Random Network Coding

Outline

- Code design and alphabet size
- The line graph and the subtree graph
- Random network coding
- Coherent vs. noncoherent network coding
A network $G$ with $h$ sources $S_1, \ldots, S_h$.

Each source has one message.

$N$ receivers, each one demands all the $h$ messages.

$G$ is a minimal multicast graph.
Select a coding vector for each edge of the network such that the 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 $L(G)$. 

Fragouli, Soljanin 2006
Given a minimal multicast network $G$ with $h$ sources and $N$ receivers construct its line graph $L(G)$.

The vertex in $L(G)$ which is formed from the edge $(u, v)$ of $G$ will be denoted by $uv$.

If the edge $(u, v)$ of $G$ is an input for a receiver $R$ then the vertex $uv$ in $L(G)$ is labelled by $R$. 
Code Design – the Line Graph
Given the line graph \( L(G) \) of a minimal multicast network \( G \), the subtree graph \( \Gamma = (G, S, R) \) is defined by:

Each vertex in \( L(G) \) which is either \( S_iu \) 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 \( L(G) \).
Each vertex in \( L(G) \), which is either \( S_iu \) 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 \( T_1 \) of \( \Gamma \) to a vertex \( T_2 \) of \( \Gamma \) if there is an edge from a vertex of \( T_1 \) to vertex of \( T_2 \) in \( L(G) \).
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 $L(G)$.

The maximal subtrees form a partition of the vertices in $L(G)$. 
Code Design – Line Graph
Code Design – Subtree Graph

![Subtree Graph Diagram]
Code Design – Subtree Graph

Each receiver labels $h$ nodes.

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

The graph has been made minimal – edge removed then multicast is lost.
For a network with 2 sources the 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.

$\Gamma$ has $n - 2$ coding subtrees since two nodes contain the sources.
For a network with 2 sources the 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.

$\Gamma$ has $n - 2$ coding subtrees since two nodes contain the sources.
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.
Code Design – Subtree Graph

\[
\begin{align*}
S_1 R_2 & \rightarrow [1, 0] \\
R_1 R_3 & \rightarrow [1, 1] \\
R_2 R_3 & \rightarrow [0, 1] \text{ or } [1, \beta] \\
S_2 R_1 & \rightarrow [0, 1] \\
\end{align*}
\]

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

Theorem
For networks with 2 sources and \( N \) receivers \( q \geq \left\lfloor \sqrt{2N - 7/4 + 1/2} \right\rfloor \) is sufficient, and for some networks, necessary.

Fragouli, Soljanin 2006

Theorem
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.
Coding for Networks with 2 Sources

Let $V$ be the following set of $q + 1$ vectors
\{\[0, 1\], [1, 0]\} ∪ \{[1, α^i] : 0 ≤ i ≤ q − 2\}, where
$α$ is a primitive element of $𝔽_q$.

Any two vectors in $V$

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

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

\[ \Gamma - \text{the subtree graph with } n \text{ vertices.} \]

\[ \Omega - \text{an undirected graph with the } n \text{ vertices as } \Gamma. \]

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.
Coding for Networks with 2 Sources

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

\[\Gamma\]

\[\Omega\]

\[\text{flow}\]
Coding for Networks with 2 Sources

\[ \begin{aligned}
S_1 & \rightarrow R_2 \\
S_2 & \rightarrow R_1 \\
R_1 & \rightarrow R_2 \\
R_2 & \rightarrow R_3 \\
R_3 & \rightarrow R_2
\end{aligned} \]

\[ \begin{aligned}
\Omega & \rightarrow [1, 0] \\
\Omega & \rightarrow [0, 1]
\end{aligned} \]
Bounds on the Alphabet Size

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 which shares a receiver or a child with each parent.
Lemma

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

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

Each receiver is labelled in at most two nodes. Flow edges exist between vertices with a common child, i.e. between parents of a coding subtree and there are $n - 2$ coding subtrees.
Bounds on the Alphabet Size

Theorem

For networks with 2 sources and $N$ receivers, $q \geq \left\lfloor \sqrt{2N - 7/4} + 1/2 \right\rfloor$ 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.
**Theorem**

For 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.

