Optimal Coded Information Network Design and Management via Improved Characterizations of the Binary Entropy Function

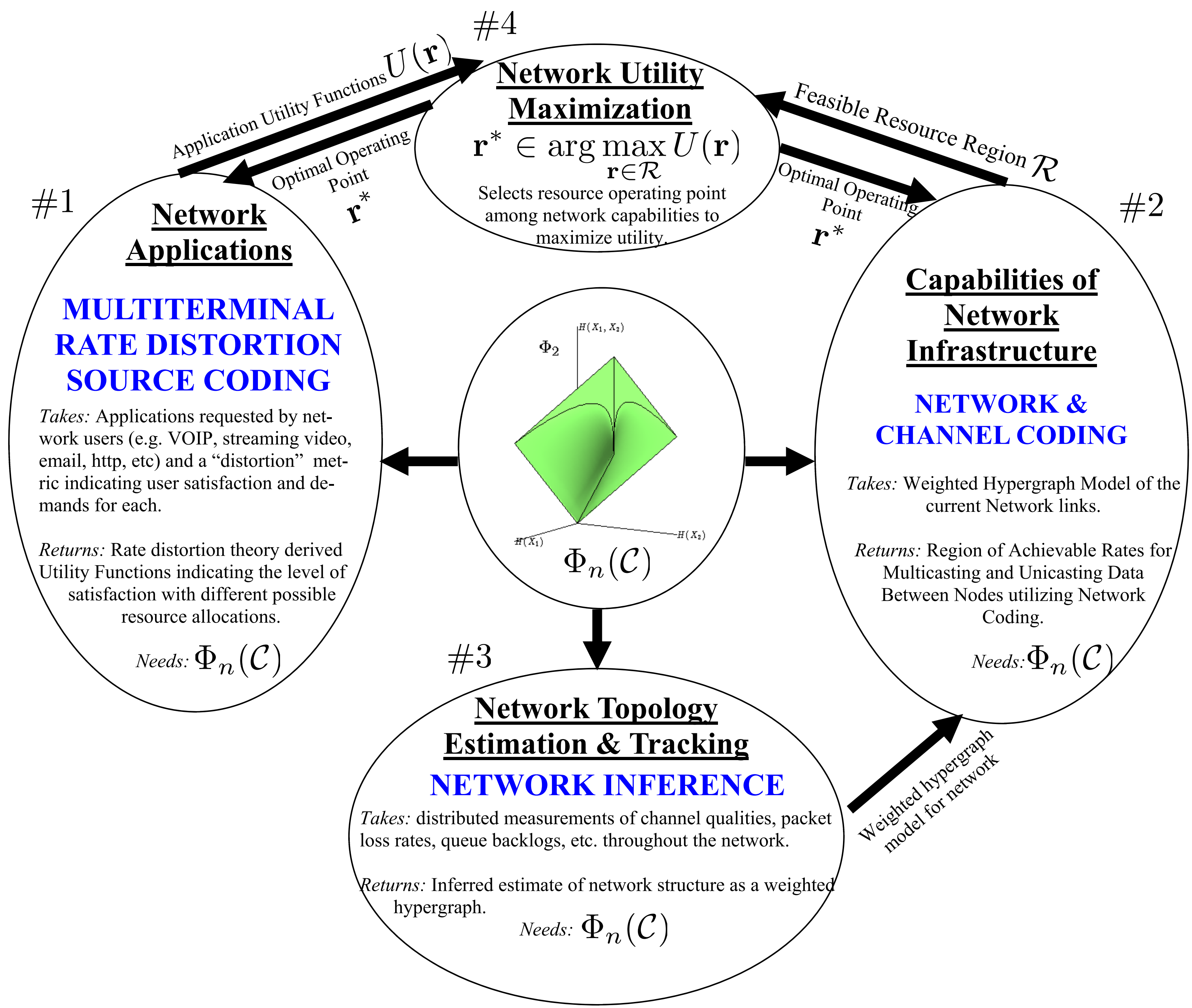
John MacLaren Walsh & Steven Weber
Department of Electrical and Computer Engineering
Drexel University
Philadelphia, PA
jwalsh@ece.drexel.edu & sweber@ece.drexel.edu

Some work in collaboration with Drakontas LLC under an Air Force STTR.

