Gauss, Cayley and Projective Linear Groups

Babak Hassibi

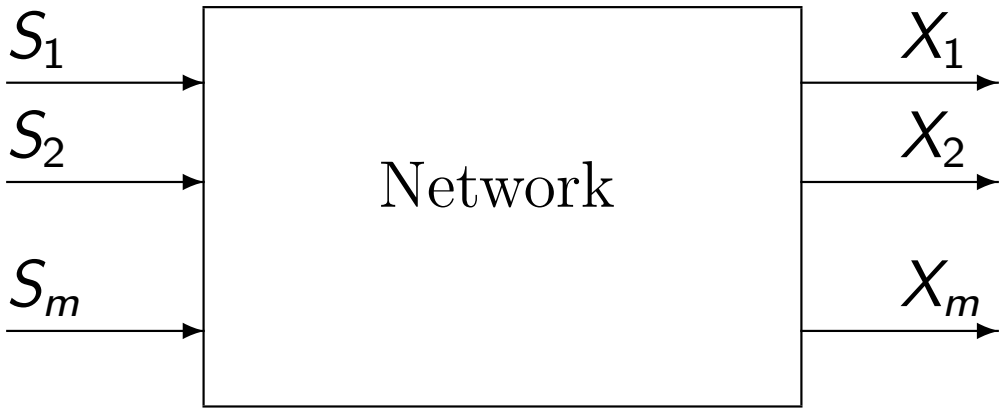
Joint work with Sormeh Shadbakht, Wei Mao and Matthew Thill

Department of Electrical Engineering
California Institute of Technology, Pasadena, CA 91125

First International Workshop on Entropy and Information Inequalities
April, 16, 2013, Chinese University of Hong Kong

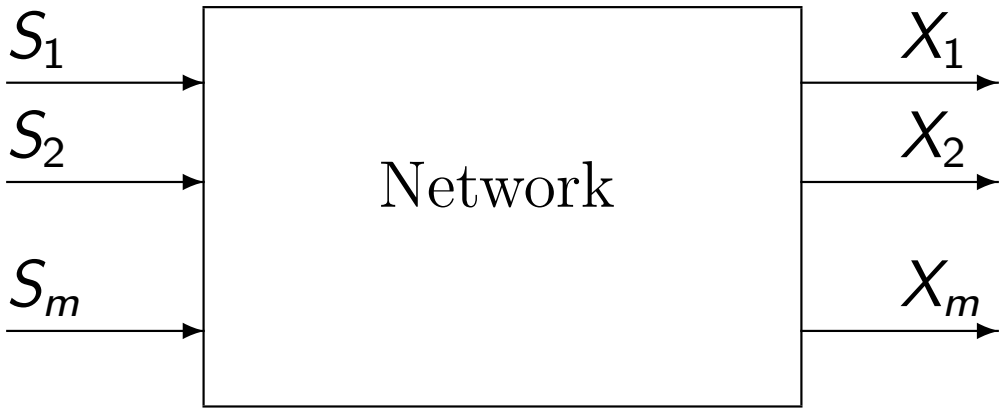
A Generic Network Problem

Consider the following acyclic discrete memory-less network and assume that each source needs to transmit to its corresponding destination at rate R_i, $i = 1, 2, \ldots, m$:



A Generic Network Problem

Consider the following acyclic discrete memory-less network and assume that each source needs to transmit to its corresponding destination at rate R_i, $i = 1, 2, \ldots, m$:



It is not terribly hard to show that (cf. Ahlswede) the *rate region* for reliable communication is

$$\mathcal{R} = \mathrm{cl}\left\{R_i, i = 1, \ldots, m \mid R_i < \frac{1}{T}\left(H(X_i^T) - H(X_i^T|S_i^T)\right)\right\} \quad \text{as } T \to \infty$$

A Generic Network Problem

Equivalently, if we are interested in optimizing a certain linear combination of the rates, we must solve

$$\lim_{T\to\infty} \sup_{p(S_i^T) \text{ and network operations}} \sum_{i=1}^{m} \alpha_i \frac{1}{T} \left(H(X_i^T) - H(X_i^T|S_i^T)\right)$$

A Generic Network Problem

Equivalently, if we are interested in optimizing a certain linear combination of the rates, we must solve

$$\lim_{T\to\infty} \sup_{p(S_i^T) \text{ and network operations}} \sum_{i=1}^{m} \alpha_i \frac{1}{T} \left(H(X_i^T) - H(X_i^T|S_i^T)\right)$$

This problem is notoriously difficult, since

A Generic Network Problem

Equivalently, if we are interested in optimizing a certain linear combination of the rates, we must solve

$$\lim_{T\to\infty} \sup_{p(S_i^T) \text{ and network operations}} \sum_{i=1}^{m} \alpha_i \frac{1}{T} \left(H(X_i^T) - H(X_i^T|S_i^T)\right)$$

This problem is notoriously difficult, since

- it is infinite-dimensional (what is called an *infinite-letter characterization*)

A Generic Network Problem

Equivalently, if we are interested in optimizing a certain linear combination of the rates, we must solve

$$\lim_{T\to\infty} \sup_{p(S_i^T) \text{ and network operations}} \sum_{i=1}^{m} \alpha_i \frac{1}{T} \left(H(X_i^T) - H(X_i^T|S_i^T)\right)$$

This problem is notoriously difficult, since

- it is infinite-dimensional (what is called an *infinite-letter characterization*)
- for any T, the problem is highly non-convex in the $p(S_i^T)$ and the "network operations"

A Generic Network Problem

Equivalently, if we are interested in optimizing a certain linear combination of the rates, we must solve

$$\lim_{T\to\infty} \sup_{p(S_i^T) \text{ and network operations}} \sum_{i=1}^{m} \alpha_i \frac{1}{T} \left(H(X_i^T) - H(X_i^T|S_i^T)\right)$$

This problem is notoriously difficult, since

- it is infinite-dimensional (what is called an *infinite-letter characterization*)
- for any T, the problem is highly non-convex in the $p(S_i^T)$ and the "network operations"

Ergo: No one does it this way!

Normalized Entropy Vectors

Consider n discrete random variables with alphabet-size N. For any set $\mathcal{S} \subseteq \{1, \ldots, n\}$, we have the normalized entropy $h_{\mathcal{S}} = \frac{1}{\log N} H(X_i, i \in \mathcal{S})$. *The $2^n - 1$ dimensional vector obtained from these entropies, is called a* normalized entropy vector.

Normalized Entropy Vectors

Consider n discrete random variables with alphabet-size N. For any set $\mathcal{S} \subseteq \{1, \ldots, n\}$, we have the normalized entropy $h_{\mathcal{S}} = \frac{1}{\log N} H(X_i, i \in \mathcal{S})$. *The $2^n - 1$ dimensional vector obtained from these entropies, is called a* normalized entropy vector.

Conversely, any $2^n - 1$ dimensional vector which can be regarded as the entropy vector of some collection of n random variables, for some value of N, is called *normalized entropic*.

Normalized Entropy Vectors

Consider n discrete random variables with alphabet-size N. For any set $\mathcal{S} \subseteq \{1, \ldots, n\}$, *we have the* normalized entropy $h_{\mathcal{S}} = \frac{1}{\log N} H(X_i, i \in \mathcal{S})$. *The* $2^n - 1$ *dimensional vector obtained from these entropies, is called a* normalized entropy vector.

Conversely, any $2^n - 1$ dimensional vector which can be regarded as the entropy vector of some collection of n random variables, for some value of N, is called *normalized entropic*.

We will denote space of normalzied entropic vectors is denoted by Ξ_n^*.

Normalized Entropy Vectors

Consider n discrete random variables with alphabet-size N. For any set $\mathcal{S} \subseteq \{1, \ldots, n\}$, *we have the* normalized entropy $h_{\mathcal{S}} = \frac{1}{\log N} H(X_i, i \in \mathcal{S})$. *The* $2^n - 1$ *dimensional vector obtained from these entropies, is called a* normalized entropy vector.

Conversely, any $2^n - 1$ dimensional vector which can be regarded as the entropy vector of some collection of n random variables, for some value of N, is called *normalized entropic*.

We will denote space of normalzied entropic vectors is denoted by Ξ_n^*.

We have focused on *normalized* entropy, since it is what somes up in

$$\sum_{i=1}^{m} \alpha_i \frac{1}{T} \left(H(X_i^T) + H(S_i^T) - H(X_i^T, S_i^T) \right).$$

and since it makes the the space Ξ_n^* compact (a finite region), $h_{\mathcal{S}} \leq |\mathcal{S}|$.

Convexity of Ξ_n^*

We should not that, *for any fixed N*, the set of normalized entropy vectors is highly non-convex. However,

Convexity of Ξ_n^*

We should not that, *for any fixed N*, the set of normalized entropy vectors is highly non-convex. However,

Theorem

The closure of the space of entropic vectors, $\bar{\Xi}_n^*$ is compact and convex.

Convexity of Ξ_n^*

We should not that, *for any fixed N*, the set of normalized entropy vectors is highly non-convex. However,

Theorem

The closure of the space of entropic vectors, $\bar{\Xi}_n^*$ is compact and convex.