**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: $E(\Omega) \leq N + n - 2$.

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

- **$h = 2$** messages.
- **$N$** receivers.

1. **$h > 2$** messages.

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

   necessary and sufficient.

2. **$N_{2,r,2}$**

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

   Field size: $q \geq r - 1 \geq \left\lceil \sqrt{2N} \right\rceil$
Introduction to Network Coding, Bounds and Constructions

A SHORT BREAK
Network nodes independently and randomly select linear mappings from inputs links onto outputs links over some field.

In a multicast solvable network with $N$ receivers in which the coefficients for the linear combinations are chosen independently and uniformly over $\mathbb{F}_q$, the success probability that all the $N$ receivers will obtain the information sent by the source node is at least $(1 - N/q)^\eta$ for $q > N$, where $\eta$ is the maximum number of coding points employed by a receiver.
Let \( f(z_1, z_2, ..., z_\eta) \) be a polynomial over \( \mathbb{F}_q \) such that
- \( f(z_1, z_2, ..., z_\eta) \) is not identically zero;
- The degree of a variable in a term of \( f \) is at most \( d \); the total degree of a term is at most \( d\eta' \);
- \( q > d \).

If the values of \( z_1, z_2, ..., z_\eta \) are chosen uniformly at random from \( \mathbb{F}_q \) then
\[
\Pr\{f(z_1, z_2, ..., z_\eta) = 0\} \leq 1 - \left(1 - \frac{d}{q}\right)^{\eta'}
\]
Let \( f(z_1, z_2, ..., z_\eta) \) be a nonzero polynomial over \( \mathbb{F}_q \) such that the sum of degrees of all the variables in a term of \( f(z_1, z_2, ..., z_\eta) \) is at most \( d \). If values \( a_1, a_2, ..., a_\eta \in \mathbb{F}_q \) are chosen uniformly at random from a subset \( A \) of \( \mathbb{F}_q \), then the probability that \( f(a_1, a_2, ..., a_\eta) = 0 \) is at most \( \frac{d}{|A|} \).
The proof is by induction on $\eta$. For $\eta = 1$, $f$ is a polynomial in a single variable of degree at most $d$ and hence it can have at most $d$ zeros and the claim follows.

Suppose that the claim is true for all the polynomials with at most $\eta - 1$ variables, $\eta > 1$. Let $f$ be a polynomial with $\eta$ variables, where the sum of degrees in a term is at most $d$. It can be written as

$$f(z_1, z_2, \ldots, z_\eta) = \sum_{i=0}^{k} z_1^i f_i(z_2, \ldots, z_\eta)$$

where $k \leq d$ is the highest degree of $z_1$ in term of $f$. 

Proof
Random Network Coding

Proof

\[ f(z_1, z_2, \ldots, z_\eta) = \sum_{i=0}^{k} z_1^i f_i(z_2, \ldots, z_\eta) \]

where \( k \leq d \) is the highest degree of \( z_1 \) in term of \( f \). \( f_k(z_2, \ldots, z_\eta) \) is not identically zero and the sum of degrees of its terms is at most \( d - k \). Hence, by the assumption the probability that \( f_k(z_2, \ldots, z_\eta) = 0 \) is at most \( \frac{d-k}{|A|} \). If \( f_k(a_2, \ldots, a_\eta) \neq 0 \) we define

\[ g(z_1) = f(z_1, a_2, \ldots, a_\eta) = \sum_{i=0}^{k} z_1^i f_i(a_2, \ldots, a_\eta). \]