Drexel
UNIVERSITY

Overview

1. Network Utility Maximization (NUM) Mantra for Network Management
   (a) Problem Setup
   (b) Key Issues for Success
2. Answering NUM's Needs with Information Theory
   (a) Utility Functions $\Leftrightarrow$ Multiterminal Rate Distortion Source Coding
   (b) Feasible Resource Region $\Leftrightarrow$ Network Coding
   (c) Network Monitoring (Inference and Tracking) $\Leftrightarrow$ Indirect Multiterminal Rate Distortion Source Coding
3. Difficult Deep Entropy Geometry Theory Problem at Core
   (a) Motivation: Optimization in Distribution Space v.s. Optimization in Entropy Space
   (b) Different Flavors of the Set of Entropic Vectors $\bar{\Gamma}_n^*, \bar{\Omega}_n^*, \Phi_n$


#4
Network Utility Maximization
$\mathbf{r}^* \in \arg\max_{\mathbf{r}\in\mathcal{R}} U(\mathbf{r})$
Selects resource operating point among network capabilities to maximize utility.
Application Utility Functions $U(\mathbf{r})$
Optimal Operating Point $\mathbf{r}^*$
Feasible Resource Region $\mathcal{R}$
Optimal Operating Point $\mathbf{r}^*$
#1
Network Applications
MULTITERMINAL RATE DISTORTION SOURCE CODING
Takes: Applications requested by network users (e.g. VOIP, streaming video, email, http, etc) and a "distortion" metric indicating user satisfaction and demands for each.
Returns: Rate distortion theory derived Utility Functions indicating the level of satisfaction with different possible resource allocations.
Needs: $\Phi_n(\mathcal{C})$
Φ_2
$H(X_1, X_2)$
$H(X_1)$
$H(X_2)$
$\Phi_n(\mathcal{C})$
#2
Capabilities of Network Infrastructure
NETWORK & CHANNEL CODING
Takes: Weighted Hypergraph Model of the current Network links.
Returns: Region of Achievable Rates for Multicasting and Unicasting Data Between Nodes utilizing Network Coding.
Needs: $\Phi_n(\mathcal{C})$
#3
Network Topology Estimation & Tracking
NETWORK INFERENCE
Takes: distributed measurements of channel qualities, packet loss rates, queue backlogs, etc. throughout the network.
Returns: Inferred estimate of network structure as a weighted hypergraph.
Needs: $\Phi_n(\mathcal{C})$
Weighted hypergraph model for network

Network Utility Maximization (NUM) Mantra for Network Management

Network Utility Maximization [1, 2, 3] (NUM) Mantra: Provide methods for distributing constrained (finite) network resources among users and applications in such as manner as to maximize the aggregate utility over the network.

From this level of idealized generality, the problem can be broken up into four parts:

1. Determine *utility functions* $U_i(\mathbf{r})$ for different users and applications which accurately reflect their happiness when given different levels of resources $\mathbf{r}$. Aggregate these into a global utility function $U(\mathbf{r})$.

2. Determine the set $\mathcal{R}$ of *feasible resource allocations* which may be supported by a current estimated model for the network's constraints.

3. Create *distributed network inference algorithms* which can monitor and track the estimated network model.

4. Create *distributed controllers* which allocate resources in the network according to

$$\mathbf{r}^* \in \arg\max_{\mathbf{r}\in\mathcal{R}} U(\mathbf{r})$$

NUM Meets Information Theory

Much of NUM literature uses simple models for the utility functions and feasible resource allocations, assumes the network model is perfectly known, and focusses on developing and analyzing distributed controllers.