- One simple proof uses *time-sharing*

Convexity of Ξ_n^*

We should not that, *for any fixed N*, the set of normalized entropy vectors is highly non-convex. However,

Theorem

The closure of the space of entropic vectors, $\bar{\Xi}_n^*$ is compact and convex.

- One simple proof uses *time-sharing*
- It should also be clear that

$$\text{cone}(\bar{\Xi}_n^*) = \bar{\Gamma}_n^*.$$

Networks and Entropy

But what does all this say about our network problem?

Networks and Entropy

But what does all this say about our network problem?

Well, networks put two types of constraints on entropy vectors:

1. topological constraints

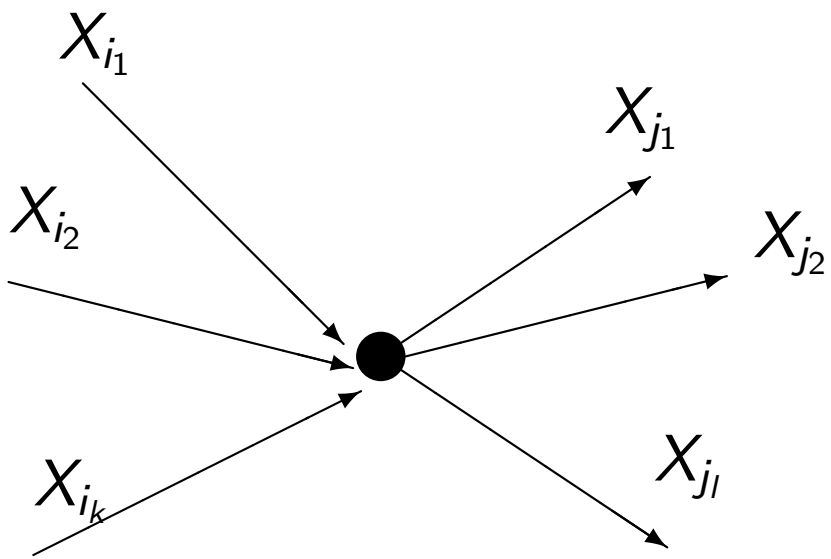
Networks and Entropy

But what does all this say about our network problem?

Well, networks put two types of constraints on entropy vectors:

1. topological constraints
2. channel constraints

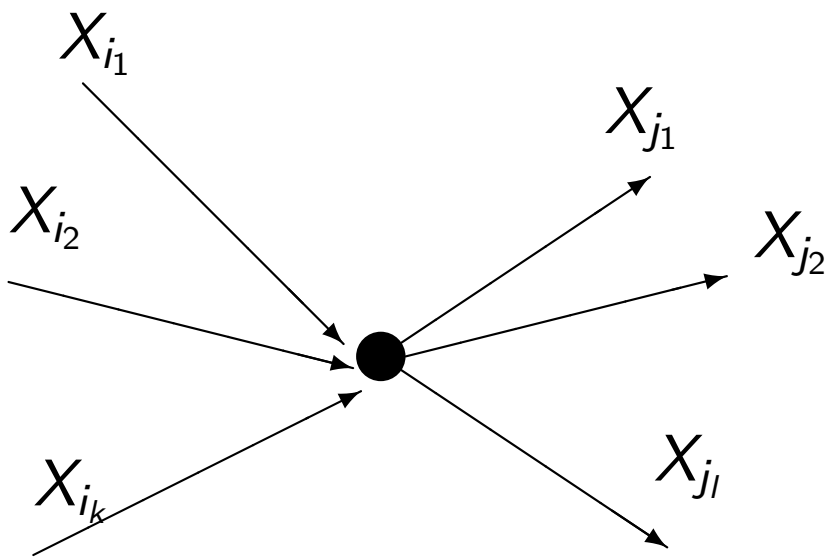
Topological Constraints



Assume the signals $X_{i_1}, \ldots, X_{i_k}$ arrive at a non-source node and the signals $X_{j_1}, \ldots, X_{j_l}$ are transmitted. This can be represented as the following linear constraints on the entropy vector:

$$h(X_{j_q}, X_{i_1}, \ldots, X_{i_k}) - h(X_{i_1}, \ldots, X_{i_k}) = 0 \quad q = 1, \ldots l$$

Topological Constraints



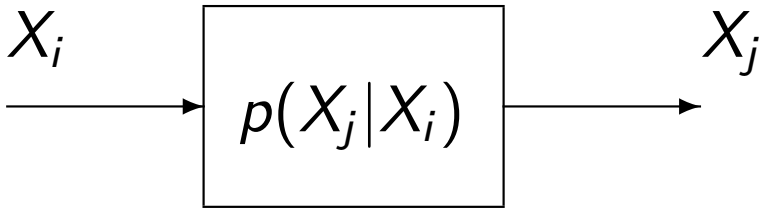
Assume the signals $X_{i_1}, \ldots, X_{i_k}$ arrive at a non-source node and the signals $X_{j_1}, \ldots, X_{j_l}$ are transmitted. This can be represented as the following linear constraints on the entropy vector:

$$h(X_{j_q}, X_{i_1}, \ldots, X_{i_k}) - h(X_{i_1}, \ldots, X_{i_k}) = 0 \quad q = 1, \ldots l$$

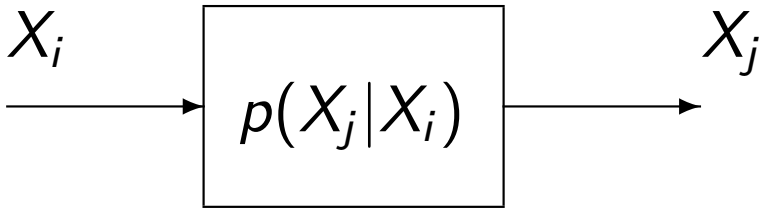
At source nodes, if S_i and S_j are independent,

$$h(S_i, S_j) - h(S_i) - h(S_j) = 0.$$

Channel Constraints



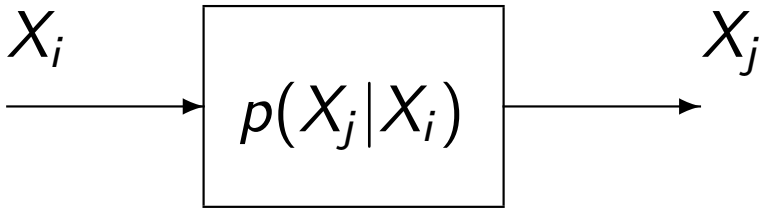
Channel Constraints



Channel constraints do not translate directly to entropies. What they do is constrain the joint distribution of all random variables in the network

$$p(X_i, X_j) = p(X_j|X_i)p(X_i),$$

Channel Constraints



Channel constraints do not translate directly to entropies. What they do is constrain the joint distribution of all random variables in the network

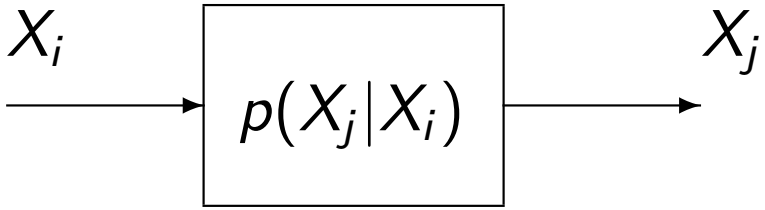
$$p(X_i, X_j) = p(X_j|X_i)p(X_i),$$

or, equivalently,

$$\int \prod_{k\neq i,j} dX_k \; p(X_1, \ldots, X_n) = p(X_j|X_i) \int \prod_{k\neq j} dX_k \; p(X_1, \ldots, X_n),$$

which is a linear constraint on the joint distribution. **Thus, the space of entropic vectors remains convex under channel constraints.**

Wired Networks

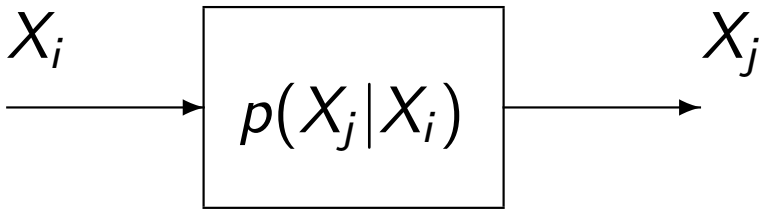


In this case, things simplify considerably. In fact, the only inequality we need is

$$h(X_i) + h(X_j) - h(X_i, X_j) \leq C_{ij},$$

where C_{ij} is the Shannon capacity of the link.

