Outstanding Reading

- Read all chapters assigned from IT-I (EE514, Winter 2012).
- Read chapter 8 in the book.
- Read chapter 9 in the book.
- Read chapter 10 in the book (chapter on rate distortion theory).
- Read chapter 14 in the book (Kolmogorov complexity).
- Read chapter 13, section on Lempel Ziv compression, in the book.
- Read chapter 15 in C&T.
- Read chapters 4, 5, 9, 10, 15 in El Gamal & Kim.
Please do use our discussion board (https://catalyst.uw.edu/gopost/board/bilmes/27386/) for all questions, comments, so that all will benefit from them being answered.

On Final Presentations

- Your task is to give a 15-20 minute presentation that summarizes 2-3 related and significant papers that come from IEEE Transactions on Information Theory (or a very related area).
- The papers must not be ones that we covered in class, although they can be related.
- You need to do the research to find the papers yourself (i.e., that is part of the assignment).
- The papers must have been published in the last 10 years (so no old or classic papers).
- Your grade will be based on how clear, understandable, and accurate your presentation is.
- This is a real challenge and will require significant work! Many of the papers are complex. To get a good grade, you will need to work very hard to present very complex ideas in an extremely simple yet still precise way.
- Again, don’t expect this to be easy, you might need to try a few topics until you find one that is suitable.
Final Presentation Milestones

All submissions done in PDF file format via our dropbox (https://catalyst.uw.edu/collectit/dropbox/bilmes/21171)

- Final presentations: Monday, June 4th from 2:00pm - 5:00pm in EEB-417 (my lab).
- What to turn in by Monday 1:00pm:
  1. your slides from your presentation
  2. a short at most 4 page summary of the papers
  3. Turn in electronically in PDF format via our dropbox (https://catalyst.uw.edu/collectit/dropbox/bilmes/21171)

General Network Information Theory

- Very important part of modern IT (still currently being actively researched).
- The most general case (first). We have an arbitrary network:

```
X_1, Y_1   X_2, Y_2
  /
X_m, Y_m   X_3, Y_3
  /
  /
  /
X_4, Y_4
```

- Each sender $X_i$ is trying to communicate simultaneously with each receiver $Y_i$ (i.e., for all $i$, $X_i$ is sending to $\{Y_i\}_i$)
- The $X_i$ are not necessarily independent.
Broadcast Channel (BC)

- One sender, many receivers
- Sender might have separate message for each receiver, or might have a mix of common messages (for all, or some receivers) and a unique message to each receiver.
- We do not consider privacy considerations here in this discussion (although it is important, sorry).
- Our goal: understand the limits of achievability of rates so that there exists vanishingly small chance of error as $n \to \infty$.
- Ex: cell-phone towers: separate message to each cell phone over essentially the same channel (lots of interference potential between messages).
- Ex: classroom lecture to students (plural). Goal is to provide information at an appropriate rate to both good (and bad (but of course there is no such thing 😊)) students. Actually, it’s all the fault of the professor and poorly prepared lectures! 😊
Some definitions

- One sender, two receivers, sender has messages for both receivers, simultaneous broadcast.
- The information for each receiver is independent.
- Channel: $(X, Y_1, Y_2), p(y_1, y_2|x)$
- Memoryless: $\Pr(y^{1:n}_1, y^{2:n}_1|x^{1:n}) = \prod_{i=1}^{n} p(y^1_i, y^2_i|x_i)$.
- Code sequence: $((2^{nR_1}, 2^{nR_2}), n)$ code for Broadcast channel consists of encoder:
  \[ X : \{1, 2, \ldots, 2^{nR_1}\} \times \{1, 2, \ldots, 2^{nR_2}\} \to X^n \]  
  (1)
  and two decoders:
  \[ g_1 : Y^{1:n}_1 \to \{1, \ldots, 2^{nR_1}\}, \quad g_2 : Y^{2:n}_1 \to \{1, \ldots, 2^{nR_2}\} \]  
  (2)
- Probability of error:
  \[ P_e^{(n)} = \Pr [(g_1(Y^{1:n}_1) \neq W_1) \lor (g_2(Y^{2:n}_1) \neq W_2)] \]  
  (3)
  where $W_1, W_2$ are both uniformly distributed (and thus independent).