Network information theory can have a big, helpful, augmenting impact on NUM by giving a deep theoretical basis for these topics typically glossed over. Namely, we wish to convincingly show

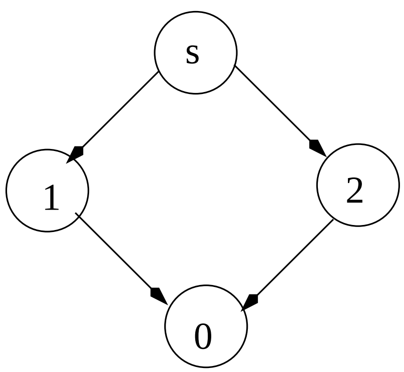
- *utility functions* can be determined with *multi-terminal rate distortion source coding theory*
- the *region of feasible resource allocations*, given a current hypergraph model for the network, can be determined with *channel and network coding theory*
- the performance of the best *distributed network inference algorithms* can be determined with *indirect multiterminal rate distortion source coding theory*

Note we will focus on *rate* (as is typical) as the resource of interest, but important work reconciles this idea with other important resources (e.g. delay and priority).

Utility Functions $\Leftrightarrow$ Multiterminal Rate Distortion Theory

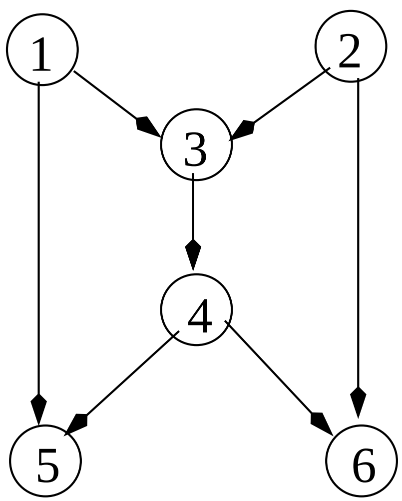
- Utility function represents happiness of an application/user with a given resource allocation.
- Consider an application, such as voice or video unicast.
- Analogous idea from information theory is the (negative of the) distortion rate function for a code, which represents reproduction fidelity attainable given a communication rate constraint.
- Note: selection of utility function should consider the "happiness" the *most efficient* use of the resources would yield (gives incentive for using resources optimally).
- Distortion - rate function is perfect fit: best avg. distortion attainable (over all possible codes = resource uses) under a rate constraint.

Utility Functions ⇔ Multiterminal Rate Distortion Theory, II



- Consider an application, such as voice or video *multicast*. (right)
- 0 can receive both what 1 and 2 receives.
- for reproduction at 0 to be good want description from 1 and 2 to contain complementary (different) source information
- for reproduction at 1 and 2 to be good want the information at 1 and 2 to contain similar (source representation) source information.
- The tradeoff between R_1, R_2 and D_0, D_1, D_2 is given by *multiple descriptions rate distortion source coding theory*.
- Utility function for this application can then be selected as a function of the three distortions (e.g. p-norm $\|[D_0, D_1, D_2\|_p$ to trade min sum distortion for min max distortion).

