Lecture 10
Computability for Alphabet Size
Computability for Alphabet Size

Outline

- Lower and upper bounds are close
- Determining the alphabet size is NP-hard
- Complexity classification
- Maximum alphabet with no solution
Routing Multicast Rate

**What is the best multicast rate that can be achieved by (fractional) routing?**

A multicast tree in an acyclic directed network $G = (V, E)$ is a directed subtree of $G$, that span the source node and all the receivers of $G$.

Let $K$ denote the set of multicast trees and let $K_e$ denote the subset of $K$ with the trees that contain the edge $e \in E$. 

**REMINDER**
Routing Multicast Rate

For each tree $k \in K$ assign a flow rate $f_k$ which means that every edge of $k$ has a flow rate $f_k$.

The maximum rate that we can have is:

$$\max_{k \in K} \sum_{k \in K} f_k$$

Subject to

$$\sum_{k \in K_e} f_k \leq 1$$

for all $e \in E$

$$f_k \geq 0$$

for all $k \in K$
Routing Multicast Rate

The number of multicast trees in $G$ may be very large and hence generally it will be computationally infeasible to solve this linear program. Except for small networks.

Optimal fractional packing of distribution trees is NP-hard.

Jain, Mahdian, Salavatipour 2003
Let $G = (V, E)$ be an undirected graph, in which each edge has some weight, and let $T$ be a subset of vertices from $V$.

A Steiner tree is a tree, with minimum sum for the weights of the edges, which contains all the vertices of $T$.

The vertices in this tree which are not contained in $T$ are called Steiner vertices.
Steiner Trees

The Steiner tree problem in graphs is defined as a decision problem:

- an undirected graph $G = (V, E)$;
- a subset of the vertices $T \subseteq V$, called terminal nodes;
- a number $k \in \mathbb{N}$.

Instance

Is there a subtree $G'$ of $G$ that includes all the vertices of $T$ (a spanning tree for $G'$) and contains at most $k$ edges?

Question

The Steiner tree problem is NP-complete.

Theorem

Multicasting is NP-complete.

REMINDER
The vertices of the set \( T \) are points in the plane and they should be connected only by horizontal and vertical lines, where the sum of the lengths of these lines is minimized.

The decision problem for the rectilinear Steiner tree problem is also NP-complete.
Bounds on the Alphabet Size

Consider $h = 2$ messages.

The encoding for the edges can be taken from $[0, 1]$, $[1, 0]$, $[1, \alpha^i]$, $0 \leq i \leq q - 2$, where $\alpha$ is a primitive element in $\mathbb{F}_q$.

Any field of size $q \geq \left\lceil \sqrt{2N - 7/4} + 1/2 \right\rceil$ is sufficient and sometimes necessary.

Fragouli, Soljanin 2006
Bounds on the Alphabet Size

$h = 2$ messages.

$N$ receivers.

Field size $q \geq \left\lceil \sqrt{2N - \frac{7}{4}} + \frac{1}{2} \right\rceil$ necessary and sufficient.

$h > 2$ messages.

Field size $q \geq \left\lceil \sqrt{2N - \frac{7}{4}} + \frac{1}{2} \right\rceil$ necessary.

$N_{2,r,2}$

$N = \binom{r}{2}$

Field size $q \geq r - 1 \geq \left\lceil \sqrt{2N} \right\rceil$
Bounds on the Alphabet Size

REMINDER

\( h > 2 \) messages.

Field size \( q \geq \left\lfloor \sqrt{2N - \frac{7}{4}} + \frac{1}{2} \right\rfloor \) necessary and sufficient

LIF Algorithm

Any field of size \( q \geq N \) is sufficient.

\( \Theta(\log N) \) bits are necessary and sufficient to store an alphabet symbol.
Determining the Alphabet Size

**Theorem**

Rasala Lehman, Lehman 2004

Deciding whether there exists a linear network code with alphabet of size $q$ for a multicast information flow problem is NP-hard, where $q$ is a prime power.