Wired Networks



In this case, things simplify considerably. In fact, the only inequality we need is

$$h(X_i) + h(X_j) - h(X_i, X_j) \leq C_{ij},$$

where C_{ij} is the Shannon capacity of the link.
Furthermore, if we use channel coding to make the link error-free, X_i and X_j can be replaced by the single random variable Z_{ij} and the constraint

$$h(Z_{ij}) \leq C_{ij}.$$

Convex Formulation of the Network Problem

Theorem

The problem of determining the capacity of an acyclic, memoryless wired network can be reduced to the optimization problem

$$\max \sum_{i=1}^{m} \alpha_i \left(h(X_i) + h(S_i) - h(X_i, S_i) \right),$$

subject to $h \in \bar{\Gamma}_n^*$ *and*

- $h(S_1, \ldots, S_m) = \sum_{i=1}^{m} h(S_i)$, *for sources*
- $h(X_{out}, X_{In}) - h(X_{In}) = 0$, *for topological constraints*
- $h(X_i) \leq C_i$, *for channel constraints*

Convex Formulation of the Network Problem

Theorem

The problem of determining the capacity of an acyclic, memoryless wired network can be reduced to the optimization problem

$$\max \sum_{i=1}^{m} \alpha_i \left(h(X_i) + h(S_i) - h(X_i, S_i) \right),$$

subject to $h \in \bar{\Gamma}_n^*$ *and*

- $h(S_1, \ldots, S_m) = \sum_{i=1}^{m} h(S_i)$, *for sources*
- $h(X_{out}, X_{In}) - h(X_{In}) = 0$, *for topological constraints*
- $h(X_i) \leq C_i$, *for channel constraints*

Thus, by going to the space of entropy vectors, we have circumvented both the *infinite-letter characterization* problem, as well as the *non-convexity*.

Remarks

- Network information theory is essentially the problem of characterizing $\bar{\Gamma}_n^*$.

Remarks

- Network information theory is essentially the problem of characterizing $\bar{\Gamma}_n^*$.
- Wired network problems reduce to convex optimization over $\bar{\Gamma}_n^*$, similar to how network flow problems reduce to linear programming.

Remarks

- Network information theory is essentially the problem of characterizing $\bar{\Gamma}_n^*$.
- Wired network problems reduce to convex optimization over $\bar{\Gamma}_n^*$, similar to how network flow problems reduce to linear programming.
- To do so, three issues need to be addressed:
  - given a vector in R^{2^n-1}, is it entropic?

Remarks

- Network information theory is essentially the problem of characterizing $\bar{\Gamma}_n^*$.
- Wired network problems reduce to convex optimization over $\bar{\Gamma}_n^*$, similar to how network flow problems reduce to linear programming.
- To do so, three issues need to be addressed:
  - given a vector in R^{2^n-1}, is it entropic?
  - given an entropic vector, find an alphabet size and joint distribution that achieves it (or comes arbitrarily close to it).

Remarks

- Network information theory is essentially the problem of characterizing $\bar{\Gamma}_n^*$.
- Wired network problems reduce to convex optimization over $\bar{\Gamma}_n^*$, similar to how network flow problems reduce to linear programming.
- To do so, three issues need to be addressed:
  - given a vector in R^{2^n-1}, is it entropic?
  - given an entropic vector, find an alphabet size and joint distribution that achieves it (or comes arbitrarily close to it).
  - can these be done in a *distributed* way?

Remarks

- Network information theory is essentially the problem of characterizing $\bar{\Gamma}_n^*$.
- Wired network problems reduce to convex optimization over $\bar{\Gamma}_n^*$, similar to how network flow problems reduce to linear programming.
- To do so, three issues need to be addressed:
  - given a vector in R^{2^n-1}, is it entropic?
  - given an entropic vector, find an alphabet size and joint distribution that achieves it (or comes arbitrarily close to it).
  - can these be done in a *distributed* way?
- The framework results in an explosion in the number of variables.

Remarks

- Network information theory is essentially the problem of characterizing $\bar{\Gamma}_n^*$.
- Wired network problems reduce to convex optimization over $\bar{\Gamma}_n^*$, similar to how network flow problems reduce to linear programming.
- To do so, three issues need to be addressed:
  - given a vector in R^{2^n-1}, is it entropic?
  - given an entropic vector, find an alphabet size and joint distribution that achieves it (or comes arbitrarily close to it).
  - can these be done in a *distributed* way?
- The framework results in an explosion in the number of variables.
  - is this really necessary?

Remarks

- Network information theory is essentially the problem of characterizing $\bar{\Gamma}_n^*$.
- Wired network problems reduce to convex optimization over $\bar{\Gamma}_n^*$, similar to how network flow problems reduce to linear programming.
- To do so, three issues need to be addressed:
  - given a vector in R^{2^n-1}, is it entropic?
  - given an entropic vector, find an alphabet size and joint distribution that achieves it (or comes arbitrarily close to it).
  - can these be done in a *distributed* way?
- The framework results in an explosion in the number of variables.
  - is this really necessary?

This is what we will focus on for the rest of the talk.

Entropy and Groups

Given a finite group G, and $G_1, \ldots, G_n$ of its subgroups, the $2^n - 1$-dimensional vector whose components are

$$v_{\mathcal{S}} = \log \frac{|G|}{|\bigcap_{\alpha \in \mathcal{S}} G_\alpha|}.$$

for all $\mathcal{S} \subseteq \{1, \ldots, n\}$, is *entropic*.

Entropy and Groups

Given a finite group G, and $G_1, \ldots, G_n$ of its subgroups, the $2^n - 1$-dimensional vector whose components are

$$v_{\mathcal{S}} = \log \frac{|G|}{|\cap_{\alpha \in \mathcal{S}} G_\alpha|}.$$

for all $\mathcal{S} \subseteq \{1, \ldots, n\}$, is *entropic*.

Conversely, any entropic vector for some collection of n random variables, *can be scaled* to correspond to some finite group and n of its subgroups [Chan and Yeung].

Abelian Groups and the Ingleton Inequality

One may ask what types of groups are needed to characterize $\bar{\Gamma}_n^*$? Here is an important result.

Abelian Groups and the Ingleton Inequality

One may ask what types of groups are needed to characterize $\bar{\Gamma}_n^*$? Here is an important result.

Theorem (Chan)

If G is an Abelian group, then the resulting entropy vectors satisfy the Ingleton bound

$$h_{ij} + h_{ik} + h_{il} + h_{jk} + h_{jl} \geq h_{ijk} + h_{ijl} + h_{kl} + h_i + h_j.$$

Abelian Groups and the Ingleton Inequality

One may ask what types of groups are needed to characterize $\bar{\Gamma}_n^*$? Here is an important result.

Theorem (Chan)

If G is an Abelian group, then the resulting entropy vectors satisfy the Ingleton bound

$$h_{ij} + h_{ik} + h_{il} + h_{jk} + h_{jl} \geq h_{ijk} + h_{ijl} + h_{kl} + h_i + h_j.$$

The Ingleton bound was first discovered in the context of representable matroids.

Abelian Groups and the Ingleton Inequality

One may ask what types of groups are needed to characterize $\bar{\Gamma}_n^*$? Here is an important result.

Theorem (Chan)

If G is an Abelian group, then the resulting entropy vectors satisfy the Ingleton bound

$$h_{ij} + h_{ik} + h_{il} + h_{jk} + h_{jl} \geq h_{ijk} + h_{ijl} + h_{kl} + h_i + h_j.$$

The Ingleton bound was first discovered in the context of representable matroids. **It is known that entropy can violate the Ingleton bound (more on this in a moment) and so Abelian groups are not sufficient.**

Where is This All Coming From?

Ans: Statistical mechanics and typical sequences

Where is This All Coming From?

Ans: Statistical mechanics and typical sequences

- Suppose we have T particles that can be in one of N states with probability p_i, $i = 1, 2, \ldots, N$.

Where is This All Coming From?

Ans: Statistical mechanics and typical sequences

- Suppose we have T particles that can be in one of N states with probability p_i, $i = 1, 2, \ldots, N$.
- Then the *typical* micro-states will be those for which

$$T_i = Tp_i.$$

Where is This All Coming From?

Ans: Statistical mechanics and typical sequences

- Suppose we have T particles that can be in one of N states with probability p_i, $i = 1, 2, \ldots, N$.
- Then the *typical* micro-states will be those for which

$$T_i = Tp_i.$$

- Since all typical micro-states are equally likely, this gives a quasi-uniform distribution.

Where is This All Coming From?

Ans: Statistical mechanics and typical sequences

- Suppose we have T particles that can be in one of N states with probability p_i, $i = 1, 2, \ldots, N$.
- Then the *typical* micro-states will be those for which

$$T_i = Tp_i.$$

- Since all typical micro-states are equally likely, this gives a quasi-uniform distribution.
- The entropy is simply the log of the number of typical microstates

$$\log \frac{T!}{T_1! T_2! \ldots T_N!}, \qquad T_i = Tp_i, \quad \sum_{i=1}^{N} T_i = T.$$

Where is This All Coming From?