Region of Feasible Resource Allocations ⇔ Network Coding Theory



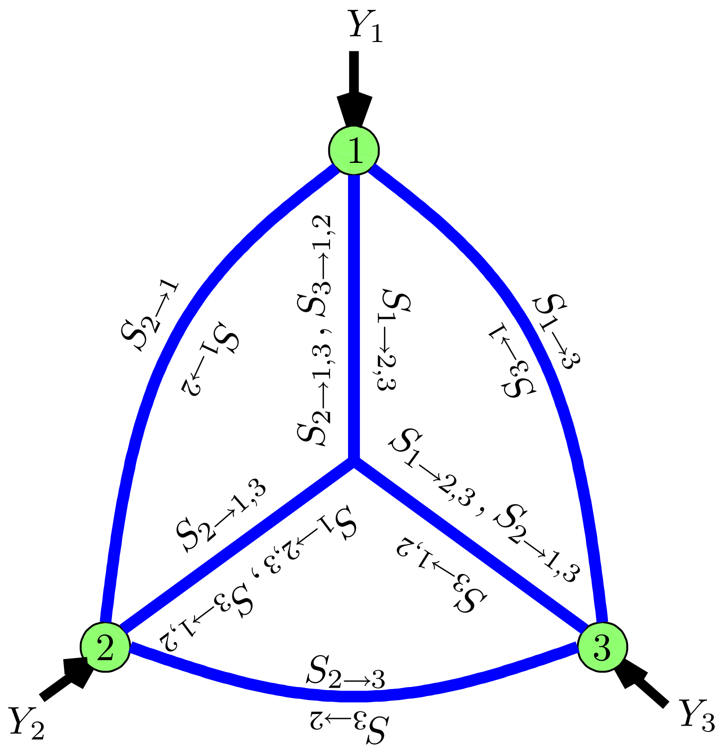
- Another question: feasible set of resource allocations $\mathcal{R}$?
- One solution (wired network model): sum up rates of all applications flowing on a link and compare with link capacity.
- *Network coding* on graph to right shows that this is suboptimal if nodes can process and combine flows/packets. The capacity region of simultaneously supportable multicasts describes the region of feasible vectors $\mathcal{R}$ under network coding.
- Wireless Capacity Regions not known exactly in many cases. Hypergraph models have been helpful in this respect [4]. Other more elegant approaches include power with rate among resources to adapt in this context, e.g. [5, 6] (bringing in power control to NUM).

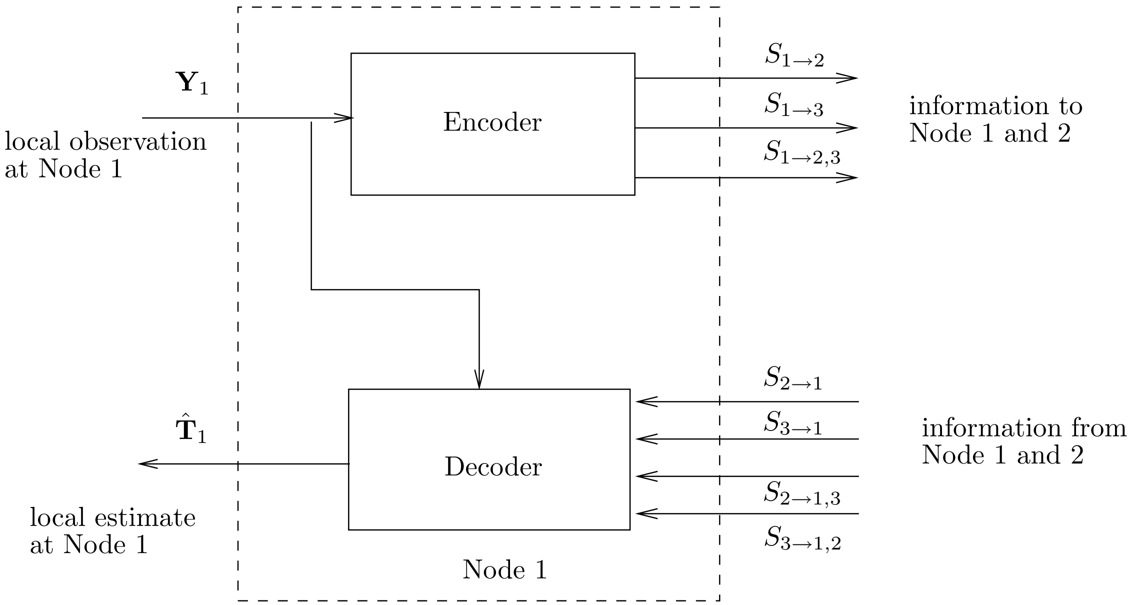
Distributed Network Inference ⇔ Indirect Multiterminal Rate Distortion Theory

What is one (naive) model for the essence of network tomography/monitoring?

- Have an underlying weighted hypergraph T_n reflecting operating network interconnections and link capacities (regions) that we would like to know (at time n).
- Have a collection of observations $Y_{i,n}$ at each node i in the network (perhaps in some cases due to intentional network "pinging").
- Need to try to share some processed information from these observations with other nodes in the network, with each node ultimately trying to from an estimate $\hat{T}_{i,n}$.
- Bandwidth for communication between the nodes about this is limited. Must live in region $\mathcal{R}$ of simultaneously supportable multicast rates $R_{i \rightarrow \mathcal{A}}$.
- Problem of choosing best encodings allowing an long run avg. estimation error D is an indirect multiterminal rate distortion problem with side information.