**Proof**

The proof is by a reduction from the coloring decision problem on undirected graphs.

Let $G' = (V', E')$ be an undirected graph. An information flow graph $G$ is constructed as follows.
Determining the Alphabet Size

Proof

From an undirected graph \( G' = (V', E') \), construct an information flow graph \( G \).

- **nodes of** \( G \)
- **a single source** \( s \)
- **an intermediate node** \( v_i \) for each vertex \( v'_i \in V' \)
- **a receiver** \( t_{ij} \) for each edge \( (v'_i, v'_j) \in E' \)
- **edges of** \( G \)

There is an edge \( s \rightarrow v_i \) for each vertex \( v'_i \in V' \)

There are edges \( v_i \rightarrow t_{ij} \) and \( v_j \rightarrow t_{ij} \) for each \( (v'_i, v'_j) \in E' \).
Determining the Alphabet Size

**Proof**

- nodes of $G$
- a single source $s$
- an intermediate node $v_i$ for each vertex $v'_i \in V'$
- a receiver $t_{ij}$ for each edge $(v'_i, v'_j) \in E'$
- edges of $G$
- there are edges $v_i \rightarrow t_{ij}$ and $v_j \rightarrow t_{ij}$ for each $(v'_i, v'_j) \in E'$
- there is an edge $s \rightarrow v_i$ for each vertex $v'_i \in V'$

Two messages $M_1$ and $M_2$ are available at the source and are demanded by each receiver.

We will show that $G'$ is $q + 1$ colorable if and only if $G$ is solved by a linear network code with an alphabet of size $q$. 
Let $G' = (V', E')$ be an undirected graph. An information flow graph $G$ is constructed as follows. The nodes of $G$ consists of a of a single source $s$, an intermediate node $v_i$ for each vertex $v_i' \in V'$, and a receiver $t_{ij}$ for each edge $(v_i', v_j') \in E'$. There is an edge $s \to v_i$ for each vertex $v_i' \in V'$, and there are edges $v_i \to t_{ij}$ and $v_j \to t_{ij}$ for each $(v_i', v_j') \in E'$. The source has two messages $M_1$ and $M_2$ which are demanded by each receiver. We show that $G'$ is $q + 1$ colorable if and only if $G$ is solved by a linear network code with an alphabet of size $q$.

Assume $G'$ is $q + 1$ colorable with colors \{1, 2, ..., $q + 1$\}

Let $A_1, A_2, ..., A_{q+1}$ be the set of $q + 1$ pairwise orthogonal $q \times q$ squares ($q - 1$ are Latin)

Let $c(i)$ denote the color of vertex $v_i'$

the edges $s \to v_i$ and $v_i \to t_{ij}$ we send the symbol $A_{c(i)}(M_1, M_2)$
Determining the Alphabet Size

**Proof**

Let $G' = (V', E')$ be an undirected graph. An information flow graph $G$ is constructed as follows. The nodes of $G$ consists of a of a single source $s$, an intermediate node $v_i$ for each vertex $v'_i \in V'$, and a receiver $t_{ij}$ for each edge $(v'_i, v'_j) \in E'$. There is an edge $s \to v_i$ for each vertex $v'_i \in V'$, and there are edges $v_i \to t_{ij}$ and $v_j \to t_{ij}$ for each $(v'_i, v'_j) \in E'$. The source has two messages $M_1$ and $M_2$ which are demanded by each receiver. We show that $G'$ is $q + 1$ colorable if and only if $G$ is solved by a linear network code with an alphabet of size $q$.

