Introduction to Network Coding, Bounds and Constructions

Tuvi Etzion

Lecture 9

Information Flow Decomposition
Information Flow Decomposition

Outline

- Code design and alphabet size
- The line graph, the subtree and code graphs
Information Flow Decomposition

The butterfly network

source flow

S₁ R₁
S₂ R₂

R₁ R₂

coding flow

Information Flow Decomposition
Information Flow Decomposition

two sources and two receivers

two possible information flow decompositions

\[ S_1 R_1 S_2 R_2 \]

\[ S_1 R_1 R_2 \quad S_2 R_1 R_2 \]

only routing

coding is required
A network $G$ with $h$ sources $S_1, \ldots, S_h$.

Emphasis $h = 2$!

Each source has one message.

$N$ receivers, $R_1, \ldots, R_N$ each one demands all the $h$ messages.

$G$ is (generally) a minimal multicast graph.

Fragouli, Soljanin, Shokrollahi 2004
Select a coding vector for each edge of the network such that the transfer matrices $C_1, C_2, \ldots, C_N$ will be of full rank.

The coding vector of the edge $e$ is in the linear span of the coding vectors of the input edges to the parent node of $e$.

The coding vectors should be selected only for the coding points.
For a directed graph $G$ the line graph $L(G)$ is defined as follows.

An edge $e$ in $G$ is a vertex $v(e)$ in $L(G)$.

If the edge $e_1$ in $G$ is an input of a vertex $v$ and the edge $e_2$ in $G$ is an output of the vertex $v$ in $G$, then there is an edge from $v(e_1)$ to $v(e_2)$ in $L(G)$.

The vertex in $L(G)$ which is formed from the edge $(u, v)$ of $G$ will be denoted by $uv$. 
Given a (minimal) multicast network $G$ with $h$ sources and $N$ receivers construct its line graph $L(G)$.

Fragouli, Soljanin 2006

A vertex denoted by $S_i u$ is called a source node.

If the edge $(u, v)$ of $G$ is an input for a receiver $R$ then the vertex $uv$ in $L(G)$ is labelled by $R$.

A vertex with at least two inputs is called a coding point.
Given a multicast network $G$ with $h$ sources and $N$ receivers construct the graph $\ell(G)$ as follows. For each path $(S_i, R_j)$ construct the line graph $L(S_i, R_j)$. Last vertex in $L(S_i, R_j)$ is labelled by $R_j$. This is the receiver node for $(S_i, R_j)$. 

$\ell(G) = \bigcup_{1 \leq i \leq h \atop 1 \leq j \leq N} L(S_i, R_j)$

$h \cdot N$ receiver nodes

Is $\ell(G) = L(G')$? where $G'$ is the minimal multicast network of $G$ obtained from union of the paths $(S_i, R_j)$?
Code Design – the Line Graph

\[ x, y \]

\[ S \]

\[ A \rightarrow B \rightarrow C \rightarrow R_1 \]

\[ D \rightarrow R_2 \]

\[ E \rightarrow R_3 \]

\[ S_1 \]

\[ A \rightarrow B \rightarrow C \rightarrow E \rightarrow R_1 \rightarrow S_2 \]

\[ D \rightarrow R_1 \]

\[ F \rightarrow R_2 \]

\[ G \rightarrow H \rightarrow K \rightarrow R_3 \]
Code Design – Line Graph
Given the line graph $L(G)$ of a minimal multicast network $G$, the subtree graph $\Gamma = (G, S, R)$ is defined by:

The vertices of $L(G)$ are partitioned into disjoint subsets $T_i$ such that:

- Each $T_i$ contains either exactly one source node or exactly one coding point;

- Every other node belongs to the $T_i$ containing its first ancestral coding or source node.
The vertices of $L(G)$ are partitioned into disjoint subsets $T_i$ such that:

- Each $T_i$ is a tree since the only nodes with two or more input edges in $L(G)$ are the coding points.
- The same linear combination of source symbols flows through all vertices of the same subtree.

- Each $T_i$ is a tree since the only nodes with two or more input edges in $L(G)$ are the coding points.
- The same linear combination of source symbols flows through all vertices of the same subtree.

- Each $T_i$ contains either exactly one source node or exactly one coding point.
- Every other node belongs to the $T_i$ containing its first ancestral coding or source node.

