Combinatorial Reasoning in Information Theory

Noga Alon, Tel Aviv U.

ISIT 2009
Combinatorial Reasoning is crucial in Information Theory

Google lists 245,000 sites with the words “Information Theory” and “Combinatorics”
The Shannon Capacity of Graphs

The (and) product $G \times H$ of two graphs $G=(V,E)$ and $H=(V',E')$ is the graph on $V \times V'$, where $(v,v') \neq (u,u')$ are adjacent iff $(u=v \text{ or } uv \in E)$ and $(u'=v' \text{ or } u'v' \in E')$.

The $n$-th power $G^n$ of $G$ is the product of $n$ copies of $G$. 
Shannon Capacity

Let $\alpha(G^n)$ denote the independence number of $G^n$.

The Shannon capacity of $G$ is

$$c(G) = \lim_{n \to \infty} [\alpha(G^n)]^{1/n} \quad (= \sup_n [\alpha(G^n)]^{1/n})$$
A channel has an input set $X$, an output set $Y$, and a fan-out set $S_x \subset Y$ for each $x \in X$.

The graph of the channel is $G=(X,E)$, where $xx' \in E$ iff $x,x'$ can be confused, that is, iff $S_x \cap S_{x'} \neq \emptyset$. 

$\alpha(G)$ = the maximum number of distinct messages the channel can communicate in a single use (with no errors)

$\alpha(G^n)$ = the maximum number of distinct messages the channel can communicate in $n$ uses.

$c(G)$ = the maximum number of messages per use the channel can communicate (with long messages)
There are several upper bounds for the Shannon Capacity:

Combinatorial [ Shannon(56)]

Geometric [ Lovász(79), Schrijver (80)]

Algebraic [Haemers(79), A (98)]
Theorem (A-98): For every $k$ there are graphs $G$ and $H$ so that $c(G), c(H) \leq k$ and yet

$$c(G + H) \geq k^{\Omega(\log k / \log \log k)}$$

where $G+H$ is the disjoint union of $G$ and $H$. This answers a problem of Shannon.
Multiple channels and privileged users (A+Lubetzky 07):

For any fixed $t$ and any family of subsets $\mathcal{F}$ of \{1, 2, $\ldots$, $t$\}, there are graphs $G_1, G_2, \ldots, G_t$ so that $c(\sum_{i \in I} G_i)$ is “large” if $I$ contains a member of $\mathcal{F}$, and is “small” otherwise.

For example, the capacity can be large iff $I$ is of size at least $k$. 

Not much is known on the Shannon capacity of graphs:

What is the maximum possible capacity of a disjoint union of two graphs, each having capacity at most $k$?

Is the maximum possible capacity of a graph $G$ with independence number 2 bounded by an absolute constant?

Is the problem of deciding if the Shannon capacity of a given input graph is at least a given real $x$ decidable?

What is the expected value of $c(G(n,\frac{1}{2}))$?
Broadcasting with side information

A sender holds a word \( x = x_1 x_2 \ldots x_n \) of \( n \) blocks \( x_i \), each consisting of \( t \) bits, which he wants to broadcast to \( m \) receivers.

Each receiver \( R_i \) is interested in \( x_{f(i)} \) and has a prior side information consisting of some other blocks \( x_j \).

\[ \beta_t = \text{minimum number of bits that have to be transmitted to allow each } R_i \text{ to recover } x_{f(i)} \]

\[ \beta = \lim_{t \to \infty} \frac{\beta_t}{t} \quad ( = \inf \frac{\beta_t}{t} ). \]
Motivation [Birk and Kol (98)]
[Bar-Yossef, Birk, Jayram and Kol (06)]:

Applications such as Video on Demand
The representing directed hypergraph:

**Vertices**: $[n] = \{1, 2, \ldots, n\}$ (indices of blocks)

**Edges**: For each receiver $R_j$ there is a directed edge $(f(j), N(j))$, where $N(j)$ is the set of indices of all blocks known to $R_i$.

