**COMPSCI 650 Applied Information Theory** April 14, 2016

# Lecture 21

*Instructor: Arya Mazumdar* *Scribe:* [redacted]

## 1 Review

In the network coding problems, we have network such as butterfly network (Figure 1.a), which was mentioned in last class. The network can be directed or undirected. Each edge in the network can carry one package one time, and there is no delay for the transmission. In butterfly network, we have 2 sources ($S_1$ and $S_2$) and 2 receivers($R_1$ and $R_2$), which is multiple-cast model. The receiver can only be one. Or there is no receiver and the message is sent to every node in the network, which is Broadcast model. In butterfly network, if both $x_1$ and $x_2$ is sent by the common edge, there must be a queue for them to send the message in order. Instead, we can send $x_1 + x_2$, and the receivers can figure out the message is $x_1$ or $x_2$ by linear equations.

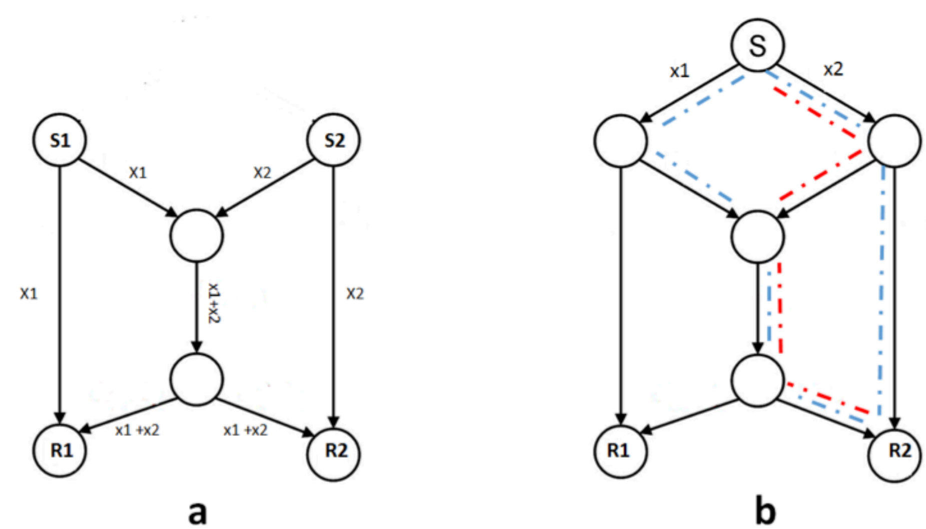


**Figure 1**: Networks for Network Coding

In the Max flow- min cut problem, there are two sets: one set is the sources and another is the receivers. And the multiple-source network can be converted to single-source network. There are many cuts between the source and a receiver. A min cut is a cut of minimum size of all possible cuts, which cause the receiver disconnected from the source. For example, in Figure 1 b, the min cut of $R_2$ is 2, although there are 3 transmission paths from the source to $R_2$.

During each time-slot, the maximum number of messages that you can send from source to a receiver is equal to the min cut of this receiver: <max flow>=<min cut>. That is because: if $|min - cut| =$ h, then there exist only h edge-disjoint paths from the sources to the receiver. If message number is large than h, only h messages can be transmitted through h edge-disjoint paths at the same time.

## 2 The Main Network Coding Theorem

### 2.1 Statement of the main theorem

**Theorem 1** *Consider a directed acyclic graph $G = (V,E)$ with unit capacity edges, h unit rate sources located on the same vertex of the graph and M receivers. Assume that the value of the min-cut to each receiver is h. Then there exists a multicast transmission scheme over a large enough finite field $\mathbb{F}_q$, in which intermediate network nodes linearly combine their incoming information symbols over $\mathbb{F}_q$, that delivers the information from the sources simultaneously to each receiver at a rate equal to h.*