The vertices of $L(G)$ are partitioned into disjoint subsets $T_i$ such that:

- Each $T_i$ contains either exactly one source node or exactly one coding point.
- Every other node belongs to the $T_i$ containing its first ancestral coding or source node.
For each receiver $R_j$, there exist in $L(G)$, $h$ vertex-disjoint paths from the source subtrees. For each receiver $R_j$, the $h$ receiver nodes labelled with $R_j$ belong to distinct subtrees. Each subtree contains at most $N$ receiver nodes.

If $T_i$ starts with a source node then it will be called a source subtree.

If $T_i$ starts with a coding point then it will be called a coding subtree.
Each vertex in $L(G)$, which is either $S_{i,u}$ or has in-degree greater than one, is a root of a maximal subtree $T$.

Each maximal subtree is a vertex in $\Gamma$. A vertex in $\Gamma$ is labelled with the source and receivers that it contains in $L(G)$.

There is an edge from a vertex $T_1$ to a vertex $T_2$ in $\Gamma$ if there is an edge from a vertex of $T_1$ to vertex of $T_2$ in $L(G)$.

The maximal subtrees form a partition of the vertices in $L(G)$.

subtree decomposition
Code Design – Line Graph
Code Design – Subtree Graph

Source subtrees

Coding subtrees

\[ S_1 R_2 \]

\[ S_2 R_1 \]
Code Design – the Subtree Graph

Given the minimal multicast network \( G \), the code graph (identical to the subtree graph) \( \Gamma = (G, S, R) \) is defined by:

There is a vertex in \( \Gamma \) for each source in \( G \) and for each coding point in \( G \). These vertices will be called source nodes and coding nodes.

There is an edge between two nodes of \( \Gamma \) if there is a path between the corresponding objects in \( G \).
There is a vertex in $\Gamma$ for each source in $G$ and for each coding point in $G$. These vertices will be called source nodes and coding nodes.

There is an edge between two nodes of $\Gamma$ if there is a path between the corresponding objects in $G$.

A source node is labelled by the related source.

A node $v$ is labelled by a receiver $R$ if the coding vector of $v$ (or the related message) will appear in the transfer matrix of $R$. 

Code Design – the Subtree Graph
Code Design – the Code Graph
All networks that have identical code graphs are equivalent from coding point of view.
Introduction to Network Coding, Bounds and Constructions

A SHORT BREAK
Each receiver labels $h$ nodes.

Nodes labelled by $R_i$ are connected to the sources by $h$ node-disjoint paths (multicast property).

The graph has been made minimal - edge removed then multicast property is lost.

Coding vector of length $h$ should be assigned to each node.

Minimal subtree graph
For each node select a vector from $\mathbb{F}_q^h$ such that:

- The assignment to the node with $S_i$ is $e_i$.
- Vectors of the nodes which are labelled by the same receiver are linearly independent.
- The vector assigned to a node is in the linear span of the vectors assigned to its parents.
- The vector assigned to a node is not in the linear span of the vectors assigned to a proper subset of its parents.
Code Design – Subtree Graph

Coding vectors

\[ C(T_1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \]

\[ C(T_2) = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \]

\[ C(T_3) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \]

\[ C(T_4) = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \]

Transfer matrices

\[ C(R_1) = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \]

\[ C(R_2) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \]

\[ C(R_3) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \]

\[ \beta \in \mathbb{F}_q \setminus \{0, 1\} \]

\[ S_1 R_2 \]

\[ S_2 R_1 \]

\[ T_1 \]

\[ T_2 \]

\[ T_3 \]

\[ T_4 \]

\[ [1, 0] \]

\[ [0, 1] \]

\[ [1, 1] \]

\[ [0, 1] \] (or \[ [1, \beta] \])
Vectors assigned to parents of a subtree must be linearly independent.

A coding subtree has at least 2 and at most $h$ parents.

The coding vector assigned to a child cannot be a subspace spanned by a proper subset of his parents.

If a coding subtree $T$ has $2 \leq P \leq h$ parents, then $P$ vertex-disjoint paths exist from $P$ source nodes to $T$. 
For a network with 2 sources, a minimal subtree graph \( \Gamma \) with \( n \) nodes has the following properties:

1. A parent and a child subtree have a child or a receiver in common, or both.
2. Each coding subtree contains at least two receiver nodes.
3. Each source subtree contains a receiver node.