More Definitions

- Achievable: A rate pair $(R_1, R_2)$ is said to be achievable for BC if there exists sequence of $((2^{nR_1}, 2^{nR_2}), n)$ codes with $P_e^{(n)} \to 0$.
- With common information to be sent to both receivers, i.e., $W_0$ to be received by both, while $W_i$ to be received by receiver $i$ ($i = 1, 2$):
- Code sequence: $((2^{nR_0}, 2^{nR_1}, 2^{nR_2}), n)$ code for Broadcast channel with common information consists of encoder:
  \[ X : \{2^{nR_0}\} \times \{2^{nR_1}\} \times \{2^{nR_2}\} \to X^n \]  
  and two decoders:
  \[ g_1 : Y^{1:n}_1 \to \{2^{nR_0}\} \times \{2^{nR_1}\}, \quad g_2 : Y^{2:n}_1 \to \{2^{nR_0}\} \times \{2^{nR_2}\} \]  
  (4)
- Probability of error:
  \[ P_e^{(n)} = \Pr [(g_1(Y^{1:n}_1) \neq (W_0, W_1)) \lor (g_2(Y^{2:n}_1) \neq (W_0, W_2))] \]  
  (5)
  where $W_0, W_1, W_2$ are both uniformly distributed (and thus independent).
More Definitions

- **Achievable**: A rate pair \((R_1, R_2)\) is said to be achievable for BC if there exists sequence of \(((2^nR_1, 2^nR_2), n)\) codes with \(P_e^{(n)} \to 0\).

- **Achievable**: A rate triple \((R_0, R_1, R_2)\) is said to be achievable for BC in a similar way.

- **Capacity region** is the closure of all achievable rates.

**On Capacity Region**

**Theorem 2.1**

The capacity region of achievable rates in a broadcast channel depends only on \(p(y_1|x)\) and \(p(y_2|x)\).

**Proof sketch.**

- \(Y_1\) (resp. \(Y_2\)) error depends only on \(p(x, y_1)\) (resp. \(p(x, y_2)\)) not on \(p(x, y_1, y_2)\).

- \(A \triangleq (g_1(Y_{1:n}^1) \neq W_1) \lor (g_2(Y_{1:n}^2) \neq W_2)\), \(B_1 \triangleq (g_1(Y_{1:n}^1) \neq W_1)\), \(B_2 \triangleq (g_2(Y_{1:n}^2) \neq W_2)\).

- Then \(B_1 \Rightarrow A\) and \(B_2 \Rightarrow A\), so \(p(B_1) \leq p(A)\) and \(p(B_2) \leq p(A)\).

- Hence \(\max \{p(B_1), p(B_2)\} \leq p(A)\).

- \(\max \{P_1^1, P_2^2\} \leq P_1^{1+2} \leq P_1^1 + P_2^2\), so while \(1 + 2\) error depends on the full joint \(p(x, y_1, y_2)\), to drive down the \(1 + 2\) error, it is necessary and sufficient to drive down both \(P_1^1, P_2^2\) individually.

- this essentially completes the proof.
Simple Bounds on Capacity

- Given a BC $p(y_1, y_2|x)$, define $C_j = \max_{p(x)} I(X; Y_j)$ for $j = 1, 2$, capacities of DMCs $p(y_1|x)$ and $p(y_2|x)$.
- These give us a time-sharing “inner bound” of the capacity region of a BC.
- On the other hand, $R_1 + R_2 \leq C_{12} = \max_{p(x)} I(X; Y_1, Y_2)$ gives us outer bound.
- Neither bound is tight in general.