\( g(z_1) \) is a nonzero polynomial of degree \( k \) and hence the probability that \( g(a_1) = 0 \) is at most \( \frac{k}{|A|} \).
Hence, by the assumption the probability that 
\[ f_k(z_2, \ldots, z_\eta) = 0 \] is at most \( \frac{d-k}{|A|} \). If \( f_k(a_2, \ldots, a_\eta) \neq 0 \),

\[ g(z_1) = f(z_1, a_2, \ldots, a_\eta) = \sum_{i=0}^{k} z_1^i f_i(a_2, \ldots, a_\eta). \]

g\(z_1\) is a nonzero polynomial of degree \( k \) and hence the probability that \( g(a_1) = 0 \) is at most \( \frac{k}{|A|} \).

Let \( B \) be the event that \( g(a_1) = f(a_1, a_2, \ldots, a_\eta) = 0 \).

Let \( C \) be the event that \( f_k(a_2, \ldots, a_\eta) = 0 \).

\[ \Pr(C) \leq \frac{d-k}{|A|}, \quad \Pr(B|\overline{C}) \leq \frac{k}{|A|}. \]

\[ \Pr(B) = \Pr(B|C)\Pr(C) + \Pr(B|\overline{C})\Pr(\overline{C}) \]

\[ \leq \Pr(C) + \Pr(B|\overline{C}) \leq \frac{d-k}{|A|} + \frac{k}{|A|} = \frac{d}{|A|}. \]
A multicast network is solvable if and only if the min-cut for each receiver is at least the number of messages.

A network with \( h \) sources, where each one has exactly one message, is solvable if and only if the min-cut to each receiver is at least \( h \).
Multicast Networks

Network nodes independently and randomly select linear mappings from inputs links onto outputs links over some field.

Theorem

In a multicast solvable network with $N$ receivers in which the coefficients for the linear combinations are chosen independently and uniformly over $\mathbb{F}_q$, the success probability that all the $N$ receivers will obtain the information sent by the source node is at least $(1 - \frac{N}{q})^\eta$ for $q > N$, where $\eta$ is the total number of coefficients in the coding points.
Let $f(z_1, z_2, \ldots, z_\eta)$ be a polynomial over $\mathbb{F}_q$ such that

- $f(z_1, z_2, \ldots, z_\eta)$ is not identically zero;
- The degree of a variable in a term of $f$ is at most $d$;
- $q > d$.

If the values of $z_1, z_2, \ldots, z_\eta$ are chosen uniformly at random from $\mathbb{F}_q$ then

$$\Pr\{f(z_1, z_2, \ldots, z_\eta) = 0\} \leq 1 - \left(1 - \frac{d}{q}\right)^\eta$$

Ho, Médard, Koetter, Karger, Effros, Shi, Leong, 2006
Let $f(z_1, z_2, ..., z_\eta)$ be a polynomial over $\mathbb{F}_q$
- $f(z_1, z_2, ..., z_\eta)$ is not identically zero;
- Degree of a variable in term of $f$ is at most $d$;
- $q > d$.

If the values of $z_1, z_2, ..., z_\eta$ are chosen uniformly at random from $\mathbb{F}_q$ then

$$\Pr\{f(z_1, z_2, ..., z_\eta) = 0\} \leq 1 - \left(1 - \frac{d}{q}\right)^\eta$$

**Proof**

The proof is by induction on $\eta$. For $\eta = 1$, $f$ is a polynomial in a single variable of degree at most $d$. An element of $\mathbb{F}_q$ is a root of $f$ with probability at most

$$\frac{d}{q} = 1 - \left(1 - \frac{d}{q}\right)^1$$
Proof

The proof is by induction on $\eta$. For $\eta = 1$, $f$ is a polynomial in a single variable of degree at most $d$. An element of $\mathbb{F}_q$ is a root of $f$ with probability at most

$$d/q = 1 - (1 - d/q)^1$$

For $\eta > 1$, we assume that the claim holds for polynomials with fewer than $\eta$ variables. We express $f$ as

$$f(z_1, \ldots, z_{\eta}) = z_\eta^{d_1} f_1(z_1, \ldots, z_{\eta-1}) + f_2(z_1, \ldots, z_{\eta}),$$

where $d_1 \leq d$ and $f_1$ is not identically zero polynomial.