Thus $\beta_t = \beta_t(H)$, $\beta_{s+t}(H) \leq \beta_s(H) + \beta_t(H)$, $\beta(H) = \inf \frac{\beta_t(H)}{t}$

**In words**: $\beta$ is the average asymptotic number of encoding bits needed per bit in each input block.
Example:
\[ H=([5],\{(1,2),(2,1),(3,\{4,5\}),(4,\{3,5\}),(5,\{3,4\}\}) \]

\[ \beta_t(H) \geq 2t \] as receivers \( R_1, R_3 \) have to get \( x_1, x_3 \)

\[ \beta_t(H) \leq 2t \] : it is enough to transmit
\[ x_1 \oplus x_2, x_3 \oplus x_4 \oplus x_5. \]
A related parameter: $\beta^*_t = \beta_1(t \cdot H)$, where $t \cdot H$ is the disjoint union of $t$ copies of $H$.

In words: $\beta^*_t(H)$ is the minimum number of bits required if the network topology is replicated $t$ independent times.

$$\beta^*(H) = \lim_{t \to \infty} \frac{\beta^*_t(H)}{t} = \inf_t \frac{\beta^*_t(H)}{t}.$$

Facts: $\alpha(H) \leq \beta(H) \leq \beta^*(H) \leq \beta_1(H)$

$$\alpha(H) = \max\{|S| : S \subset V, \forall v \in S \exists e = (v, J) \in E, J \cap S = \emptyset\}.$$
Computing $\beta^*(H)$

Given $H=([n],E)$ and $t=1$, two input strings $x, y \in \{0, 1\}^n$ are confusable if there exists a receiver (directed edge) $(i,J)$ in $E$ such that $x_j = y_j$ for all $j$ in $J$, and yet $x_i \neq y_i$.

Let $\lambda$ ($=\lambda_n$) denote the maximum cardinality of a set of input strings that are pairwise non-confusable.

**Theorem 1** [A, Lubetzky, Stav(08)]:

$$\beta^*(H) = \lim_{t \to \infty} \frac{\beta_1(tH)}{t} = n - \log_2 \lambda.$$
Corollary [ALS]: For $H = \overline{C_5} = (Z_5, E)$ where $E = \{(i, \{i - 1, i + 1\}) : i \in Z_5\}$, $\beta_1(H) = 3, \beta^{\ast}(H) = 5 - \log_2 5 = 2.68$. 

This implies that in Network Coding (introduced by Ahlswede, Cai, Li and Yeung), nonlinear capacity may exceed linear capacity [Earlier examples given by Dougherty, Freiling and Zeger (05), and by Chan and Grant (07).]
Theorem 2 [A, Weinstein (08)]: For every constant $C$ there is an (explicit) hypergraph $H$ for which $\beta^*(H)<3$ and $\beta_1(H) > C$. 

$R_1^{(r)}, \ldots, R_m^{(r)} \quad R_1^{(b)}, \ldots, R_m^{(b)} \quad R_1^{(g)}, \ldots, R_m^{(g)} \quad R_1^{(l)}, \ldots, R_m^{(l)}$
Theorem 3 [A,Hassidim,Lubetzky,Stav,Weinsten (08)]:

For every constant $C$ there is an (explicit) hypergraph $H$ for which $\beta(H)=2$ and yet $\beta^*(H)>C$. 
Codes for disjoint unions of hypergraphs

Definition: The confusion graph $C(H)$ of a directed hypergraph $H=([n],E)$ describing a broadcast network is the undirected graph on $\{0,1\}^n$, where $x,y$ are adjacent iff for some $e=(i,J)$ in $E$, $x_i \neq y_i$ and yet $x_j = y_j$ for all $j \in J$.

Here each block is of length 1, the vertices are all possible input words, and two are adjacent iff they are confusable.

Observation 1: The number of words in the optimal code for $H$ is $\chi(C(H))$. That is, $\beta_1(H) = \lceil \log_2 \chi(C(H)) \rceil$. 
Definition: The OR graph product of $G_1$ and $G_2$ is the graph on $V(G_1) \times V(G_2)$, where $(u,v)$ and $(u',v')$ are adjacent if either $uu' \in E(G_1)$ or $vv' \in E(G_2)$ (or both).

Let $G^{\lor k}$ denote the $k$-fold OR product of $G$.