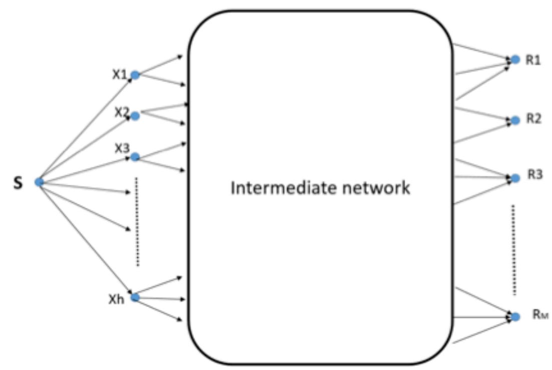


**Figure 2**: linear network

In the case of general network (Figure 2), we assume that there is no delay for data transmission, which means during each time slot all nodes simultaneously receive all their inputs and send their outputs. In the general network, the source S sends h signals to different vertexes $(x_1, x_2, \cdots, x_h)$. Then the signals are transmitted through a intermediate network. And there are M receivers $(R_1, R_2, \cdots, R_M)$. When multiple receivers are using the network simultaneously, their sets of paths may overlap. The intermediate network nodes combine their incoming information flows, so you can get the information of each receiver at the same rate simultaneously in any arbitrary network if you use network coding. Such transmission schemes can be referred as linear network coding.

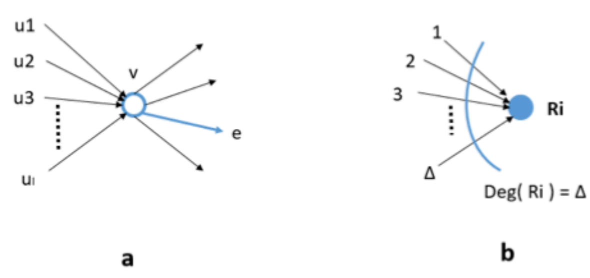


**Figure 3**: intermediate node(a) and receiver $R_i$ in linear network

Suppose there is a intermediate vertex v in the intermediate network(Figure 3a). There are l incoming edges for v $(u_1, u_2, \cdots, u_l)$, and how can we decide what information is sent in each out-coming edge e of v?

The theorem additionally claims that it is sufficient for intermediate nodes to perform linear operations over a finite field $\mathbb{F}_q$. Define the local coding vector with dimension $1 \times l$:

$$C_e = [\alpha_1, \alpha_2, \cdots, \alpha_l]$$

The local coding vector $C_e$ associated with an edge e is the vector of coefficients over $\mathbb{F}_q$ with which we multiply the incoming information to edge e. And the information for out-coming edge e of intermediate node v is the linear combination of all the in-coming information. So for every out-coming edge e, the information is:

$$< C_e, u >= \sum_{i=1}^{l} \alpha_i u_i \tag{1}$$

Assume each receiver in the network has the same min cut, which is:

$$(S, R_i) = h \quad i \in [1, M]$$

Suppose the degree of Receiver $R_i$ is $\Delta$ ($\Delta \geq h$)(Figure 3b). The linear combining can enable the receivers to get the information at rate h from the source. Each incoming edge of a receiver contains linear combination of the source information $(x_1, x_2, \cdots, x_h)$. If we can find all the local coding vectors for all the edges in the network, we can find the global coding vectors. Define the global coding vector for receiver $R_i$ is :

$$C_i^{(e)} = [\sigma_1^{(e)}, \sigma_2^{(e)}, \cdots, \sigma_h^{(e)}] \quad e \in [1, \Delta]$$

The global coding vector $C_i^{(e)}$ associated with an edge e is the vector of coefficients of the source symbols that flow (linearly combined) through edge e to a receiver. The dimension of $C_i^{(e)}$ is $1 \times h$.