Distributed Network Inference ⇔ Indirect Multiterminal Rate Distortion Theory



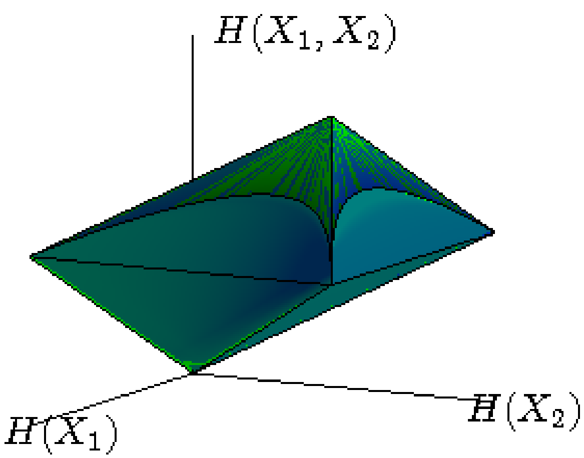


What is keeping us from using these techniques?

- Each of these additions
  - multi-terminal rate distortion region (for utility functions)
  - multi-terminal network coding rate region (for feasible resource region)
  - indirect multi-terminal rate distortion for networked inference (network tomography and monitoring)

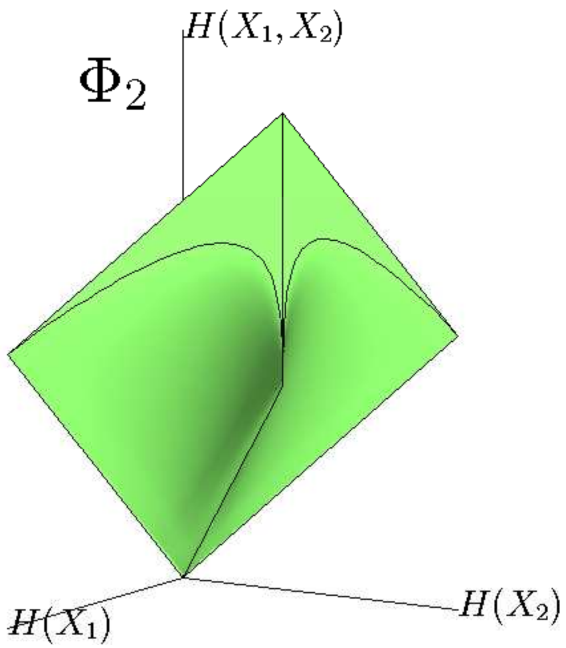
  depend on simple characterizations a fundamental geometric object in information theory: the *set of entropic vectors* $\bar{\Gamma}_n, \bar{\Omega}^*_n, \Phi_n$, and it's restriction under distribution constraints