Ans: Statistical mechanics and typical sequences

- Suppose we have T particles that can be in one of N states with probability p_i, $i = 1, 2, \ldots, N$.
- Then the *typical* micro-states will be those for which

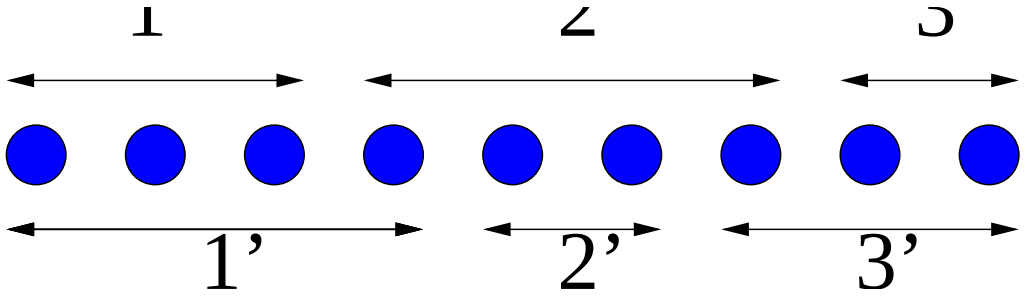
$$T_i = Tp_i.$$

- Since all typical micro-states are equally likely, this gives a quasi-uniform distribution.
- The entropy is simply the log of the number of typical microstates

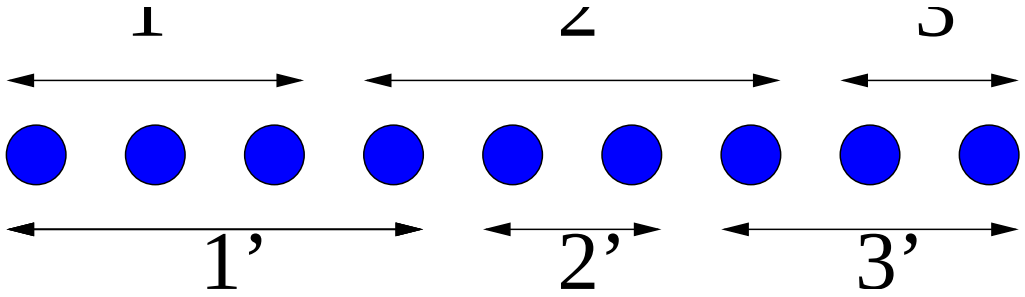
$$\log \frac{T!}{T_1! T_2! \ldots T_N!}, \qquad T_i = Tp_i, \quad \sum_{i=1}^{N} T_i = T.$$

One can think of the numerator as the size of the symmetric group S_T of T elements and the denominator as the size of a certain subgroup of S_T.

Entropy and Partitions

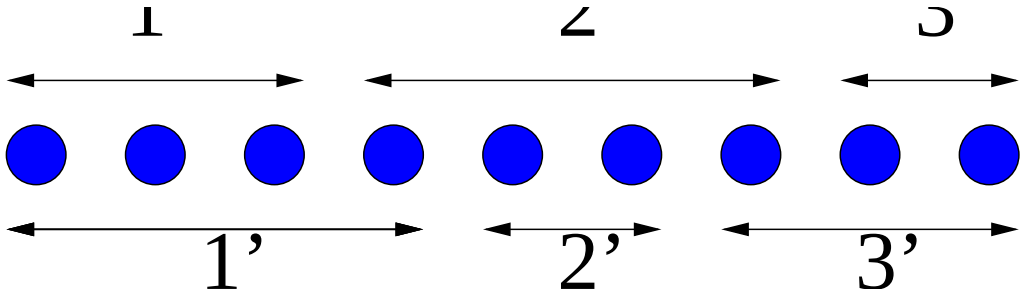


Entropy and Partitions



$$\begin{cases} T_1 = 3 \quad , \quad T_2 = 4 \quad , \quad T_3 = 2 \\ h_1 = \log \frac{9!}{3!4!2!} = \log 1260 = 10.3\text{bits} \end{cases}$$

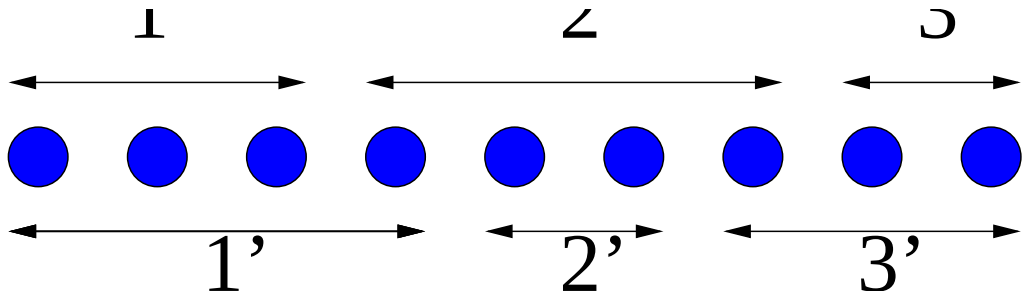
Entropy and Partitions



$$\begin{cases} T_1 = 3 \quad , \quad T_2 = 4 \quad , \quad T_3 = 2 \\ h_1 = \log \frac{9!}{3!4!2!} = \log 1260 = 10.3\text{bits} \end{cases}$$

$$\begin{cases} T_{1'} = 4 \quad , \quad T_{2'} = 2 \quad , \quad T_{3'} = 3 \\ h_2 = \log \frac{9!}{4!2!3!} = \log 1260 = 10.3\text{bits} \end{cases}$$

Entropy and Partitions



$$\begin{cases} T_1 = 3 \quad , \quad T_2 = 4 \quad , \quad T_3 = 2 \\ h_1 = \log \frac{9!}{3!4!2!} = \log 1260 = 10.3\text{bits} \end{cases}$$

$$\begin{cases} T_{1'} = 4 \quad , \quad T_{2'} = 2 \quad , \quad T_{3'} = 3 \\ h_2 = \log \frac{9!}{4!2!3!} = \log 1260 = 10.3\text{bits} \end{cases}$$

$$\begin{cases} T_{11'} = 3 \quad , \quad T_{21'} = 1 \quad , \quad T_{22'} = 2 \quad , \quad T_{23'} = 1 \quad , \quad T_{33'} = 2 \\ h_{12} = \log \frac{9!}{3!1!2!1!2!} = \log 15120 = 13.9\text{bits} \end{cases}$$

Staking Out the Entropy Region

- Take a set of size T and for each random variable partition it into N sets

Staking Out the Entropy Region

- Take a set of size T and for each random variable partition it into N sets
- The entropies and joint entropies can be computed from the partitions and their various intersections

Staking Out the Entropy Region

- Take a set of size T and for each random variable partition it into N sets
- The entropies and joint entropies can be computed from the partitions and their various intersections
- By making *local* changes to the partitions, we can move from one entropy vector to the next

Staking Out the Entropy Region

- Take a set of size T and for each random variable partition it into N sets
- The entropies and joint entropies can be computed from the partitions and their various intersections
- By making *local* changes to the partitions, we can move from one entropy vector to the next
- As T and N grow, one can stake out the entire entropic region to desired accuracy

Staking Out the Entropy Region

- Take a set of size T and for each random variable partition it into N sets
- The entropies and joint entropies can be computed from the partitions and their various intersections
- By making *local* changes to the partitions, we can move from one entropy vector to the next
- As T and N grow, one can stake out the entire entropic region to desired accuracy
- This idea can be used to perform random walks on entropy vectors and thereby MCMC methods for entropy optimization

Maximizing the Ingleton Bound via MCMC

$$I = h_{ij} + h_{ik} + h_{il} + h_{jk} + h_{jl} - h_{kl} - h_{ijk} - h_{ijl} - h_i - h_j$$

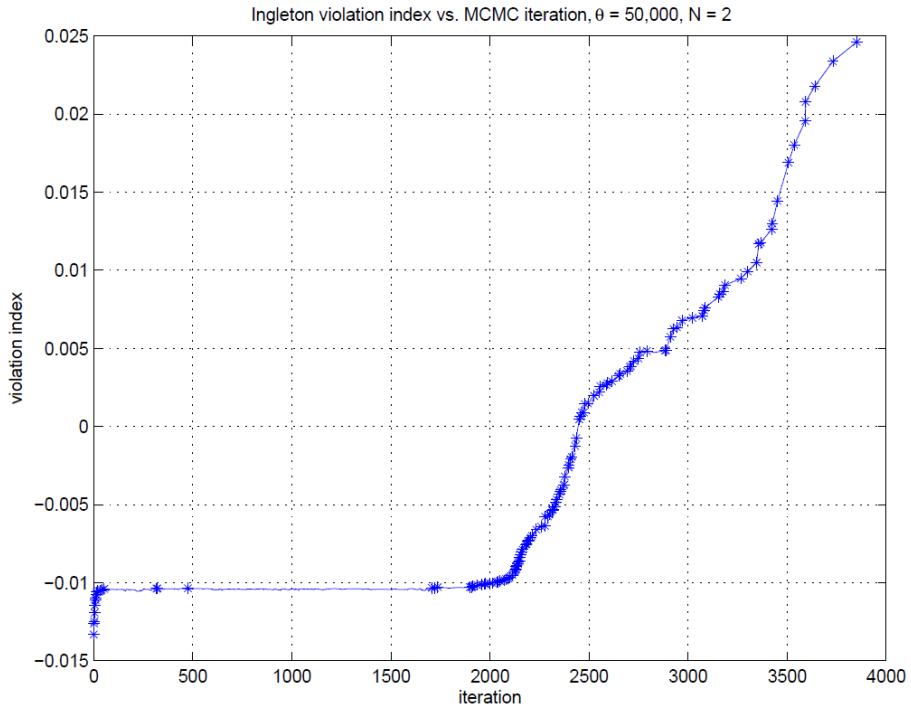
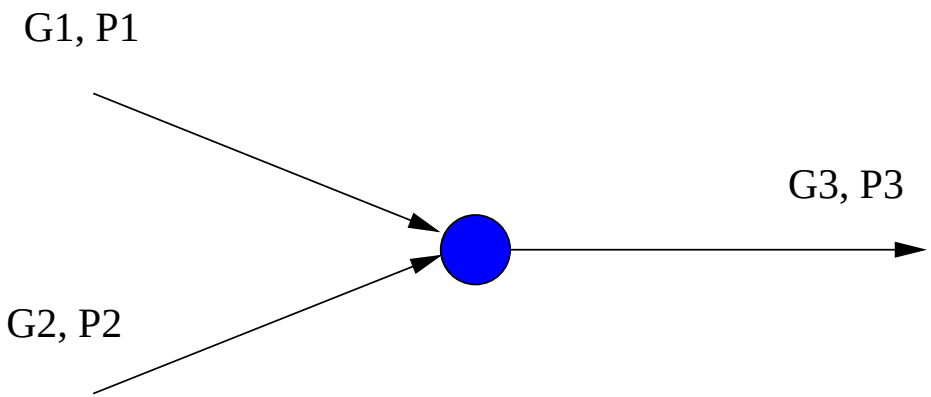