Theorem \( h = 2 \)

\( \Gamma \) has \( n - 2 \) coding subtrees since two nodes are source subtrees.

Fragouli, Soljanin 2006
Decentralized network coding assigns coding vectors to subtrees by taking into account only the local information available at the subtree, namely, which receiver nodes it contains and which coding vectors have been assigned to its parents subtrees.

The set of coding vectors which will be used for a network with $h$ sources, are taken from the columns of the generator matrix of an $h$-dimensional MDS code or the elements of $\text{PG}(h - 1, q)$. 
Code Design – Subtree Graph

Smallest Alphabet Code

Decentralized Code

\[ [1, 0] \]
\[ [0, 1] \]
\[ S_1 R_2 \]
\[ S_2 R_1 \]
\[ [1, 1] \]
\[ R_1 R_3 \]
\[ [1, \beta] \]
\[ \beta \in \mathbb{F}_q \setminus \{0, 1\} \]
\[ R_2 R_3 \]

field size?

\[ [1, 0] \]
\[ [0, 1] \]
\[ S_1 R_2 \]
\[ S_2 R_1 \]
\[ [1, 1] \]
\[ R_1 R_3 \]
\[ [1, 1] \]
\[ R_2 R_3 \]
\[ [0, 1] \]
Multicasting is Solvable

**Theorem**

For networks with 2 sources and \( N \) receivers

\[ q \geq \left\lceil \sqrt{2N - \frac{7}{4}} + \frac{1}{2} \right\rceil \]

is sufficient, and for some networks, necessary.

**Theorem**

Fragouli, Soljanin 2006

For networks with \( h \) sources and \( N \) receivers \( q \geq N \) is sufficient.

No network for which \( q > O(\sqrt{N}) \) is necessary is known.
Let $V$ be the following set of $q + 1$ vectors
$\{[0, 1], [1, 0]\} \cup \{[1, \alpha^i] : 0 \leq i \leq q - 2\}$, where $\alpha$ is a primitive element of $\mathbb{F}_q$.

Any two vectors in $V$

1. Linearly independent.
2. Span $V$.

$\Rightarrow$ Vectors in $V$ can be treated as colors.
**Coding for Networks with 2 Sources**

\( \Gamma \) - the subtree graph with \( n \) vertices.

\( \Omega \) - an undirected graph with the \( n \) vertices as \( \Gamma \).

Two vertices of \( \Omega \) are connected by an edge if:

1. the corresponding vertices of the subtrees of \( \Gamma \) are labelled with the same receiver.

2. the corresponding vertices of the subtrees of \( \Gamma \) have a common child.
Two vertices of $\Omega$ are connected by an edge if:

1. If the corresponding vertices of the subtrees of $\Gamma$ are labelled with the same receiver.

2. If the corresponding vertices of the subtrees of $\Gamma$ have a common child.

Two vertices of $\Omega$ are connected by an edge if the corresponding subtrees of $\Gamma$ cannot have the same coding vector.
Coding for Networks with 2 Sources
Coding for Networks with 2 Sources

\[ S_1 R_2 \]  [1, 0]  \[ 1, 1 \]

\[ S_2 R_1 \]  \[ 0, 1 \]  \[ S_2 R_1 \]

\[ \Omega \]  \[ \text{flow} \]  \[ \text{flow} \]

\[ R_2 \]  \[ R_2 \]  \[ R_1 \]

\[ R_1 R_3 \]  \[ R_2 R_3 \]  \[ R_1 R_3 \]

\[ R_3 \]  \[ R_3 \]  \[ \]
Coding for Networks with 2 Sources

For a network with 2 sources, a minimal subtree graph $\Gamma$ with $n$ nodes has the following properties:
1. A parent and a child subtree have a child or a receiver in common, or both.
2. Each coding subtree contains at least two receiver nodes.
3. Each source subtree contains a receiver node.

Lemma

Every vertex in $\Omega$ has degree at least 2

Proof

For source nodes: If $n = 3$ then the two sources have exactly one child which shares a receiver with each parent. If $n > 3$ then the two source subtrees have at least one child (in common) which shares a receiver or a child with each parent.
Lemma 1

Every vertex in $\Omega$ has degree at least 2.

