protocol for $\langle G_D^{\mathcal{F}}, s\rangle$ of completion time $APX(G_D^{\mathcal{F}}) = 1 + k \cdot (D-1)$, for some $k \geq 1$. By hypothesis, it holds that

$$\frac{APX(G_D^{\mathcal{F}})}{OPT(G_D^{\mathcal{F}})} = \frac{1 + k \cdot (D-1)}{1 + OPT(\mathcal{F}) \cdot (D-1)} \leq r, \tag{2}$$

which implies that

$$\frac{k}{OPT(\mathcal{F})} \leq \frac{r[1 + OPT(\mathcal{F}) \cdot (D-1)] - 1}{OPT(\mathcal{F}) \cdot (D-1)} \leq \frac{rD - 1}{D-1}.$$

Finally, by applying Claim 3, we can construct (in polynomial time) a selective family for $\mathcal{F}$ of size at most $k \leq OPT(\mathcal{F}) \frac{rD-1}{D-1}$. Hence the theorem follows. □

By making use of Theorem 6 and Theorem 7, we can easily obtain the following result (whose proof is here omitted).

Corollary 8 *For any constant $D \geq 2$, it is* NP*-hard to approximate* MIN D-BROADCAST *within a factor less than* $2 - 1/D$*. Moreover, for any positive integer $c \geq 1$,* MIN $(\log^{c/(c+1)} n)$-BROADCAST *cannot be approximated by a factor less than 2 (unless* NP $\subseteq$ DTIME$[n^{\log^c n}]$*).*

Remark 9 *The $O(D + \log^5 n)$ broadcasting protocol of [11] implies that* MIN $D(n)$-BROADCAST *is in* APX *for any $D(n) \in \Omega(\log^5 n)$.*

5 Open Problems

The main open problem which is related to this paper consists of determining whether the de-randomization techniques can also be applied to the probabilistic construction of selective families given in [9]. We suspect that this is not true and, hence, that, in order to constructively achieve the upper bound of [9], alternative techniques have to be used.

References

1. N. Alon,, A. Bar-Noy, N. Linial, and D. Peleg (1991), A lower bound for radio broadcast, *J. Comput. System Sci.*, 43, 290-298.
2. N. Alon and J. Spencer (1992), *The probabilistic method*, Wiley.
3. R. Bar-Yehuda, O. Goldreich, and A. Itai (1992), On the time-complexity of broadcast operations in multi-hop radio networks: an exponential gap between determinism and randomization, *J. Comput. System Sci.*, 45, 104–126.
4. I. Chlamtac and S. Kutten (1985), On Broadcasting in Radio Networks - Problem Analysis and Protocol Design, *IEEE Trans. on Communications*, 33, 1240–1246.
5. I. Chlamtac, O. Weinstein (1991), The Wave Expansion Approach to Broadcasting in Multihop Radio Networks, *IEEE Trans. on Communications*, 39, 426–433.
6. B.S. Chlebus, L. Gąsieniec, A.M. Gibbons, A. Pelc, and W. Rytter (2000), Deterministic broadcasting in unknown radio networks, *Proc. of 11th ACM-SIAM SODA*, 861–870.
7. B. S. Chlebus, L. Gąsieniec, A. Ostlin, and J. M. Robson (2000), Deterministic radio broadcasting, *Proc. of 27th ICALP*, LNCS, 1853, 717–728.
8. M. Chrobak, L. Gąsieniec, W. Rytter (2000), Fast broadcasting and gossiping in radio networks, *Proc. of 41st IEEE FOCS*, 575–581.
9. A.E.F. Clementi, A. Monti, and R. Silvestri (2001), Selective Families, Superimposed Codes, and Broadcasting on Unknown Networks, *Proc. 12th ACM-SIAM SODA*, 709–718.
10. R. Gallager (1985), A perspective on multiaccess channels, *IEEE Trans. Inform. Theory*, 31, 124–142.
11. I. Gaber and Y. Mansour (1995), Broadcast in Radio Networks, *Proc. 6th ACM-SIAM SODA*, 577–585.
12. R. Motwani and P. Raghavan (1995), *Randomized Algorithms*, Cambridge University Press.
13. C.H. Papadimitriou (1994), *Computational Complexity*, Addison Wesley.
14. C.H. Papadimitriou and M. Yannakakis (1991), Optimization, approximation, and complexity classes, *J. Comput. System Sci.*, 43, 425–440.
15. A. Pelc (2000), Broadcasting in Radio Networks, unpublished manuscript.
16. L.G. Roberts (1972), Aloha Packet System with and without Slots and Capture, ASS Notes 8, Advanced Research Projects Agency, Network Information Center, Stanford Research Institute.
17. T.J. Schaefer (1979), The complexity of satisfiability problems, *Proc. 10th ACM STOC*, 216–226.