Assume $G'$ is $q + 1$ colorable with colors $\{1, 2, \ldots, q + 1\}$

$A_1, A_2, \ldots, A_{q+1}$ a set of $q + 1$ pairwise orthogonal $q \times q$ squares

Let $c(i)$ denote the color of vertex $v'_i$

On $s \to v_i$ and $v_i \to t_{ij}$ we send the symbol $A_{c(i)}(M_1, M_2)$

$t_{ij}$ receives $A_{c(i)}(M_1, M_2)$ and $A_{c(j)}(M_1, M_2)$

colors on adjacent vertices are distinct implies $c(i) \neq c(j)$

$M_1, M_2$ can be reconstructed by $t_{ij}$

$M_1, M_2$ can be reconstructed by $t_{ij}$
Determining the Alphabet Size

Proof
Assume that there exists a linear network coding solution for the information flow problem $G$ with an alphabet of size $q$.

Each edge $s \rightarrow v_i$ carries a nonzero linear combination $\alpha M_1 + \beta M_2$.

All such linear combinations can be partitioned into $q + 1$ equivalence classes, where pair of linear combinations are in the same equivalent class if and only if they are dependent.

Each class is assigned a different color.
Determining the Alphabet Size

**Proof**
Assume there exists a linear network coding solution for the information flow problem $G$ with an alphabet of size $q$

$s \rightarrow v_i$ carries a nonzero linear combination $\alpha M_1 + \beta M_2$

All linear combinations can be partitioned into $q + 1$ equivalence classes, where pair of linear combinations are in the same equivalent class if and only if they are dependent

Each class is assigned a different color

$v'_i \in V'$ is colored with the color associated with the equation $s \rightarrow v_i$

The vertices of the edge $(v'_i, v'_j) \in E'$ are colored differently

$t_{ij}$ must obtain $M_1$ and $M_2$ and hence $s \rightarrow v_i$ and $s \rightarrow v_j$ must have independent linear combinations

$G'$ is colored with $q + 1$ distinct colors
For a given undirected graph $H$, an $H$-coloring of an undirected graph $G$ is a homomorphism $h : G \rightarrow H$ such that $h(v)$ and $h(u)$ are adjacent vertices of $H$ if $v$ and $u$ are adjacent vertices of $G$.

**Theorem**

$H$-coloring is NP-hard if $H$ is not a bipartite graph and it is solvable in polynomial time if $H$ is bipartite.

Hell, Nestril 1990
Determining the Alphabet Size

Theorem
Deciding whether there exists a network code with alphabet of size $q$ for a multicast information flow problem is NP-hard, where $q$ is a prime power.

Proof
The proof is by using a reduction from the $H$-coloring problem on undirected graphs.

The vertices of the graph $H$ consists of all $q \times q$ arrays with elements from an alphabet of size $q$.

Two vertices (arrays) are joined by an edge if the corresponding arrays are orthogonal.

Clearly the graph is not bipartite.
Determining the Alphabet Size

Proof

Similar to the previous proof. Let $G' = (V', E')$ be an undirected graph. An information flow graph $G$ is constructed as follows. The nodes of $G$ consists of a single source $s$, a node $v_i$ for each $v'_i \in V'$, and a receiver $t_{ij}$ for each $(v'_i, v'_j) \in E'$. There is an edge $s \rightarrow v_i$ for each $v'_i \in V'$, and there are edges $v_i \rightarrow t_{ij}$ and $v_j \rightarrow t_{ij}$ for each $(v'_i, v'_j) \in E'$. Two messages $M_1$ and $M_2$ are available at the source and are demanded by each receiver. We show that $G'$ is $H$-colorable if and only if $G$ is solved by a network code with an alphabet of size $q$. 
Determining the Alphabet Size

Proof

Let $G' = (V', E')$ be an undirected graph. An information flow graph $G$ is constructed as follows. The nodes of $G$ consists of a of a single source $s$, an intermediate node $v_i$ for each vertex $v'_i \in V'$, and a receiver $t_{ij}$ for each edge $(v'_i, v'_j) \in E'$. There is an edge $s \rightarrow v_i$ for each vertex $v'_i \in V'$, and there are edges $v_i \rightarrow t_{ij}$ and $v_j \rightarrow t_{ij}$ for each $(v'_i, v'_j) \in E'$. Two messages $M_1$ and $M_2$ are available at the source and are demanded by each receiver. We show that $G'$ is $H$-colorable if and only if $G$ is solved by a linear network code with an alphabet of size $q$.