The global coding vectors associated with the input edges of a receiver node define a system of linear equations that the receiver can solve to determine the source information. Thus $R_i$ has to solve the following linear equations ($R_i = A_i \times x$):

$$\underbrace{\begin{bmatrix} \sigma_1^{(1)} & \sigma_2^{(1)} & \cdots & \sigma_h^{(1)} \\ \sigma_1^{(2)} & \sigma_2^{(2)} & \cdots & \sigma_h^{(2)} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_1^{(\Delta)} & \sigma_2^{(\Delta)} & \cdots & \sigma_h^{(\Delta)} \end{bmatrix}}_{A_i} \times \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix}}_{x} \tag{2}$$

$A_i$ is the matrix whose ith row is the global coding vector of $R_i$ on the path of (S, $R_i$), and x is the source signal. We know $A_i$ and $R_i$, and we want to find what the source signal x is. To retrieve the source information x, The receiver $R_i$ has to be transmitted from the h sources. Therefore, choosing global coding vectors so that all $A_i$ are full rank will enable all receivers to recover the source symbols from the information they receive.Rank($A_i$) = h for all $A_i$. We can assume that $A_i$ is $h \times h$ dimensions, so that the source information x can be retrieved. If $\Delta > h$, it means there are $\Delta - h$ rows in $A_i$ in not linear independent and they are redundant.

## 2.2 Algebraic Statement of the Main Theorem

*In linear network coding, there exist values in some large enough finite field $\mathbb{F}_q$ for the components $\{\alpha_k\}$ of the local coding vectors, such that all matrices $A_i$, $1 \leq i \leq M$, defining the information that the receivers observe, are full rank.*

The entries of $A_i$ are functions of all local coding factors $\{\alpha_k\}$ of interval edges in the network. Since there are a lot of interval edges in the network, k is a large. And if we have to show that each $A_i$ is "invertible", which means we have to prove that:

$$det(A_i) \neq 0 \quad \forall i \in [1, M] \tag{3}$$

which means:

$$f(\{\alpha_k\}) = det(A_1)det(A_2)\cdots det(A_M) \neq 0 \tag{4}$$

Thus, we can choose $(\alpha_1, \alpha_2, \cdots, \alpha_E)$(E is the number of edges) to make sure $f(\{\alpha_k\})$ is not equal to 0. That means we need to show that there exists an assignment of coefficient: $z_1 = \alpha_1, z_2 = \alpha_2, \cdots, z_E = \alpha_E$ which makes $f(\alpha_1, \alpha_2, \cdots, \alpha_E) \neq 0$.

Because we only perform linear operations at intermediate nodes of the network, each entry of matrix $A_i$ is a polynomial in $\{\alpha_k\}$.When dealing with polynomials over a finite field, we have to keep in mind that even a polynomial not identically equal to zero over that field can evaluate to zero on all of its elements. For example,the polynomial $det(A) = x(x+1)$ over $\{\alpha_k\}$. The determinants are always equal to 0, since in binary field, 0(0+1)=0 and 1(1+1)=0.

Actually, to design the assignment to satisfy (3) and (4) is difficult and remain an open problem in network coding. But it has been shown that take a large field $\mathbb{F}$ and choose the coefficient randomly, (3) or (4) can be reached. The Sparse Zeros Lemma tells us that we can find values for the coefficients $\{\alpha_k\}$ over a large enough finite field such that the condition (3) or (4) holds, and thus concludes our proof.

## 2.3 Sparse Zero Lemma

*Suppose the degree of any variable in the polynomial $f(z_1, z_2, \cdots, z_E)$ is less then or equal to d, then for any $\mathbb{F}_q$ $(q > d)$, there exists an assignment of variables $z_1 = \alpha_1, z_2 = \alpha_2, \cdots, z_E = \alpha_E \in \mathbb{F}_q$, such that $f(\alpha_1, \alpha_2, \cdots, \alpha_E) \neq 0$.*

**Proof** We prove the lemma by induction in the number of variables k:

1. Basic step: k = 1. We have a polynomial in a single variable of degree at most d. The lemma holds immediately since such polynomials can have at most d roots.