$$\Pr(f = 0) = \Pr(f_1 = 0) \cdot \Pr(f = 0|f_1 = 0) + \Pr(f_1 \neq 0) \cdot \Pr(f = 0|f_1 \neq 0)$$
Random Network Coding

**Proof**

For $\eta > 1$, we assume that the claim holds for polynomials with fewer than $\eta$ variables. We express $f$ as

$$f(z_1, \ldots, z_\eta) = z_\eta^{d_1} f_1(z_1, \ldots, z_{\eta-1}) + f_2(z_1, \ldots, z_\eta),$$

where $d_1 \leq d$ and $f_1$ is not identically zero polynomial.

$$\Pr(f = 0) = \Pr(f_1 = 0) \cdot \Pr(f = 0|f_1 = 0) + \Pr(f_1 \neq 0) \cdot \Pr(f = 0|f_1 \neq 0)$$

1. $\Pr(f_1 = 0) \leq 1 - (1 - d/q)^{\eta-1}$ by the induction hypothesis.
2. $\Pr(f = 0|f_1 = 0) \leq 1$.
3. $\Pr(f = 0|f_1 \neq 0) \leq d/q$, as a polynomial in $z_\eta$. 
Random Network Coding

Proof

\[ f(z_1, \ldots, z_\eta) = z_\eta^{d_1} f_1(z_1, \ldots, z_{\eta-1}) + f_2(z_1, \ldots, z_\eta), \]
where \( d_1 \leq d \) and \( f_1 \) is not identically zero polynomial.

\[ \Pr(f = 0) = \Pr(f_1 = 0) \cdot \Pr(f = 0|f_1 = 0) \]
\[ + \Pr(f_1 \neq 0) \cdot \Pr(f = 0|f_1 \neq 0) \]

1. \( \Pr(f_1 = 0) \leq 1 - (1 - \frac{d}{q})^{\eta-1} \) by the induction hypothesis.
2. \( \Pr(f = 0|f_1 = 0) \leq 1. \)
3. \( \Pr(f = 0|f_1 \neq 0) \leq \frac{d}{q}, \) as a polynomial in \( z_\eta. \)

\[ \Pr(f = 0) \leq \Pr(f_1 = 0) + (1 - \Pr(f_1 = 0)) \cdot \frac{d}{q} \] by 2, 3
\[ = \Pr(f_1 = 0)(1 - \frac{d}{q}) + \frac{d}{q} \]
\[ \leq (1 - (1 - \frac{d}{q})^{\eta-1})(1 - \frac{d}{q}) + \frac{d}{q} \] by 1
\[ = 1 - (1 - \frac{d}{q})^{\eta} \]
Models for Network Coding

Coherent network coding - the source and the destinations nodes know the topology of the network and the network code.

Noncoherent network coding - the source and the destinations nodes don’t know the topology of the network and the network code.
The source sends $h$ messages $X = (x_1, x_2, \ldots, x_h)^t$. A receiver obtains the message $Y = (y_1, y_2, \ldots, y_h)^t$, $Y = A \cdot X$, where $A$ is a $h \times h$ transfer matrix.

Now assume that $x_i$ is a message (packet) of length $N$ over $\mathbb{F}_q$. Hence, $X, Y$ are $h \times N$ matrices over $\mathbb{F}_q$ and the transfer matrix $A$ is a $h \times h$ matrix. The receiver obtains $Y = A \cdot X$. 

Coherent Network Coding

one source, one receiver
Research Problems

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?
Research Problems

Improve the lower bound on the alphabet size for solvable multicast networks with $h$ messages?

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