![Diagram showing Simple Bounds on Capacity]

Superposition Coding Inner Bound

- As mentioned, one receiver may be cleaner (or stronger) than other.
- The weaker receiver recovers its intended message, and the stronger receiver recovers its intended message using successive cancellation decoding (as we saw for the MAC channel).
- As an example, consider the following pair of BSCs. I.e., BSC($p_1$) and BSC($p_2$) where w.l.o.g., $p_1 < p_2 < 1/2$.
- This can be seen as a BC as in the following, inner bound on right:

\[
Z_1 \sim \text{Bernoulli}(p_1)
\]

\[
X \rightarrow Z_2 \sim \text{Bernoulli}(p_2)
\]

\[
Y_1 = X + Z_1
\]

\[
Y_2 = X + Z_2
\]

![Diagram showing Superposition Coding Inner Bound]
Superposition Coding Inner Bound

- As seen above, time division can achieve the inner bound on the above right figure (just time-share the channel at any proportions), but this is non-ideal. Can we do better?
- Superposition coding allows us to do better and it uses auxiliary random variables.
- Let $U \sim \text{Bernoulli}(1/2)$ and for $\alpha \in [0, 1/2]$ let $V \sim \text{Bernoulli}(\alpha)$ be independent r.v.s.
- Let $X = U + V \mod 2$ (i.e., $0 + 0 = 0$, $0 + 1 = 1$, $1 + 1 = 0$), and define $\oplus$ as componentwise mod 2 addition of vectors or scalars (thus $X = U \oplus V$).
- Choose rates $R_1$ and $R_2$.
- Code generation: randomly generate $|W_2| = 2^{nR_2}$ sequences of the form $u_{1:n}(w_2)$ for each $w_2 \in [2^{nR_2}]$, based on Bernoulli($1/2$). These will act as “cloud centers” as we will see.

Further code generation: randomly generate $2^{nR_1}$ sequences $v_{1:n}(w_1)$ for each $w_1 \in [2^{nR_1}]$, here generated Bernoulli($\alpha$).

To broadcast message $(w_1, w_2)$ the sender sends $x_{1:n}(w_1, w_2) \triangleq u_{1:n}(w_2) \oplus v_{1:n}(w_1)$.

Receiver 2: needs to recover $w_2$, and has access to $y_{1:n}^2 = u_{1:n}(w_2) \oplus v_{1:n}(w_1) \oplus z_{1:n}^2 = u_{1:n}(w_2) \oplus (v_{1:n}(w_1) \oplus z_{1:n}^2)$ where $z_{1:n}^2$ is the BSC channel noise, and moreover via the second grouping, we can consider $(v_{1:n}(w_1) \oplus z_{1:n}^2)$ as greater noise.

Thus, for receiver 2, message destined for receiver 1 is seen as noise.

$P_{e_2}^{(n)}$, probability of error for receiver 2 $\rightarrow 0$ as $n \rightarrow \infty$ as long as $R_2 < 1 - H(\alpha \ast p_2)$ where $\alpha \ast p_2 \triangleq \alpha(1 - p_2) + (1 - \alpha)p_2$ is the noise rate of the “effective” BSC between sender and receiver 2.
Superposition Coding Inner Bound

- Now, for receiver 1 (the cleaner one), we use successive cancellation decoding (as we did for the MAC).
- Receiver 1: needs to recover $w_1$, and has access to $y_{1:n} = u_{1:n}(w_2) \oplus v_{1:n}(w_1) \oplus z_{1:n} = u_{1:n}(w_2) \oplus (v_{1:n}(w_1) \oplus z_{1:n})$
  where $z_{1:n}$ is the BSC channel noise for receiver 1.
- We can use this to recover $w_2$ treating $v_{1:n}(w_1)$ as part of the noise (but presumably receiver 1’s rate can handle this).
- Given $w_2$ we have access to $u_{1:n}(w_2)$ and can form:
  $$\hat{y}_{1:n} = u_{1:n}(w_2) \oplus v_{1:n}(w_1) \oplus z_{1:n} = v_{1:n}(w_1) \oplus z_{1:n}$$
  which is a cleaner signal for receiver 1.
- We then decode $\hat{y}_{1:n}$ to recover $w_1$.
- Probability of error $\to 0$ as long as
  $$R_1 \leq H(\alpha \ast p_1) - H(p_1)$$
  $$R_2 \leq 1 - H(\alpha \ast p_2)$$
  for some $\alpha \in [0, 1/2]$.
- This bound is larger than time-sharing.