Observation 2: For any pair $H_1$ and $H_2$ of directed hypergraphs, the confusion graph of their disjoint union is the OR product of $C(H_1)$ and $C(H_2)$.

Theorem [McEliece+Posner(71), Berge+Simonovits(74)]: For every graph $G \lim_{t \to \infty} [\chi(G^{\lor t})]^{1/t} = \chi_f(G)$

where $\chi_f(G)$ is the fractional chromatic number of $G$. 
For a directed hypergraph $H=([n], E)$, $G=C(H)$ is a Cayley Graph and thus $\chi_f(G) = |V(G)|/\alpha(G)$.

Hence $\chi_f(C(H)) = 2^n / \lambda(H)$, where $\lambda(H)$ is the maximum cardinality of a set of pairwise non-confusable words.

Therefore $\beta^*(H) = \log[2^n / \lambda(H)] = n - \log \lambda(H)$ proving Theorem 1.
Confusion graphs with $\chi$ much bigger than $\chi_f$:
An auxiliary graph: put $n=2^k$, and let $G$ be the Cayley Graph on $(\mathbb{Z}_2)^k$ in which $i,j$ are adjacent if the Hamming distance $|i \oplus j|$ between $i$ and $j$ is at least $k - \frac{\sqrt{k}}{100}$.

**Fact 1:** $\chi(G) > \frac{\sqrt{k}}{100}$ and $\chi_f(G) < 2.05$.

**Proof:** The Knéser Graph $K(k, \frac{k}{2} - \frac{\sqrt{k}}{200})$ is a subgraph of $G$, and hence by Lovász [78] the chromatic number is at least $k - 2\left(\frac{k}{2} - \frac{\sqrt{k}}{200}\right) + 2 > \frac{\sqrt{k}}{100}$.

The fractional chromatic number is small as that’s a Cayley Graph and there is an independent set of size

$$\sum_{i<k/2-\sqrt{k}/200} \binom{k}{i} > \frac{2^k}{2.05}.$$
Let $H=([n], E)$ be the directed hypergraph on $[n]$, where $N=2^k$, and for each $i, j$ satisfying $|i \oplus j| \geq k - \frac{\sqrt{k}}{100}$, $(i, V-\{i, j\}), (j, V-\{i, j\})$ are directed edges.

Let $C=C(H)$ be the corresponding confusion graph. This is the Cayley Graph on $(\mathbb{Z}_2)^n$ whose generators are all vectors $e_i$ and all $e_i \oplus e_j$ with $|i \oplus j| \geq k - \frac{\sqrt{k}}{100}$. 


**Fact 2:** \( \chi(C) \geq \chi(G) \geq \frac{\sqrt{\log n}}{100} \).

**Proof:** The induced subgraph of C on the vectors \( e_i \) is G.

**Fact 3:** \( \chi_f(C) \leq 2\chi_f(G) < 4.1. \)

**Proof (sketch):** Let I be a maximum independent set in G. Then for all \( j \in \mathbb{Z}_2^k \) and all \( \epsilon \in \{0, 1\} \), the set \( I_j = \{ u \in \mathbb{Z}_2^n : |u| \equiv \epsilon (\text{mod } 2), j \oplus \sum_i iu_i \in I \} \) is an independent set of C, and its expected size for random \( j, \epsilon \) is \( \frac{2^n}{2\chi_f(G)} \).

Thus, there is a directed hypergraph H on [n] so that \( \beta_1(H) \geq \log(\frac{\sqrt{k}}{100}) = \Omega(\log \log n) \), \( \beta^*(H) \leq \log_2 4.05 < 3. \)
This proves Theorem 2.
Open:

$\beta(H) = ?$

In particular:

Is the problem “is $\beta(H) < x$” decidable?

Remark (Kleinberg+Lubetzky):
$\beta(C_5) = \beta(H) = 2.5$ for $H = \{Z_5, \{(i, \{i - 1, i + 1\}) : i \in Z_5\}$
Conclusions:

**Combinatorics** is a powerful tool for tackling Problems in **Information Theory**

**Two representative examples:**
-- The Shannon capacity of graphs
-- Broadcasting with side information

**Information Theory** is a powerful tool for investigating **Combinatorial** problems