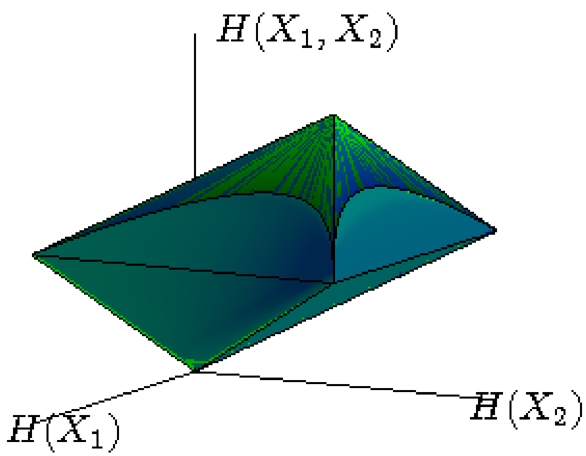
Variants on the Set of Entropic Vectors



- $\mathbf{X} := [X_1, \ldots, X_N]$, $\mathbf{X}_{\mathcal{A}} := [X_i | i \in \mathcal{A}]$
- $\mathbf{h}(p_X) := [H(\mathbf{X}_{\mathcal{A}}) | \mathcal{A} \subseteq \{1, \ldots, N\}]$
- $\Gamma_N^* := \mathbf{h}(\mathcal{D})$ [7]
- $\Phi_N := \mathbf{h}(\mathcal{D}_2)$
- $\Omega_N^* := \bigcup_{M=2}^{\infty} \frac{1}{log_2(M)} \mathbf{h}(\mathcal{D}_M)$ [8, 9]
- tight half spaces for $\bar{\Gamma}_N^*$ or $\text{conv}(\Phi_N)$ (tangent) give linear information inequalities (set of such is dual representation)
- Shannon correctly characterized for $N = 2, 3$, but for $N \geq 4$ we don't know what this set is!
- Matùš [10] Recently showed that there are an uncountably infinite number of such inequalities for $N \geq 4$
- time to find a new trick!

An Idea We are Pursuing: Restrict to Binary!





- $\Phi_N := \mathbf{h}(\mathcal{D}_2)$
- convex hull matches $\bar{\Omega}_n^*$ and convex cone matches $\bar{\Gamma}_n^*$
- good thing: we've shown that we can recursively compute boundaries of this one!
- bad thing: no longer convex, would have to convex convex hull
- HOWEVER *can* tell which faces of outer bounds are tight
  - enumerate vertices of outer bounding polytope
  - check which ones are in Φ_n

Why is $\Phi_n(\mathcal{C})$ Important?

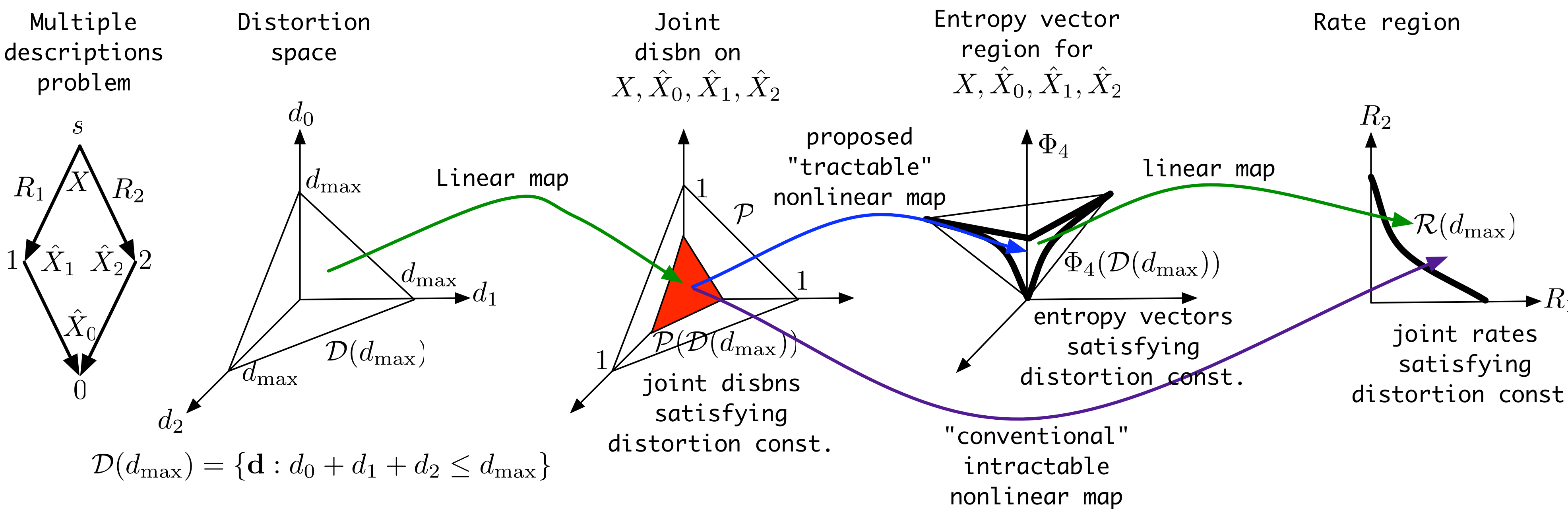
Network information theory theorems involve 2 things:

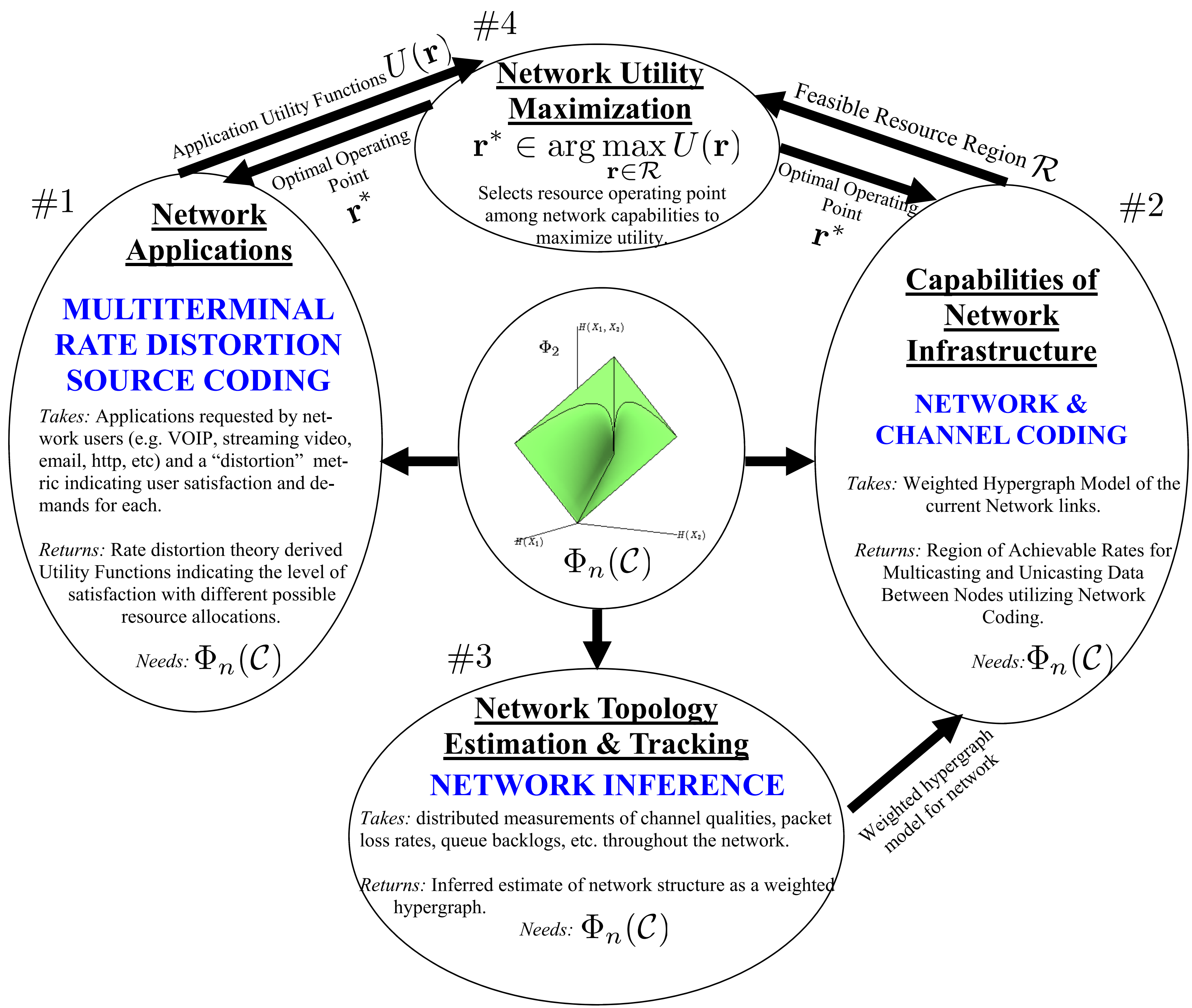
- rate inequalities in terms of weighted sums of entropies among random variables