This leads to a better inner bound for $(R_1, R_2)$ of the form

$$R_1 \leq H(\alpha \ast p_1) - H(p_1)$$  \hspace{1cm} (6)

$$R_2 \leq 1 - H(\alpha \ast p_2)$$  \hspace{1cm} (7)

for some $\alpha \in [0, 1/2]$.
We can view this as clouds, we have $2^{nR_2}$ “clouds” and within each cloud lies $2^{nR_1}$ codewords. We use $U$ (more precisely $u_{1:n}(w_2)$) to indicate cloud id.

- Total rate is $R_1 + R_2$
- Receiver 2 can distinguish only clouds, not codewords within clouds.
- Receiver 1 can distinguish both cloud id and codeword within cloud, and does so by first identifying the cloud id, and then the codeword within the cloud.

In previous figure, each cloud has $2^{nR_1}$ codewords.

- Also, there are $2^{nR_2}$ clouds.
- Thus, a total of $2^n(R_1 + R_2)$ codewords.
- Cloud radius $\approx n\alpha$

There is a tradeoff: We can fit more clouds in the large circle (increasing the rate $R_2$) but then the cloud radius would need to decrease (decreasing the rate $R_1$).

I.e., there exists tradeoff between size of clouds and number of possible clouds.
Superposition Coding Inner Bound Theorem

**Theorem 3.1**

A rate pair \((R_1, R_2)\) is achievable for the discrete memoryless BC \(p(y_1, y_2|x)\) if:

\[
R_1 < I(X; Y_1|U) \quad (8)
\]
\[
R_2 < I(U; Y_2) \quad (9)
\]
\[
R_1 + R_2 < I(X; Y_1) \quad (10)
\]

for some pmf \(p(u, x)\).

- Note: \(U\) is an auxiliary random variable that does not correspond to the channel variables.
- Unlike time sharing variable before (which helped us to mix different rates), here \(U\) is more important in establishing the capacity region.
- Also, note that we establish this bound for joint \(p(u, x)p(y_1, y_2|x)\), i.e., \(U \to X \to (Y_1, Y_2)\).

**Proof of Theorem**

**proof of Theorem 3.1.**

- We first prove achievability (i.e., if rates satisfy above equations, \(\exists\) code that has vanishing error probability as \(n \to \infty\)).
- Fix a pmf \(p(u, x)\) and randomly generate \(2^{nR_2}\) “cloud centers” \(u_{1:n}(w_2)\). For each cloud center \(u_{1:n}(w_2)\), generate \(2^{nR_1}\) “satellite” codewords \(x_{1:n}(w_1, w_2)\).
- Receiver 2 decodes cloud center \(u_{1:n}(w_2)\) and receiver 1 decodes satellite codeword.

![Diagram showing cloud centers and satellite codewords](image-url)
Proof of Theorem

proof of Theorem 3.1.