Proof

For coding subtrees: Each coding subtree has two parents. It cannot be labelled with the same coding vector as either of its parents. Hence, in $\Omega$ there is an edge between a subtree and its parents. Thus, it has degree at least two.
Lemma 2

Every $k$-chromatic graph has at least $k$ vertices of degree at least $k - 1$.

Lemma 3

In $\Omega$ there are $N$ receiver edges and $n - 2$ flow edges.

Proof

Each receiver is labelled in two nodes. Flow edges exist between vertices with a common child, i.e. between parents of a coding subtree. There are $n - 2$ coding subtrees, each one has exactly two parents.
For networks with 2 sources and $N$ receivers, field size $q \geq \left\lfloor \sqrt{2N - 7/4} + 1/2 \right\rfloor$ is sufficient, and for some networks, necessary.

Proof: Assume that $\Omega$ has $n$ vertices and chromatic number $\chi(\Omega) = k \leq n$. We will count the number of edges in $\Omega$ in two different ways.
Coding for Networks with 2 Sources

Lemma 1: Every vertex in $\Omega$ has degree at least 2.

Lemma 2: Every $k$-chromatic graph has at least $k$ vertices of degree at least $k - 1$.

Lemma 3: $\Omega$ has $N$ receiver edges and $n - 2$ flow edges.

Proof: Assume that $\Omega$ has $n$ vertices and chromatic number $\chi(\Omega) = k \leq n$. We will count the number of edges in $\Omega$ in two different ways.

1. By the first two lemmas: $E(\Omega) \geq \frac{k(k-1)}{2} + n - k$.
2. By the third lemma: $E(\Omega) = N + n - 2$.

Hence, $N + n - 2 \geq \frac{k(k-1)}{2} + n - k$. We have $q + 1$ colors. Assign $k = q + 1$ and solve a quadratic equation for $q$, which implies $q \geq \left\lceil \sqrt{2N - \frac{7}{4}} + \frac{1}{2} \right\rceil$ is sufficient.
Bounds on the Alphabet Size

\[ h = 2 \] messages. \hspace{1cm} \text{\( N \) 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.

\[ N = \binom{r}{2} \]

\[ N_{2,r,2} \]

Field size \( q \geq r - 1 \geq \lfloor \sqrt{2N} \rfloor \)
Introduction to Network Coding, Bounds and Constructions

A SHORT BREAK
Information Flow Decomposition

\[ S_1 \rightarrow A \rightarrow C \rightarrow B \rightarrow D \rightarrow F \rightarrow H \rightarrow J \rightarrow R_1 \]

\[ S_2 \rightarrow B \rightarrow E \rightarrow G \rightarrow I \rightarrow K \rightarrow R_2 \]

\[ S_1 \rightarrow S_1A \rightarrow AD \rightarrow DF \rightarrow FH \rightarrow FJ \rightarrow R_1 \]

\[ S_1 \rightarrow S_1A \rightarrow AC \rightarrow CD \rightarrow GH \rightarrow IK \rightarrow R_2 \]

\[ S_2 \rightarrow S_2B \rightarrow BC \rightarrow CE \rightarrow GH \rightarrow IK \rightarrow R_2 \]

\[ R_1 \rightarrow IJ \rightarrow HI \rightarrow EG \rightarrow GK \rightarrow R_2 \]

\[ R_2 \rightarrow IJ \rightarrow HI \rightarrow EG \rightarrow GK \rightarrow R_2 \]
Information Flow Decomposition

The multicast property is not violated.

S1 S2 S1 S2 S1 S2 S1 S2

A B A B A B A B

C D C D C D C D

E F E F E F E F

G H G H G H G H

I J I J I J I J

K K K K K K K K

R1 R2 R1 R2 R1 R2 R1 R2
Bounds on the Alphabet Size

**Theorem**

Fragouli, Soljanin 2006

For networks with 2 sources and $N$ receivers,

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

is sufficient, and for some networks, necessary.
Bounds on the Alphabet Size

Binary solution

Each coding point (subtree) performs a binary addition of its two inputs

Using the coloring method, only three colors are required for the corresponding graph $\Omega$. Hence, a field of size $2$ is sufficient.
Bounds on the Alphabet Size

Each coding point (subtree) performs a binary addition of its two inputs.