Figure : $I < 0$ is the Ingleton bound. Maximizing it with $T = 100$ and $N = 2$ using Monte Carlo Markov chain simulation achieved .025. The best prior Ingleton-bound violating instance was .0072. (Plot: $\frac{-I}{\|h\|}$)

Optimizing Information Flow in Networks

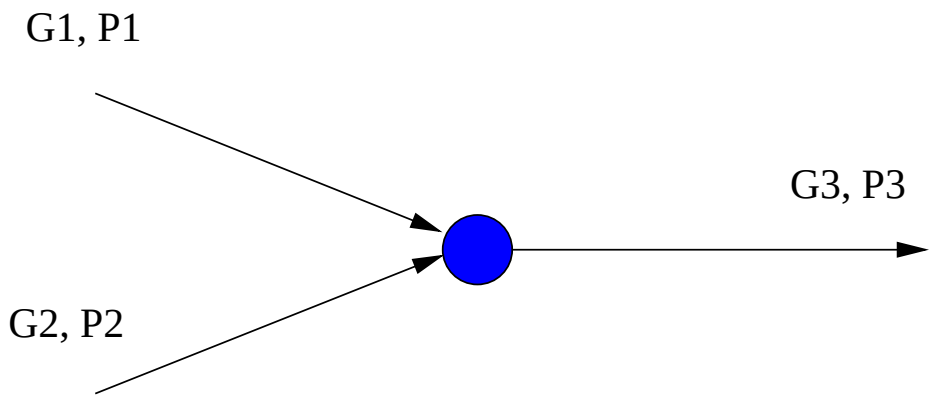
The same optimization can be done in networks, provided we respect the network topology.



$$G_3 \supseteq G_1 \cap G_2 \quad , \quad P_3 \subseteq P_1 \cap P_2$$

Optimizing Information Flow in Networks

The same optimization can be done in networks, provided we respect the network topology.

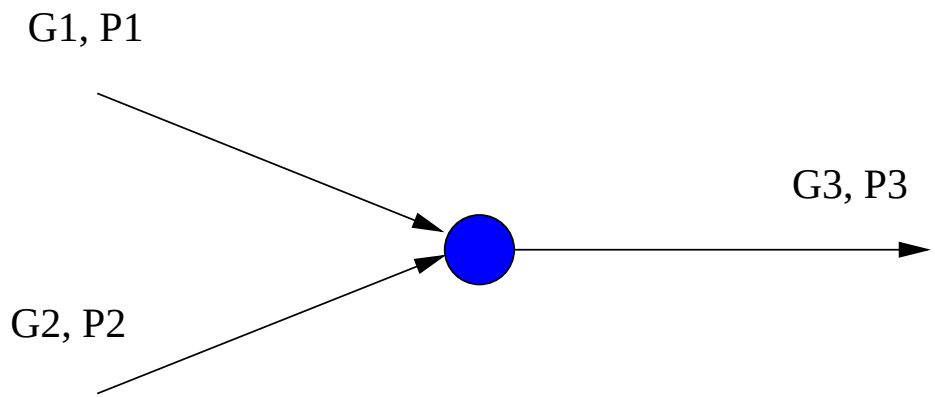


$$G_3 \supseteq G_1 \cap G_2 \quad , \quad P_3 \subseteq P_1 \cap P_2$$

- For example, the sum rate can be optimized in a *distributed* fashion

Optimizing Information Flow in Networks

The same optimization can be done in networks, provided we respect the network topology.

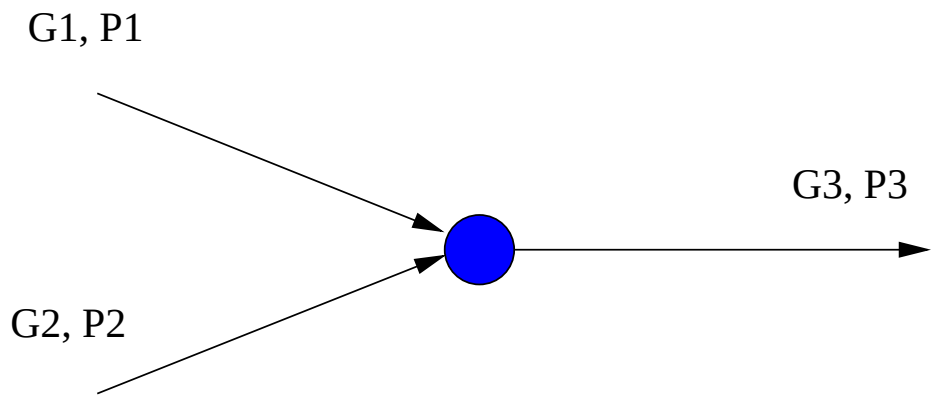


$$G_3 \supseteq G_1 \cap G_2 \quad , \quad P_3 \subseteq P_1 \cap P_2$$

- For example, the sum rate can be optimized in a *distributed* fashion
- Each edge randomly changes its partition based on information received by the sinks

Optimizing Information Flow in Networks

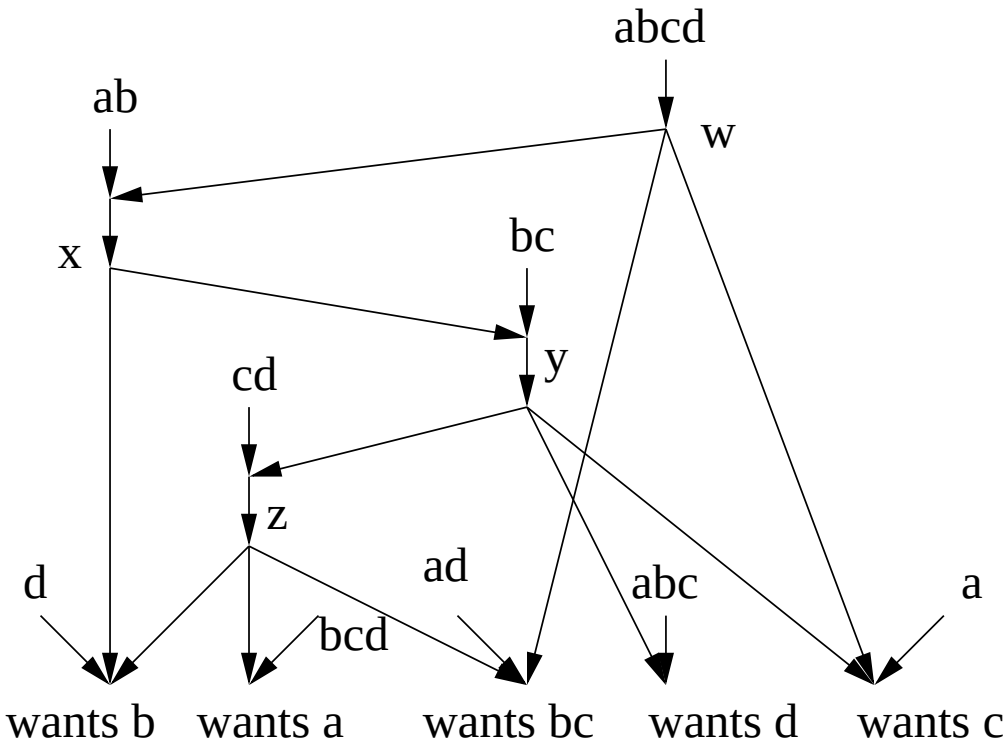
The same optimization can be done in networks, provided we respect the network topology.



$$G_3 \supseteq G_1 \cap G_2 \quad , \quad P_3 \subseteq P_1 \cap P_2$$

- For example, the sum rate can be optimized in a *distributed* fashion
- Each edge randomly changes its partition based on information received by the sinks