- Codebook generation: Fix \( p(u)p(x|u) \), randomly generate \( 2^{nR_2} \) sequences \( u_1:n(w_2) \) for \( w_2 \in [2^{nR_2}] \) each according to \( \prod_{i=1}^{n} p(u_i) \). Similarly, for each \( w_2 \), generate \( 2^{nR_1} \) sequences according to \( \prod_{i=1}^{n} p(x_i|u_i(w_2)) \).
- Encoding: to send \((w_1, w_2)\), transmit codeword \( x_{1:n}(w_1, w_2) \).
- Decoder 2 claims \( \hat{w}_2 \) if \((u_{1:n}(\hat{w}_2), y_{1:n}^{2}) \in A_{\epsilon}^{(n)} \) otherwise error.
- Decoder 1 claims \( \hat{w}_1 \) if \((u_{1:n}(w_2), x_{1:n}(\hat{w}_1, w_2), y_{1:n}^{1}) \in A_{\epsilon}^{(n)} \) for some \( w_2 \), otherwise error.
- Symmetry again allows us to assume \((w_1, w_2) = (1, 1)\)
- Error for decoder 2 decompose of:

\[
E_{21} = \left\{ (U_{1:n}(w), Y_{1:n}^{2}) \notin A_{\epsilon}^{(n)} \right\} 
\]

\[
E_{22} = \left\{ (U_{1:n}(w_2), Y_{1:n}^{2}) \in A_{\epsilon}^{(n)} \text{ for some } w_2 \neq 1 \right\} 
\] (12)

Thus, decoder error is upper bounded by

\[
\Pr(E_2) \leq \Pr(E_{21}) + \Pr(E_{22}) 
\] (13)

- AEP \( \Pr(E_{21}) \to 0 \) as \( n \to 0 \), and by
- similar argument to what we’ve seen before, \( \Pr(E_{22}) \to 0 \) as \( n \to \infty \) if \( R_2 < I(U; Y_2) - \epsilon \)
- Next, error for decoder 1. Three types of error:

\[
E_{11} = \left\{ (U_{1:n}(1), X_{1:n}(1, 1), Y_{1:n}^{1}) \notin A_{\epsilon}^{(n)} \right\} 
\] (14)

\[
E_{12} = \left\{ (U_{1:n}(1), X_{1:n}(w_1, 1), Y_{1:n}^{1}) \in A_{\epsilon}^{(n)} \text{ for some } w_1 \neq 1 \right\} 
\] (15)

\[
E_{13} = \left\{ (U_{1:n}(w_2), X_{1:n}(w_1, w_2), Y_{1:n}^{1}) \in A_{\epsilon}^{(n)} \
\text{ for some } (w_1, w_2) \neq (1, 1) \right\} 
\] (16)
Proof of Theorem

Proof of Theorem 3.1.

- Again, $\Pr(E_1) \leq \Pr(E_{11}) + \Pr(E_{12}) + \Pr(E_{13})$,
- the first $\rightarrow 0$,
- the second $\rightarrow 0$ if $R_1 < I(X; Y_1|U) - \epsilon$,
- and the third $\rightarrow 0$ if $R_1 + R_2 < I(U, X; Y_1) - \epsilon = I(X; Y) - \epsilon$
  (recall $U \rightarrow X \rightarrow Y_1$ is a Markov chain).

Discussion

- Note that we can switch receivers to get the following form of bound as well:

  \begin{align*}
  R_1 &< I(U; Y_1) \quad (17) \\
  R_2 &< I(X; Y_2|U) \quad (18) \\
  R_1 + R_2 &< I(X; Y_2) \quad (19)
  \end{align*}

- Superposition decoding is optimal for several classes of broadcast channel where one receiver is stronger than the other (e.g., Degraded broadcast channel that we briefly saw last time and that we visit again below),
- but there are cases where asking one receiver to decode both messages can strain the overall rates (see Chapter 8 in Gamal & Kim for example).
Degraded Broadcast channel

- Recall, no privacy, so $Y_1$ and $Y_2$ are both receiving all messages (i.e., the messages intended for each of $Y_1$ and $Y_2$).
- In general, one of $Y_1$ and $Y_2$ will be noisy, assume w.l.o.g. that $Y_1$ is no noisier than $Y_2$.
- Since $Y_1$ (being cleaner) has all the information about $Y_2$ that $X$ has (and is communicable to $Y_2$), we can treat this as a relay, easier to work with mathematically.
- I.e., since $Y_1$ is cleaner, can think of this as Markov chain: $X \rightarrow Y_1 \rightarrow Y_2$
- Degraded broadcast channel. A BC is physically degraded if:
  \[
p(y_1, y_2|x) = p(y_1|x)p(y_2|y_1)
  \]  
  (20)
  and stochastically degraded if $\exists p'(y_2|y_1)$ such that
  \[
p(y_2|x) = \sum_{y_1} p(y_1|x)p'(y_2|y_1)
  \]  
  (21)
  in either case, same capacity since $I(X; Y_1, Y_2) = I(X; Y_1)$.