Assume $G'$ is $H$-colorable

Send to vertex $v_i$ the symbol $A(M_1, M_2)$, where $A$ is the array associated with the vertex $h(v'_i)$.

Each receiver $t_{ij}$ gets its inputs from two vertices $v_i$ and $v_j$ such that $v'_i$ and $v'_j$ are adjacent in $G'$.
Determining the Alphabet Size

**Proof**

Assume $G'$ is $H$-colorable

Send to vertex $v_i$ the symbol $A(M_1, M_2)$, where $A$ is the array associated with the vertex $h(v'_i)$

Each receiver $t_{ij}$ gets its inputs from two vertices $v_i$ and $v_j$ such that $v'_i$ and $v'_j$ are adjacent in $G'$

Hence, $h(v'_i)$ and $h(v'_j)$ are adjacent in $H$ and are related to orthogonal arrays

Therefore, $t_{ij}$ who knows the arrays related to $v'_i$ and $v'_j$ can compute $M_1$ and $M_2$ from pair of symbols received from $v_i$ and $v_j$
Determining the Alphabet Size

Proof Assume that there exists a linear network coding solution for the information flow problem $G$ with an alphabet of size $q$.

construct an $H$-coloring $h$ for the graph $G'$

For each vertex $v_i$ in $G$, let $h(v'_i)$ be the vertex of $H$ which corresponds to the array which defines the function of $M_1, M_2$ received and sent by $v_i$

If $v'_i$ and $v'_j$ are adjacent in $G'$ then the corresponding vertices $v_i$ and $v_j$ in $G$ share the receiver $t_{ij}$. 
Determining the Alphabet Size

**Proof** Assume there exists a linear network coding solution for the information flow problem $G$ with an alphabet of size $q$ construct an $H$-coloring $h$ for the graph $G'$.

For each vertex $v_i$ in $G$, let $h(v'_i)$ be the vertex of $H$ which corresponds to the array which defines the function of $M_1, M_2$ received and sent by $v_i$. If $v'_i$ and $v'_j$ are adjacent in $G'$ then the corresponding vertices $v_i$ and $v_j$ in $G$ share the receiver $t_{ij}$. Since $t_{ij}$ can reconstruct $M_1$ and $M_2$, it follows that the arrays related to $h(v'_i)$ and $h(v'_j)$ are orthogonal and therefore adjacent in $H$. Thus, $h$ is an $H$-coloring of $G$. 
An information flow is defined by four attributes: single or multiples sources, single or multiple receivers, message distribution at sources, and message distribution at receivers.

The class of information flow problems is defined by a 4-tuple $(\alpha, \beta, \gamma, \delta)$ with the following interpretation.

- $\alpha$ is 1 if there is a single source and $n$ if there are multiple sources.
- $\beta$ is 1 if there is a single receiver and $m$ if there are multiple receivers.
The class of information flow problems is defined by a 4-tuple \((\alpha, \beta, \gamma, \delta)\) with the following interpretation.

- \(\alpha\) is 1 if there is a single source and \(n\) if there are multiple sources.
- \(\beta\) is 1 if there is a single receiver and \(m\) if there are multiple receivers.
- \(\gamma\) is \(I\) if all messages are available at every source, \(D\) if each two sources have disjoint sets of messages, and \(A\) if everything is possible.
- \(\delta\) is \(I\) if each receiver demands all the messages, \(D\) if each two receivers demand disjoint sets of messages, and \(A\) if everything is possible.
Three categories of problems

- **Trivial codes**: The problem is solved only with routing.
- **Linear codes**: If the network is solvable then there exists a linear network code solution that can be found in polynomial time.
- **Hard problems**: Determining if a linear solution over a given alphabet exists is \( \text{NP-hard} \). Some networks permit a solution with a nonlinear code.
An instance of information flow problem \((1 \text{ or } n, 1, I \text{ or } D \text{ or } A, I)\) or \((1 \text{ or } n, m, I, D)\) has a solution if and only if there is a routing solution.