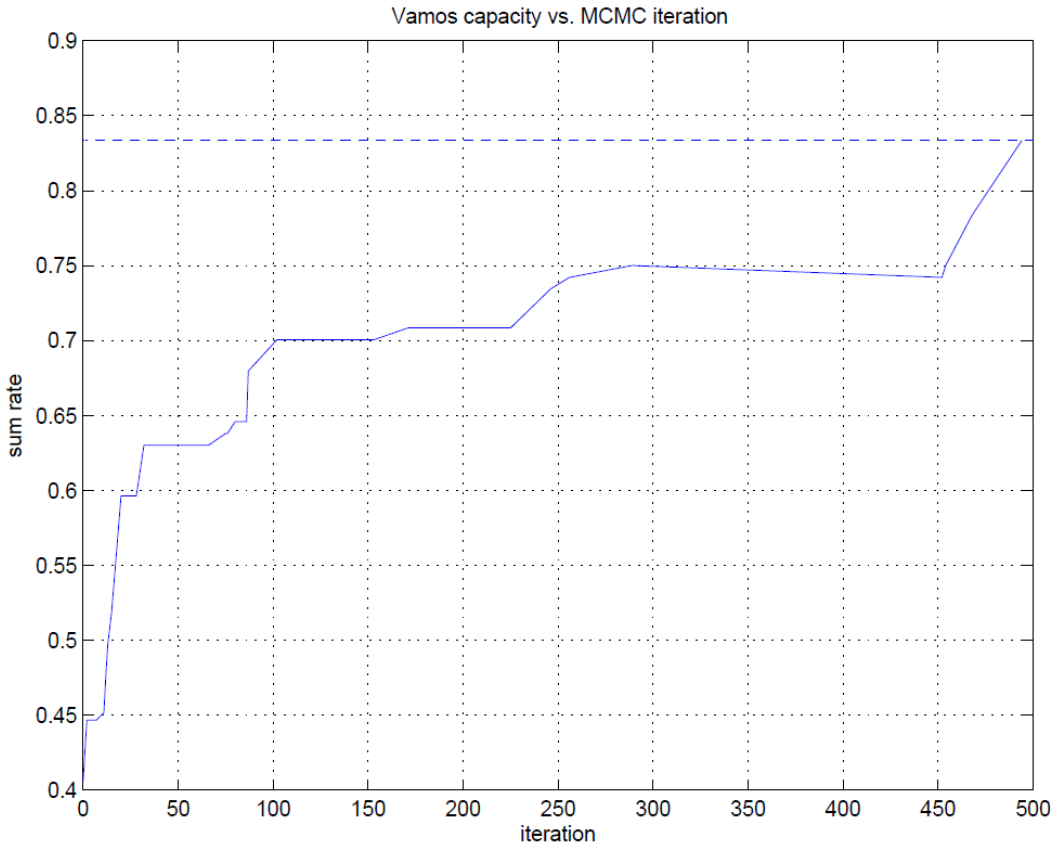
Example - The Vamos Network



- Constructed from the Vamos matroid—the *smallest non-representable matroid*—8 elements and $U(2,4)$ and $\mathcal{F}_7$ minors
- Maximum rate unknown; known to be less than $\frac{60}{11}$

Example - The Vamos Network

- Dougherty et al give a six-dimensional linear vector solution with capacity 5.
- However, using an MCMC method, we have been able to find a *nonlinear binary* solution with capacity 5 (here the search space has size 10^{12})

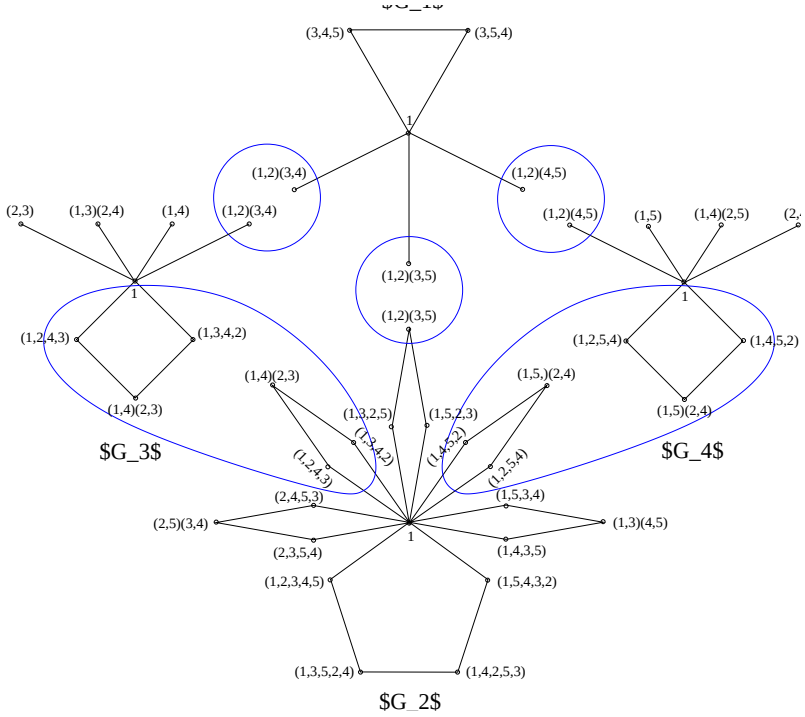


Non-Pappus Matroid and Network

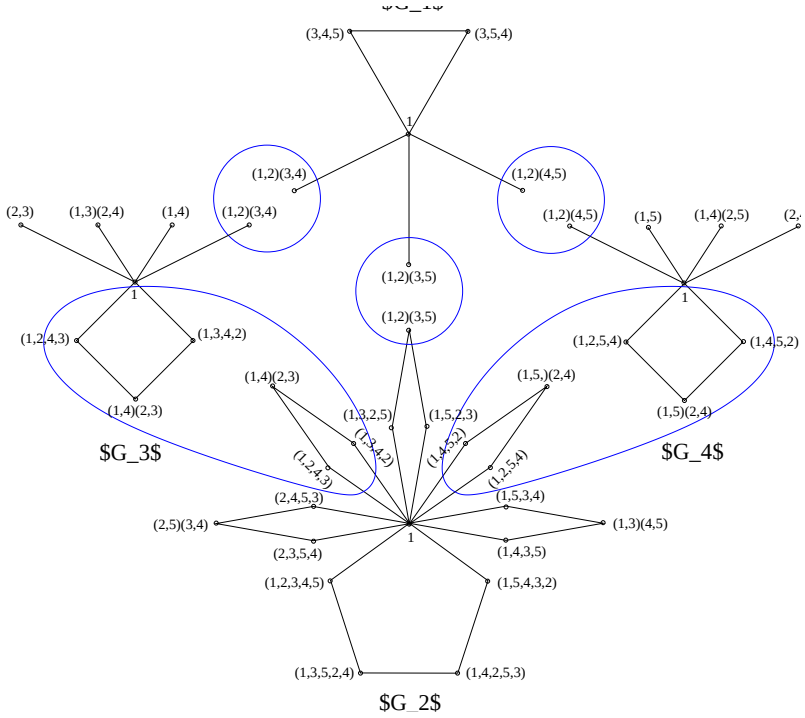
Figure : Another example of a nonrepresentable matroid.

The capacity of the corresponding network is unknown.

The Group $PGL(2, p)$

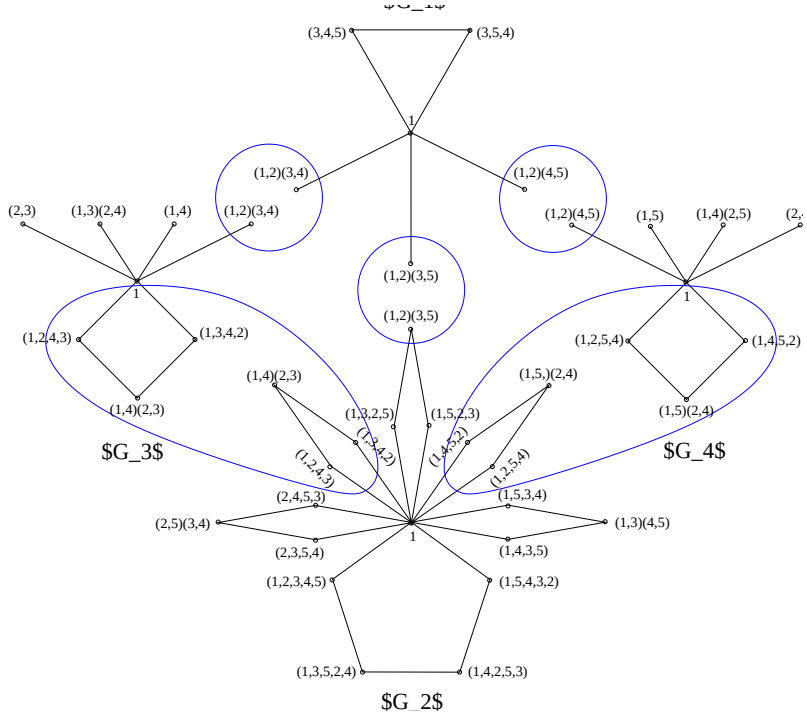


The Group $PGL(2, p)$



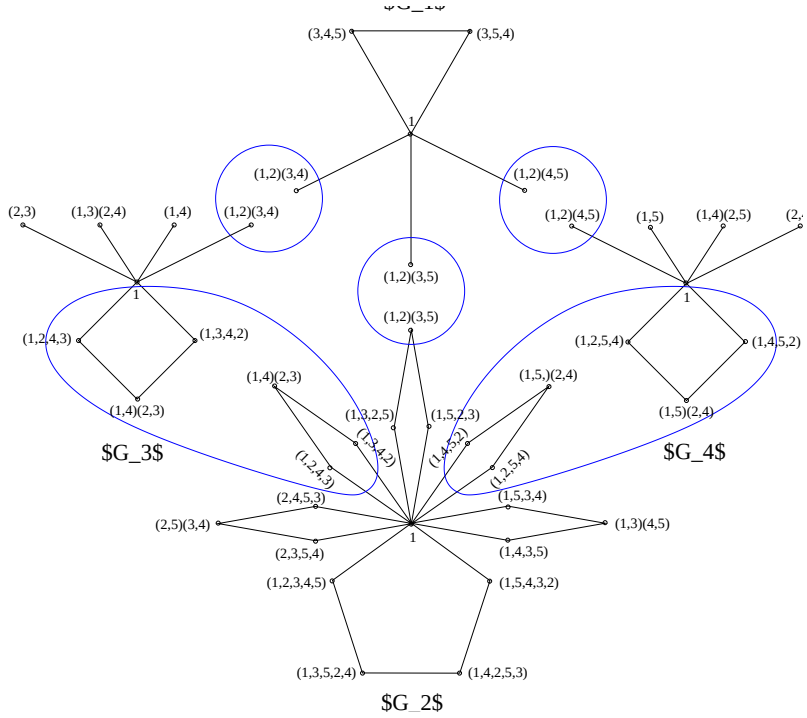
- The groups based on partitions are somewhat unstructured.