Theorem 3.2

The capacity region for sending independent information over the degraded broadcast channel $X \rightarrow Y_1 \rightarrow Y_2$ is the convex hull of the closure of all $(R_1, R_2)$ satisfying:

\[
R_2 \leq I(U; Y_2)
\]  
(22)

\[
R_1 \leq I(X; Y_1|U)
\]  
(23)

\[
R_1 + R_2 \leq I(X; Y_1)
\]  
(24)

for some joint distribution $p(u)p(x|u)p(y_1, y_2|x)$, where the auxiliary variable $U$ has cardinality bounded by $|U| \leq \min \{|X|, |Y_1|, |Y_2|\}$. 
Degraded Broadcast channel

- Note, $U$ is a “convenience” variable to show how the rates couple.
- Also, the third equation $R_1 + R_2 \leq I(X; Y_1)$ is vacuous since from the first two, we can get:

$$R_1 + R_2 \leq I(X; Y_1 | U) + I(U; Y_2)$$

$$\leq I(X; Y_1 | U) + I(U; Y_1) \text{ since } U \rightarrow X \rightarrow Y_1 \rightarrow Y_2$$

$$= I(Y_1; X, U)$$

$$= I(Y_1; X) + I(Y_1; U | X)$$

$$= I(Y_1; X)$$

- Thus, the first two constitute a tighter upper bound which is traded off by the $U$ random variable and how $U$ relates to $X$ via $p(x | u)$.
- Since we already have achievability for Superposition decoding, we immediately have achievability for degraded broadcast channel as well.

Degraded Channel: Converse

- Goal: A given sequence $(2^n R_1, 2^n R_2, n)$ codes with $P_e^{(n)} \rightarrow 0$ must have $R_1 \leq I(X; Y_1 | U)$ and $R_2 \leq I(U; Y_2)$ for some $p(u, x)$ such that $U \rightarrow X \rightarrow (Y_1, Y_2)$ is a Markov chain.
- Key is to be able to identify $U$, the auxiliary variable.
- Every $(2^n R_1, 2^n R_2, n)$ code induces joint distribution of form:

$$p(w_1, w_2, x_{1:n}, y_{1:n}^1, y_{1:n}^2) = 2^{-n(R_1 + R_2)} p(x_{1:n} | w_1, w_2) \prod_{i=1}^n p(y_i^1, y_i^2 | x_i)$$

- Fano gives us:

$$H(W_1 | Y_{1:n}^1) \leq n R_1 P_e^{(n)} + 1 \leq n \epsilon_n, \quad \text{(30)}$$

$$H(W_2 | Y_{1:n}^2) \leq n R_2 P_e^{(n)} + 1 \leq n \epsilon_n \quad \text{(31)}$$
Degraded Channel: Converse

**proof of converse of Theorem 3.2.**

- This then gives:
  \[ nR_1 \leq I(M_1; Y_1^{1:n}) + n\epsilon, \quad \text{and} \quad nR_2 \leq I(M_2; Y_2^{1:n}) + n\epsilon \quad (32) \]

- Since \( U \) represents \( W_2 \) in superposition scheme, we can choose \( U = W_2 \) and the Markov chain becomes \( U \to X_i \to (Y_1^i, Y_2^i) \).

- The proof proceeds similar to the above, but the key is how to identify \( U \) and this is done in a variety of ways (along with AEP like arguments).