An instance of information flow problem \((1 \text{ or } n, m, I \text{ or } D \text{ or } A, I)\) is solvable in polynomial time (whether there exists or there does not exist a solution).

Deciding whether there exists a linear network coding solution to information flow problem in the classes \((1 \text{ or } n, m, I, A)\) or \((n, m, I, D \text{ or } A, D \text{ or } A)\) is NP-hard.
Complexity Classification

**Theorem** An instance of information flow problem \((n, m, I, D)\), has a solution if and only if there is a routing solution.

**Proof**

Construct \(G' = (V', E')\) from \(G\)

Add to \(G\) a super source \(S^*\) and a super receiver \(R^*\)

add \(h\) edges from \(S^*\) to each source \(S_i\) of \(G\)

from each receiver \(R_i\) of \(G\) add \(\ell_i\) edges to \(R^*\), where \(\ell_i\) is the number of messages demanded by \(R_i\).

In \(G'\) all messages are in \(S^*\) and demanded by \(R^*\).
Complexity Classification

Proof

Let $G = (V, E)$ be a network with the information flow problem \((n, m, I, D)\). Construct a graph $G' = (V', E')$ by adding first to $G$ a super source $S^*$ and $h$ edges from $S^*$ to each source $S_i$ of $G$. Add a super receiver $R^*$ and from each receiver $R_i$ of $G$ add $\ell_i$ edges to $R^*$, where $\ell_i$ is the number of messages demanded by $R_i$. In $G'$ all messages are in $S^*$ and demanded by $R^*$.

Clearly, if the problem on $G$ is solvable, then the problem on $G'$ is solvable.

If the problem on $G'$ is solvable then the min-cut has size $h$.

Each such path passes through one of the sources of $G$ and $\ell_i$ paths passes through $R_i$.

Thus, these edge-disjoint paths define a routing solution for $G$.

Hence, there exists $h$ edge-disjoint paths from $S^*$ to $R^*$. 
An instance of information flow problem with a single receiver has a solution if and only if there is a routing solution.

Let $G = (V, E)$ be the network with a single receiver $R$. Construct $G' = (V', E')$ from $G$:

- Add to $G$ a super source $S^*$ and a node $\mu_i$ for each message $M_i$.
- Add an edge from $S^*$ to each node $\mu_i$ and an edge from $\mu_i$ to each source for which one of the messages was $M_i$. 

Proof

Theorem

Complexity Classification
Let $G = (V, E)$ be the network with the single receiver $R$. Construct a new graph $G' = (V', E')$ by adding to $G$ a super source $S^*$ and a node $\mu_i$ for each message $M_i$. In addition add an edge from $S^*$ to each node $\mu_i$ and an edge from $\mu_i$ to each source for which one of the messages was $M_i$.

Clearly, if the problem on $G$ is solvable, then the problem on $G'$ is solvable.

If the problem on $G'$ is solvable then the min-cut has size at least $h$.

If the maximum of flow is $h$, then one unit flow passes through each node $\mu_i$.

Hence, the $h$ edge disjoint paths of $G'$ from the $\mu_i$'s can be used to route the $M_i$'s to $R$ (the $\mu_i$ is connected to the appropriate sources of $G$).
An instance of information flow problem with every message requested by every receiver is polynomial time solvable.

Let $G = (V, E)$ be the network for which every message is requested by every receiver.

Construct $G' = (V', E')$ from $G$

- add to $G$ a super source $S^*$ and a node $\mu_i$ for each message $M_i$
- add an edge from $S^*$ to each node $\mu_i$ and an edge from $\mu_i$ to each source for which one of the messages was $M_i$. 
Complexity Classification

Proof

Let \( G = (V, E) \) be the network for which every message is requested by every receiver. Construct a new graph \( G' = (V', E') \) by adding to \( G \) a super source \( S^* \) and for each message \( M_i \) add a node \( \mu_i \). In addition add an edge from \( S^* \) to each node \( \mu_i \) and an edge from \( \mu_i \) to each source for which one of the messages was \( M_i \).