2. $k > 1$. The inductive hypothesis is that the claim holds for all polynomials with no more than k variables. We can express $f(z_1, z_2, \cdots, z_k, z_{k+1})$ as :

$$f_i(z_1, z_2, \cdots, z_k, z_{k+1}) = \sum_{i=1}^{d} f_i(z_1, z-2, \cdots, z_k) z_{k+1}^i \tag{5}$$

$$= \sum_{i=1}^{d} f_i(\alpha_1, \alpha_2, \cdots, \alpha_k) z_{k+1}^i \tag{6}$$

Since the number of variables in $f_i$ is no more than k, by the inductive hypothesis, $f_i(\alpha_1, \alpha_2, \cdots, \alpha_k) \neq 0$. Thus we can choose $z_{k+1} \neq 0$ so that make (6) not equal to 0.

∎

# 3 Introduction to Compressed sensing

## 3.1 Single Pixel Camera

### 3.1.1 Hardware design*

(*All the context in this section is from the COMPRESSIVE SINGLEPIXEL IMAGING Technical report, Andrew Thompson, University of Edinburgh,20 January 2011).

In contrast to a conventional camera which would have a vast array of photo detectors one for each pixel – a singlepixel camera is socalled because it has only a single photon detector or "pixel". The incident light field is directed onto a specialized type of Spatial Light Modulator (SLM) known as a Digital Micromirror Device (DMD) which consists of an array of tiny mirrors, one corresponding to each pixel. Each mirror can be oriented in one of two directions: the "on" position directs the light for that pixel towards the single detector, while the "of" position directs the light away from the detector. The light from all the pixels set to the "on" position is then summed and a measurement is recorded as an output voltage on the photon detector. The measurement is then subsequently digitized by an A/D convertor. A series of measurements can be obtained by flipping the mirrors and repeating the process a number of times. Clearly, the result of such a procedure is encoded measurements, and the other issue to address is how to decode these measurements and reconstruct the incident image.

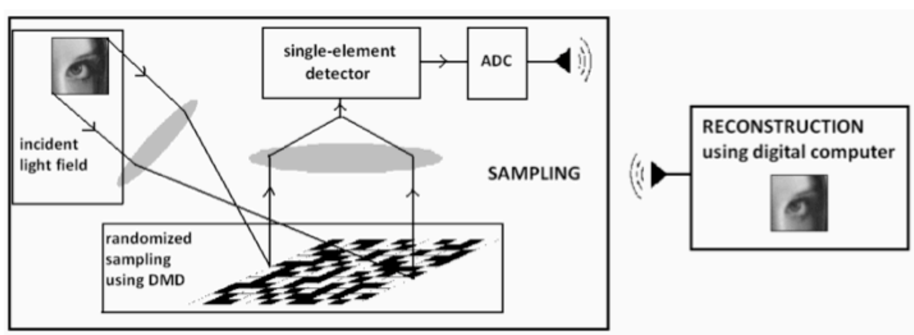


**Figure 4**: The Rice CS single pixel camera

The concept of a single pixel camera is not a new one, and it fits into the broad category of multiplexing imaging methods in which a series of consecutive measurements are made by a single detector. What sets the CS single pixel camera model apart is a novel sampling approach which means that it is possible to take fewer measurements. It therefore offers an alternative approach to obtaining compressed images: rather than taking a full set of samples and subsequently compressing, compressed samples are acquired in the first place. To achieve this, CS theory motivates the use of a random sampling procedure in which each mirror is set randomly to either the "on" or "off" position with equal probability. Figure 4 shows a diagram of the proposed camera design.

### 3.1.2 Models of Computation

Suppose then that we have an unknown discretized signal which we model as a vector $x \in R^N$, and suppose that we obtain linear measurements of x , which we can view as inner products of x with a series of "test function" vectors $\{\phi\}$. we may represent the sampling process as the matrix equation as:

$$y = \Phi x$$

where y is the vector of samples, and $\Phi$ is sampling m $\times$ N ($N >> m$) matrix whose rows are the test functions $\{\phi\}$.