The Group $PGL(2, p)$



- The groups based on partitions are somewhat unstructured.
- We have found the smallest Ingleton-violating group to be the projective linear group $PGL(2, 5)$ with 120 elements

The Group $PGL(2, p)$



- The groups based on partitions are somewhat unstructured.
- We have found the smallest Ingleton-violating group to be the projective linear group $PGL(2, 5)$ with 120 elements
- Its generalizations, $PGL(2, p)$, for $p \geq 5$, all violate Ingleton, as does the general linear group $GL(2, p)$. These can be used to construct codes stronger than linear ones.

Entropy Vectors for Continuous Random Variables

Let $X_i \in \mathcal{R}^m$, $i = 1, \ldots, n$ be vector-valued *continuous* random variables. The normalized entropy is now defined as

$$h_{\mathcal{S}} = \frac{1}{m} H(X_i, i \in \mathcal{S}),$$

and the space of normalized entropic vectors denoted by $\bar{\Gamma}^*_{c,n}$.

Theorem (Chan)

Let

$$\sum_{\alpha \subset \{1,\ldots,n\}} k_\alpha h_\alpha \geq 0,$$

be an inequality for continuous random variables. Then

$$\sum_{\alpha \subset \{1,\ldots,n\}} k_\alpha h_\alpha + \sum_{i=1}^{n} r_i (h_{i,i^c} - h_{i^c}) \geq 0,$$

for any $r_i \geq 0$ is an inequality for discrete random variables. Conversely any inequality for discrete random variables must be of this form.

Example

This implies that it is sufficient (and perhaps simpler) to study continuous random variables.
For example, take $n = 2$. The only inequality in the continuous case is

$$h_1 + h_2 - h_{12} \geq 0.$$

Thus, the inequalities for the discrete case are

$$h_1 + h_2 - h_{12} + r_1(h_{12} - h_2) + r_2(h_{12} - h_1) \geq 0,$$

for any $r_1, r_2 \geq 0$. For example:

$$\begin{aligned} r_1 = 1,\ r_2 = 0: &\qquad h_1 \geq 0 \\ r_1 = 0,\ r_2 = 1: &\qquad h_2 \geq 0 \\ r_1 = 1,\ r_2 \to \infty &\qquad h_{12} \geq h_1 \\ r_1 \to \infty,\ r_2 = 1 &\qquad h_{12} \geq h_2 \end{aligned}$$

Gaussan Random Variables

The most obvious class of continuous random variables to consider are Gaussians. In this case, we start with a $nm \times nm$ positive definite covariance matrix R. Let $R_{\mathcal{S}}$ be the principal minor determined by the rows and columns in set $\mathcal{S}$. Then we have

$$h_{\mathcal{S}} = \frac{1}{m} \log \det R_{\mathcal{S}}.$$

Thus, the study of entropy leads us to the study of determinant inequalities. This is a subject with a long history.

Determinantal Inequalities

- **Hadamard Inequality**

$$\det R_{11} \det R_{22} \geq \det \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}.$$

- **Koteljanskii Inequality**

$$\det R_{\alpha} \det R_{\beta} \geq \det R_{\alpha \cup \beta} \det R_{\alpha \cap \beta}.$$

There are perhaps 3 reasons why looking at Gaussians may be fruitful.

Reason 1: They Work for $n = 2, 3$

Let Ω_n denote the space of entropic vectors generated by vector-valued Gaussian random variables.

Theorem

$$\bar{\Omega}_2^* = \bar{\Gamma}_{c,2}^*.$$

Theorem

For $n = 3$, the closure of the cone generated by vector-valued *Gaussian entropic vectors is* $\bar{\Gamma}_{c,3}^*$.

Theorem

For $n = 3$, the closure of the convex *cone generated by* scalar-valued *Gaussian entropic vectors is* $\bar{\Gamma}_{c,3}^*$.

Reason 2: They Violate Ingleton

Linear codes over a finite field (or codes induced by finite Abelian groups) satisfy the so-called Ingleton bound:

$$h_{ij} + h_{ik} + h_{il} + h_{jk} + h_{jl} \geq h_{ijk} + h_{ijl} + h_{kl} + h_i + h_j.$$

It is known that there exist entropy vectors that violate the Ingleton bound, though in the discrete case this is not easy to do (one needs nonlinear codes or non-Abelian groups).
However, the Ingleton bound is easy to violate with Gaussians:

$$R = \begin{bmatrix} 1 & \frac{1}{4} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 1 \end{bmatrix}$$

Just check! (It is also known that Gaussians can achieve several non-Shannon inequalities.)

Reason 3: There is Hope for Characterizing $\bar{\Omega}_n^*$

A 3×3 symmetric matrix has 6 parameters and 7 principal minors. Thus, one may expect that they satisfy some equation. Very recently, Holtz and Sturmfels (2007) have shown that the principal minors $p_1, p_2, p_3, p_{12}, p_{23}, p_{31}, p_{123}$ satisfy

$$(p_{123} - p_1 p_{23} - p_2 p_{31} - p_3 p_{12} + 2p_1 p_2 p_3)^2 = 4(p_1 p_2 - p_{12})(p_2 p_3 - p_{23})(p_3 p_1 - p_{31}).$$

And they recognized this as *Cayley's* $2 \times 2 \times 2$ *hyperdeterminant*.

What is going on?

Reason 3: There is Hope for Characterizing $\bar{\Omega}_n^*$

In general, an $n \times n$ symmetric matrix has $\frac{n(n+1)}{2}$ parameters and $2^n = 1$ principal minors. Given the p_i, the diagonals of the matrix are fixed. Given the pairwise minors p_{ij} the off-diagonals are fixed, *up to a sign*.
For example for $n = 3$:

$$R = \begin{bmatrix} p_1 & \pm\sqrt{p_1 p_2 - p_{12}} & \pm\sqrt{p_1 p_3 - p_{13}} \\ \pm\sqrt{p_1 p_2 - p_{12}} & p_2 & \pm\sqrt{p_2 p_3 - p_{23}} \\ \pm\sqrt{p_1 p_3 - p_{13}} & \pm\sqrt{p_2 p_3 - p_{23}} & p_3 \end{bmatrix}.$$

In fact, it is easy to see that out of the eight possible sign combinations, only two give different values for p_{123}. Holtz and Sturmfels showed that these two different values satisfy a quadratic equation (which is the hyperdeterminant).
(One can use this general observation to determine whether $2^n - 1$ given numbers can be the principal minors of an $n \times n$ symmetric matrix.)

Cayley's Hyperdeterminant

- The standard determinant can be obtained by looking at the bilinear form,

$$\sum_{i,j} \alpha_{ij} x_i y_j,$$

  and asking whether it has a nonzero stationary point. Taking derivatives with respect to the x_i, y_j shows that this happens when the determinant of the matrix defined by α_{ij} vanishes.

- If we instead consider the multi-linear form

$$\sum_{i_1,\ldots,i_m} \alpha_{i_1 \ldots i_m} x_{i_1} \ldots x_{i_m},$$

  the condition for having a nonzero stationary point is given by setting the hyperdeterminant of α_{ij} equal to zero.

For example for $n = 3$, this is equivalent to the condition that the following 6 nonlinear equations have nonzero solutions $x_0, x_1, y_0, y_1, z_0, z_1$:

$$\begin{aligned} x_0 y_0 + p_1 x_1 y_0 + p_2 x_0 y_1 + p_{12} x_1 y_1 &= 0 \\ p_3 x_0 y_0 + p_{31} x_1 y_0 + p_{23} x_0 y_1 + p_{123} x_1 y_1 &= 0 \end{aligned}$$

$n = 4$

Define $g_{ijk} = p_{ijk} - p_i p_{jk} - p_j p_{ki} - p_k p_{ij} + 2p_i p_j p_k$.