To each receiver \( R_i \) find a flow \( F_i \) of size \( h \) in \( G' \). Now, apply the LIF algorithm to find a linear network coding solution in polynomial time.

To each receiver \( R_i \) find a flow \( F_i \) of size \( h \) in \( G' \). Now, apply the LIF algorithm to find a linear network coding solution in polynomial time.
There are mappings of a 3-CNF (conjunctive normal form) formula to an information flow problem in the classes $(1 \text{ or } n, m, I, A)$ or $(n, m, I, D \text{ or } A, D \text{ or } A)$.

**Lemma 1**

A 3-CNF formula $\phi$ is satisfiable if and only if the corresponding information flow problem has a linear network coding solution.

**Theorem**

Deciding whether there exists a linear network coding solution to information flow problem in the classes $(1 \text{ or } n, m, I, A)$ or $(n, m, I, D \text{ or } A, D \text{ or } A)$ is NP-hard.
There is mapping of a 3-CNF (conjunctive normal form) formula to an information flow problem in the class \((n, m, D, A)\). Let \(\phi\) be a 3-CNF formula over the variables \(x_1, \ldots, x_d\).

\[
c_i = (x_j \lor \overline{x_k} \lor x_\ell)
\]
There are solvable information flow problems in the classes \((1 \text{ or } n, m, I, A) \text{ or } (n, m, I, D \text{ or } A, D \text{ or } A)\) for which there is no linear network.

**Theorem**

For \((n, m, D, A)\) let \(\phi\) be an unsatisfiable 3-CNF formula.

**Proof**

By **Lemma 1**, there is no linear solution to the corresponding information flow problem.

there exists a nonlinear solution

Consider the alphabet \(\Sigma = \Gamma \times \Gamma\) for the messages, where \(\Gamma\) is arbitrary alphabet.
Complexity Classification

**Proof** For \((n, m, D, A)\) let \(\phi\) be an unsatisfiable 3-CNF formula. By the Lemma, there does not exist a linear solution to the corresponding information flow problem. However, there exists a nonlinear solution. Consider the alphabet \(\Sigma = \Gamma \times \Gamma\), where \(\Gamma\) is arbitrary alphabet, i.e. each message is a pair from \(\Gamma\).

Each source node \(s_j\) send the first symbol of the messages \(M_j\) and \(\overline{M}_j\) as a pair to \(r_j\) and send the second symbol of these messages to all its other outgoing edges.

Consider the clause \(c_i = (x_j \lor \overline{x_k} \lor x_\ell)\)

The receiver \(t_i\) demands messages \(M_j, \overline{M}_k, M_\ell\)

From nodes \(r_j, r_k, r_\ell\) it receives the first symbol of \(M_j, \overline{M}_j, M_k, \overline{M}_k, M_\ell, \overline{M}_\ell\)

Each one can send to \(t_i\) two symbols from \(\Gamma\) and hence they can provide together the three symbols that \(t_i\) needs.

The nodes \(u_i\) and \(v_i\) receive the second symbol of these six messages.
For a given network, let $q_{\text{min}}$ be the minimum field size for the existence of a linear solution.

Let $q^*_{\text{max}}$ be the maximum field size for the non-existence of a linear solution.

What are the possible gaps $q^*_{\text{max}} - q_{\text{min}}$?
Research Problems

Introduce a better hierarchy (more classes) for the complexity classification.

Introduce separate hierarchies for linear and nonlinear codes.

What are the possible gaps for $q_{\text{max}}^* - q_{\text{min}}$? Can this gap tend to infinity? Consider the gap as a function of $h$ and/or $N$. 
Introduction to Network Coding, Bounds and Constructions

END OF LECTURE 10