- See Gamal & Kim for details.

---

Gaussian Broadcast Channel

- We can define a Gaussian broadcast channel and physically degraded version as follows:

![Diagram](https://via.placeholder.com/150)
Theorem 3.3

The capacity region for the Gaussian BC is the set of rate pairs \((R_1, R_2)\) such that

\[
R_1 \leq C(\alpha S_1) \tag{33}
\]

\[
R_2 \leq C\left(\frac{(1 - \alpha)S_2}{\alpha S_2 + 1}\right) \tag{34}
\]

for some \(\alpha \in [0, 1]\) where \(C(x) = \frac{1}{2} \log(1 + x)\) is the Gaussian capacity function.

General Networks

- Each sender \(X_i\) is trying to communicate simultaneously with each receiver \(Y_i\) (i.e., for all \(i\), \(X_i\) is sending to \(\{Y_i\}_i\))
- The \(X_i\) are not necessarily independent.
- Goal: necessary and sufficient conditions for achievability as we’ve done for other channels.
- Rates from \(i\) to \(j\) are \(R^{(i \rightarrow j)}\) to send message \(W^{(i \rightarrow j)} \in \{1, 2, \ldots, 2^{nR^{(i \rightarrow j)}}\}\).
- Each encoder \(i\) gets full message \(W^{i,1:m}\) and received history \(Y^{1:k-1}_i\).
- Each decoder gets all received symbols and all sent messages.
General Networks

Theorem 4.1

If all rates \( \{ R_{i \rightarrow j} \} \) are achievable, then there exists a joint distribution \( p(x^{1:m}) \) such that

\[
\forall S \subseteq V, \sum_{i \in S, j \in V \setminus S} R_{i \rightarrow j} \leq I(X^S; Y^{V \setminus S} | X^{V \setminus S}) \tag{35}
\]

meaning, all information flow across all cuts is bounded by the conditional mutual information.

- Note \( I(X^S; Y^{V \setminus S} | X^{V \setminus S}) = H(X^V) + H(Y^{V \setminus S}, X^{V \setminus S}) - H(X^{V \setminus S}) - H(X^V, Y^{V \setminus S}) \) so as a function of \( S \) this is neither submodular nor supermodular (but is a difference).

Graphical Multicast Network

- We’ve got a general network \( G = (V, E) \), each node is a sender-receiver pair, and each edge \( (j, k) \in E \) is a noiseless communication link with capacity \( C_{j,k} \).
- A source (w.l.o.g. node 1) wants to communicate a message \( W \in [2^{nR}] \) to a set of receivers \( D \subseteq V \), all of whom get an estimate \( \hat{W}_j \) for \( j \in D \) of the message.
- Each node \( k \in [2:|V|] \) can act as a relay to help source node communicate messages to destination nodes, and general computational relay operations are allowed.
Theorem 4.2 (Network Coding Theorem)

The capacity of the graphical multicast network $G = (V, E, C)$ with destination set $D \subseteq V$ is given as:

$$C = \min_{j \in D} \min_{S \subseteq V, 1 \in S, j \notin S} C(S)$$  \hspace{1cm} (36)

where $C(S)$ is the cut value in the network (sum of cut edges).

- This is relatively easy to compute, min-cut max-flow duality.
- But many network IT instances do not have known achievable regions or capacities and this is ongoing research.

Where to go next?

- For communications, the Gamal & Kim book seems very comprehensive for network information theory.
- For machine learning, learn about graphical models and inference techniques (very related to the coding in IT settings, e.g., Gallager codes vs. Pearl’s loopy belief propagation)
- For optimization, learn more about convex optimization and discrete/combinatorial optimization (great books by Schrijver on the topic). Also, he has a great book on theory of linear programming.
- For powerful generalizations on information measures (beyond that of entropy) learn about submodular functions (next quarter).
- For particular applications (e.g., speech, language, bioinformatics, computer vision, or others) learn lots about that application (there is no substitute).