Theorem

The 15 principal minors of a 4×4 symmetric matrix satisfy the 5 equations

$$\begin{aligned} g_{123}^2 &= 4(p_1p_2 - p_{12})(p_2p_3 - p_{23})(p_3p_1 - p_{31}) \\ g_{124}^2 &= 4(p_1p_2 - p_{12})(p_2p_4 - p_{24})(p_4p_1 - p_{41}) \\ g_{134}^2 &= 4(p_1p_3 - p_{13})(p_3p_4 - p_{34})(p_4p_1 - p_{41}) \\ g_{123}g_{124}g_{134} &= 4(p_1p_2 - p_{12})(p_1p_3 - p_{13})(p_1p_4 - p_{14})g_{234} \\ p_{1234} &= poly(p_i, p_{ij}, p_{ijk}). \end{aligned}$$

If one could obtain the convex cone of the above algebraic variety, it would yield an Ingleton-bound-violating inner bound to $\bar{\Omega}_n^*$.

Entropy and Matroids

- A (poly)matroid is a set of objects along with a rank function that satisfies submodularity
- Entropy satisfies submodularity and therefore defines a polymatroid

$$H(A \cup B) + H(A \cap B) \leq H(A) + H(B)$$

- However, not all matroids are entropic
- A matroid is called *representable* if it can be represented by a collection of vectors over some (finite) field
- All representable matroids are entropic, but not all entropic matroids are representable
- When a matroid is representable, the corresponding network problem has an optimal solution which is a linear network code (over the field which represents the matroid)

The Fano Matroid

The Fano matroid has a representation only over $GF(2)$

$$A_7 = \begin{array}{c} \begin{matrix} a & b & c & d & e & f & g \end{matrix} \\ \left[\begin{matrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{matrix}\right] \end{array}$$

The Fano Network

- The sources are a, b, c and the sinks require c, b, a, respectively
- Links are unit capacity
- What is the maximum rate?

The Fano Network Solution

$$d = a + b \quad , \quad f = b + c \quad , \quad e = d + f = a + c \quad , \quad g = d + c = a + b + c$$

- Therefore the capacity is 3
- The network only has a solution on $GF(2)$

The Non-Fano Matroid

The Non-Fano matroid has a representation over every field except $GF(2)$

$$B_7 = \begin{array}{c} \begin{matrix} a & b & c & d & e & f & g \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix} \end{array}$$

The Non-Fano Network

- The sources are a, b, c and the sinks require c, b, a, respectively
- Links are unit capacity
- What is the maximum rate?

The Non-Fano Network Solution

$$d = a + b \quad , \quad e = a + c \quad , \quad f = b + c \quad , \quad g = a + b + c$$

- Therefore the capacity is 4
- The network only has a solution except on $GF(2)$

A Network with No Linear Solution

- This network has no linear coding solution with capacity 7
- The linear network coding capacity can be shown to be $\frac{70}{11} < 7$

Capacity is 7

- A non-Abelian solution can be given
- Alternatively, view a, b, c, d, e, f, g on the LHS as elements of $GF(2)^n$ and a, b, c, h, i, j, k on the RHS as elements of $GF(2^n + 1)$, such that

$$d = a \oplus b \quad , \quad f = b \oplus c \quad , \quad e = d \oplus f = a \oplus c \quad , \quad g = d \oplus c = a \oplus b \oplus c$$

$$h = a + b \quad , \quad i = a + c \quad , \quad j = b + c \quad , \quad k = a + b + c$$

- The resulting capacity is $7\frac{n}{\log(2^n+1)} \approx 7(1 - \frac{1}{n}2^{-n})$

Matroid Representations

- Unfortunately, determining whether a general matroid is representable is a classical open problem in matroid theory
- However, the question of whether a matroid is *binary representable* has a relatively simple answer
  - the matroid must have no 4-element minor such that all pairs are independent and all triples dependent—see matrix below

$$\begin{bmatrix} 1 & 0 & 1 & ? \\ 0 & 1 & 1 & ? \end{bmatrix}$$

Question: Is it possible to decompose an arbitrary network into two components: a binary representable component, and a component involving $U(2,4)$ minors (trivially representable in any other field), represent each component and then somehow "glue" the solutions together?

Binary Matroids

Theorem (Tutte 1958)

A matroid is binary representable iff it has no $U(2,4)$ minor.

Minors of a matroid are obtained by *deletion* and *contraction* of elements in the ground set.

Ternary and Quaternary Matroids

Theorem (Reid 1971;Bixby 1979;Seymour 1979)

A matroid is ternary representable iff it has no $U(2,5)$, $U(3,5)$, $\mathcal{F}_7$ or $\mathcal{F}_7^*$ minors.

Theorem (Geelen, Gerards, Kapoor 1997)

A matroid is quaternary representable iff it has no $U(2,6)$, $U(4,6)$, $\mathcal{F}_7^-$, $(\mathcal{F}_7^-)^*$, P_6, P_8 or P_8'' minors.

Binary Entropic Vectors

- For random variables deletion corresponds *marginalization* and contraction corresponds to *conditioning*

Theorem

A vector in R^{2^n-1} is the entropic vector of n linearly-related binary random variables iff

1. *it has integer entries*
2. $h(X_{\mathcal{S}}) \leq |\mathcal{S}|$
3. *it satisfies submodularity*
4. *for every $i, j, k, l \in \{1, 2, \ldots n\}$ and every $\mathcal{S} \in \{1, 2, \ldots n\} - \{i, j, k, l\}$, the 15-dimensional entropy vector corresponding to $\{X_i, X_j, X_k, X_l | X_{\mathcal{S}}\}$ not be $U(2, 4)$*

The Convex Cone of Binary Entropic Vectors

In order to solve general network problems over the binary field, we need to know the convex cone of binary entropic vectors

Theorem

A vector in R^{2^n-1} is in the convex cone of the entropic vectors of n linearly-related scalar binary random variables iff

1. *it is in the cone of matroids, $\mathcal{M}$*
2. *for every $i, j, k, l \in \{1, 2, \ldots n\}$ and every $\mathcal{S} \in \{1, 2, \ldots n\} - \{i, j, k, l\}$, the 15-dimensional entropy vector corresponding to $\{X_i, X_j, X_k, X_l | X_{\mathcal{S}}\}$ be in the convex cone of the entropic vectors of four binary random variables*

The convex cone of the entropic vectors of four binary random variables is given by the Ingleton inequality and 5 other types of inequalities.

The Capacity of Scalar Binary Linear Networks

We call a scalar binary linear network, one in which nodes either route packets, combine them via XORs or time-share between these two.

Theorem

The problem of determining the capacity of an acyclic, memoryless wired network using only scalar binary linear codes can be reduced to

$$\max \sum_{i=1}^{m} \alpha_i \left(h(X_i) + h(S_i) - h(X_i, S_i) \right),$$

subject to $h \in \mathcal{M}$ *and*

- $h(S_1, \ldots, S_m) = \sum_{i=1}^{m} h(S_i)$, *for sources*
- $h(X_{out}, X_{In}) - h(X_{In}) = 0$, *for topological constraints*
- $h(X_i) \leq C_i$, *for channel constraints*
- *the entropy vector for* $\{X_i, X_j, X_k, X_l | X_{\mathcal{S}}\}$, $\mathcal{S} \in \{1, 2, \ldots n\} - \{i, j, k, l\}$ *lies in the convex cone of the entropic vectors of four binary random variables*

Remarks

- The above problem is a linear program
- One problem is that the cone $\mathcal{M}$ is not known
- If we move towards vector-valued binary random variables, then the cone $\mathcal{M}$ is replaced by the polymatroidal cone, Γ_n
  - The problem here is that the characterization of representable vector-valued binary matroids is not know
  - The uniform matroid $U(2,4)$ is, for example, vector binary representable

$$\left[\underbrace{\begin{matrix}1&0\\0&1\\0&0\\0&0\end{matrix}}_{a}\ \underbrace{\begin{matrix}0&0\\0&0\\1&0\\0&1\end{matrix}}_{b}\ \underbrace{\begin{matrix}1&0\\0&1\\1&0\\0&1\end{matrix}}_{c}\ \underbrace{\begin{matrix}1&0\\0&1\\0&1\\1&1\end{matrix}}_{d}\right]$$

- In general, the complexity of the linear program is exponential:
  - there are $2^n - 1$ variables
  - there are $n + \binom{n}{2} 2^{n-2}$ submodular inequalities

Remarks

Conclusion:

- If the cone of matroids, $\mathcal{M}$, can be determined then finding optimal linear scalar codes over the binary, ternary and quaternary fields reduces to linear programming
  - when the number of sources and the fan-in of the network is small, the linear program is computationally tractable
- If the condition for vector binary representability can be established, then $\mathcal{M}$ can be replaced by Γ_n and we obtain a linear programming solution for finding optimal linear vector codes

The above can be done with reasonable complexity if the alphabet size, or T and N are small.

Conclusion

- Showed that a large class of network information theory problems can be cast as convex optimization problems over the convex set of *entropy vectors*.
- Thus, the problem is to characterize $\bar{\Gamma}_n^*$, the space of entropy vectors, which for $n \geq 4$ is a fundamental open problem.
- Explored connections to matroids, non-Shannon inequalities, quasi-uniform distributions, finite groups, determinantal inequalities
- Developed a distributed MCMC method (via random walks over partitions) for the design of optimal linear and nonlinear codes over small alphabet sizes
- Identified the smallest Ingleton-bound-violating group, $PGL(2,5)$
- Reduced the design of optimal linear codes over $GF(2)$, $GF(3)$ and $GF(4)$ for arbitrary networks to linear programming. Problem is to reduce the number of inequalities.