$$R_i \geq I(Y_i; U_i | U_j)$$

- constraints in terms of the probability distribution
  - given a source distribution $\sum_{\mathbf{X} \setminus X_0} p_{\mathbf{X}} = p(X_0)$
  - given conditional distributions $p(Y|X)$
  - given distortion constraints $\mathbb{E}[d(X_1, X_2)] \leq D$
  - given conditional independence (Markov Contraints)
- When weighted sums of entropies work out to be convex functions of free distributions, everything is fine. This frequently doesn't happen in multi-terminal context.

Why is $\Phi_n(\mathcal{C})$ Important?




#4
Network Utility Maximization
$\mathbf{r}^* \in \arg\max_{\mathbf{r}\in\mathcal{R}} U(\mathbf{r})$
Selects resource operating point among network capabilities to maximize utility.
Application Utility Functions $U(\mathbf{r})$
Optimal Operating Point $\mathbf{r}^*$
Feasible Resource Region $\mathcal{R}$
Optimal Operating Point $\mathbf{r}^*$
#1
Network Applications
MULTITERMINAL RATE DISTORTION SOURCE CODING
Takes: Applications requested by network users (e.g. VOIP, streaming video, email, http, etc) and a "distortion" metric indicating user satisfaction and demands for each.
Returns: Rate distortion theory derived Utility Functions indicating the level of satisfaction with different possible resource allocations.
Needs: $\Phi_n(\mathcal{C})$
#2
Capabilities of Network Infrastructure
NETWORK & CHANNEL CODING
Takes: Weighted Hypergraph Model of the current Network links.
Returns: Region of Achievable Rates for Multicasting and Unicasting Data Between Nodes utilizing Network Coding.
Needs: $\Phi_n(\mathcal{C})$
Φ_2
$H(X_1, X_2)$
$H(X_1)$
$H(X_2)$
$\Phi_n(\mathcal{C})$
#3
Network Topology Estimation & Tracking
NETWORK INFERENCE
Takes: distributed measurements of channel qualities, packet loss rates, queue backlogs, etc. throughout the network.
Returns: Inferred estimate of network structure as a weighted hypergraph.
Needs: $\Phi_n(\mathcal{C})$
Weighted hypergraph model for network

References

[1] Frank Kelly, "Charging and Rate Control for Elastic Traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.

[2] F. P. Kelley, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operations Research Society*, vol. 49, pp. 237–252, 1998.

[3] Yung Yi and Mung Chiang, "Stochastic Network Utility Maximization," *European Transactions on Telecommunications*, no. 19, pp. 421–442, 2008.

[4] D. S. Lun, "Efficient Operation of Coded Packet Networks," Ph.D. dissertation, Massachusetts Institute of Technology, 2006.

[5] D. O'Neill, A.J. Goldsmith, and S. Boyd, "Optimizing Adaptive Modulation in Wireless Networks via Utility Maximization," in *IEEE International Conference on Communications, 2008*, May 2008, pp. 3372 – 3377.

[6] Mung Chiang, "To Layer or Not To Layer: Balancing Transport and Physical Layers in Wireless Multihop Networks," in *INFOCOM*, 2004.

[7] Raymond W. Yeung, *Information Theory and Network Coding*. Springer, 2008.

[8] B. Hassibi and S. Shadbakht, "Normalized Entropy Vectors, Network Information Theory and Convex Optimization," in *IEEE Information Theory Workshop on Information Theory for Wireless Networks*, July 2007, pp. 1 – 5.

[9] ——, "On a Construction of Entropic Vectors Using Lattice-Generated Distributions," in *IEEE International Symposium on Information Theory (ISIT)*, June 2007, pp. 501 – 505.

[10] František Matúš, "Infinitely Many Information Inequalities," in *IEEE International Symposium on Information Theory (ISIT)*, June 2007, pp. 41–44.