Using the coloring method, only three colors are required for the corresponding graph $\Omega$. Hence, a field of size $2$ is sufficient.
Add a new receiver for each pair of subtrees that do not share a child.

- For $k$ subtrees, $N = \frac{k(k-1)}{2} - k + 2$.
- The chromatic number of $\Omega$ is $k$ and hence the field size is $q \geq k - 1$.

$$q \geq \left\lfloor \sqrt{2N - \frac{7}{4} + \frac{1}{2}} \right\rfloor$$
The code graph is a complete bipartite graph with parts of sizes $h$ and $q - 1$ (or $q - 1 - h$)

For each $i$, $1 \leq i \leq h$, remove all incoming edges of $A_i$, except for the one connecting it to $S_i$, to obtain a minimal multicast graph.
Information Flow Decomposition

Non-minimal subtree decomposition
Information Flow Decomposition

non-minimal multicast graph

minimal multicast graph
Information Flow Decomposition

minimal subtree decomposition
Information Flow Decomposition

\[ \Omega \text{ is a complete graph with 5 vertices} \]

\( S_1 \)

\( S_2 \)

\( [1 \ 0] \)

\( [0 \ 1] \)

\( \begin{bmatrix} 4 & 7 & 9 & 10 \end{bmatrix} \)

\( \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} \)

\( \begin{bmatrix} 1 & 5 & 6 & 7 \end{bmatrix} \)

\( \begin{bmatrix} 2 & 5 & 8 & 9 \end{bmatrix} \)

\( \begin{bmatrix} 3 & 6 & 8 & 10 \end{bmatrix} \)

minimal subtree decomposition
Three sources, no coding subtree has a child

Bipartite graph
Is there a solution in \( \text{PG}(2, 4) \)?

![Diagram of Information Flow Decomposition]

- **\( R_1R_2R_3 \)**
- **\( T_5 \)**
- **\( S_1 \): \([1 \ 0 \ 0]\)**
- **\( S_2 \): \([0 \ 1 \ 0]\)**
- **\( S_3 \): \([0 \ 0 \ 1]\)**
- **\( R_1R_2 \)**
- **\( R_3R_4 \)**
- **\( T_1 \)**
- **\( R_2R_4 \)**
- **\( T_2 \)**
- **\( R_1R_3 \)**
- **\( T_3 \)**
- **\( R_2R_4 \)**
- **\( T_4 \)
Information Flow Decomposition

\[
\begin{bmatrix}
1 & 1 & \alpha^2
\end{bmatrix}
\]

\(R_1R_2R_3\)  
\(T_5\)

\(S_1\)  
\([1 0 0]\)

\(R_1R_2\)  
\(T_1\)  
\([1 1 0]\)

\(R_3R_4\)  
\(T_2\)  
\([1 0 \alpha^2]\)

\(S_2\)  
\([0 1 0]\)

\(R_2R_4\)  
\(T_2\)  
\([1 0 \alpha^2]\)

\(S_3\)  
\([0 0 1]\)

\(R_1R_3\)  
\(T_3\)  
\([0 1 1]\)

\(R_2R_4\)  
\(T_4\)  
\([0 0 \alpha]\)

a solution in \(PG(2, 4)\)
Information Flow Decomposition

\[ C(R_1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & \alpha^2 \end{bmatrix} \]

\[ C(R_2) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & \alpha^2 \\ 1 & 1 & \alpha^2 \end{bmatrix} \]

\[ C(R_3) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & \alpha^2 \end{bmatrix} \]

\[ C(R_4) = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & \alpha^2 \\ 0 & 0 & \alpha^2 \end{bmatrix} \]

\[ C(S_1) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \]

\[ C(S_2) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \]

\[ C(S_3) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \]

a solution in \( \text{PG}(2, 4) \)
Improve the upper bound on the alphabet size for solvable multicast networks with \( h \) messages?

Improve the upper bound on the alphabet size for solvable multicast networks with 3 messages?

Can the alphabet size for solvable multicast networks with \( h \) messages be different for linear and nonlinear codes (for general networks)?