Identifying Sufficient Statistics in Information Networks

Huaiyu Dai[1] and Husheng Li[2]
[1]Department of Electrical and Computer Engineering, NC State University, Raleigh, NC 27695, E-mail: Huaiyu_Dai@ncsu.edu
[2]Department of Electrical Engineering and Computer Sciences, University of Tennessee, Knoxville, TN, E-mail: husheng@ece.utk.edu

Abstract –**Given a network targeting information transmission, this paper considers the problem of identifying a set of edges or nodes whose transmission is sufficient to determine information flow at some other parts of the network. We give a simple information theoretic characterization for this problem, which naturally leads to a graph theoretic identifying procedure.**

I. INTRODUCTION

While information theory has been successful in addressing the single link communication problem, its progress in a network setting is rather limited. Even for the simplest structure with additive Gaussian noise, capacity regions are largely unknown [1].

Advances are made in this exciting and yet challenging area through relaxations. One direction is to consider a simpler goal: instead of characterizing the complete capacity region, people focus on a specific linear functional of achievable rate tuples – transport capacity, and mainly study its scaling law with the network size [2]. Alternatively, we can focus on the interaction among networked nodes and hide the details of lower layers at the protocol stack. In particular, all communications are assumed noise free, and further non-interfering as long as links are sufficiently separated. This deterministic approach fairly accurately represents the information transmission scenario in the Internet backbone [3], and is extended to allow broadcast and multiple access in neighborhood recently [4][5] to better capture the essence of wireless communications. Being reasonable at the high SNR regime, this approach facilitates cross fertilization from graph theory and combinatorial optimization, already witnessed in recent process in network coding.

In this paper, we also adopt this deterministic approach and consider how we can effectively characterize the constraint imposed by the inherent graphical structure of a network. In a deterministic network, the randomness is only induced by the random choice of information messages, and all quantities can be viewed as functions of the joint message set. We allow nodes in the network to realize any functions based on their own messages and causal input. This implies possibility of joint channel and network coding, and information packets can be stored, replicated, reorganized, and combined if needed. In this context, we seek to identify some edges or nodes which contain sufficient information to infer quantities at some other parts of the network, in particular, some sink nodes in the network. Such identification should be helpful in determining the bottleneck of information transportation and security in the network. Essentially our work unifies and generalizes some existing approaches in literature [7][8][9] and simplifies some proofs there.

The remainder of this paper is organized as follows: Section II presents the system model. In Section III, an information theoretic interpretation of the problem is given, building connections between information dominance and conditional independence. Section IV is devoted to a discussion of identifying conditional independence in general graphs, which is then applied to our problem in Section V. Section VI concludes the paper.

II. NETWORK MODEL

We consider a network defined by a directed (possibly cyclic) graph $G=(V,E)$, where vertices denote communication devices (nodes) and an arrow (v_1,v_2) between two nodes indicates a feasible communication link from its tail v_1 to head v_2. Two-way communications between a pair of nodes, usually denoted by an undirected edge, are explicitly represented by two directed edges.

There are K independent message sets in the network,

$$W=\{W_1,\cdots,W_K\}. \tag{1}$$

Each message W_k originates at a source s_k, destined for a set of sinks

$$T_k=\{t_k(1),\cdots,t_k(D_k)\}, \tag{2}$$

and is assumed to be uniformly randomly drawn from

$$\Omega_k=\left\{1,\cdots,\left\lceil 2^{NR_k}\right\rceil\right\}, \tag{3}$$

where N is the block length. Let $W^{(v)}$ be the set of messages generated from v, and Ω the Cartesian product of all Ω_k. For each message W_k, denote its estimate at $t_k(i)$ as $\hat{W}_k^{(i)}$.

Each node $v\in V$ is associated with a transmitted variable X_v and received variable Y_v. Assuming noise-free transmission, $Y_v=h_v(\{X_u\}_{(u,v)\in E})$ is a function of its neighbors' transmission. X_v depends (causally) on Y_v, and $W^{(v)}$, and admits a function form of

This work was supported in part by the National Science Foundation under Grant CCF-0515164 and CNS-0721815.

$$X_v = g_v\left(\{X_u\}_{(u,v)\in E}, W^{(v)}\right). \quad (4)$$

The above modeling applies to both wireline and wireless networks. For point-to-point communication networks, we can further associate with each link $e \in E$ a variable X_e [1], with the understanding that $\{X_{(v,u)}\}_{(v,u)\in E}$ collectively represent X_v; similarly $Y_v = \{X_{(u,v)}\}_{(u,v)\in E}$ is a collection of incoming edge variables, and can be viewed as a function of $\{X_u\}_{(u,v)\in E}$. In a wireless network, X_v is broadcast to all nodes u such that $(v,u)\in E$, and Y_v is a (lossy) superposition of $\{X_u\}_{(u,v)\in E}$.

Each edge $e \in E$ is capacitated by $c(e)$, reflecting an upper limit on the information flow; for a two-way channel [6], several possibilities exist, the most popular of which is the push-to-talk model, i.e., a total link capacity is time-shared in two directions. For wireless networks, extra constraints, such as capacity regions of broadcast and multiple access channels, are imposed on joint transmissions.

To facilitate our discussion below, we also introduce hypothetical source and sink nodes into the network: directed edges run from source nodes $W^{(v)}$ to node v, and from $t_k(i)$ to sink node $\hat{W}_k^{(i)}$; these edges are assumed non-interfering with each other and with links in E. It is understood that each source node only transmits its message and receives no signals, while each sink node $\hat{W}_k^{(i)}$ simply receives the estimate, a function of $Y_{t_k(i)}$ and $W^{(t_k(i))}$.

Definition 1: A rate tuple $(R_1,\cdots,R_K)$ is admissible for a given capacitated network if there exist a coding scheme such that

$$P\left(\bigcup_{k,i}\{\hat{W}_k^{(i)} \neq W_k\}\right) \to 0 \quad (5)$$

as $N \to \infty$.

The distribution of W induces a joint distribution on $\{X_v\}_{v\in V}$ (equivalently $\{X_e\}_{e\in E}$ for point-to-point networks) and $\{\hat{W}_k^{(i)}\}_{k,i}$, depending on the coding schemes employed. For convenience, we also adopt the following notation for a subset of edges or nodes A: $X_A = \{X_i : i \in A\}$.

Definition 2: Given a general deterministic network, a subset of random variables A is said to be a sufficient statistic for another subset of random variables B, if B is conditionally independent of W given A.

III. INFORMATION DOMINANCE AND CONDITIONAL INDEPENDENCE

Our study is mainly motivated by some recent work in this area, [7][8][9]. In particular, the concept of information dominance is introduced in [8], which we redefine below with our notations.

Definition 3: A subset of variables A informationally dominates another subset B if for all network coding solutions and $x, y \in \Omega$,

$$A(x) = A(y)$$

implies

$$B(x) = B(y).$$

Remark: Recall that all variables in an information network are functions of the message set *W*. Note that a traditional network coding setting, where all links are non-interfering and noise-free, is assumed in [8]. This definition also applies to source edges or sink edges introduced above. A network coding solution in [8] refers to a coding scheme such that the corresponding rates of the message sets are admissible. Actually in the traditional network coding setting, zero-error decoding is imposed.

Our main results below show that Definition 2 and 3 are equivalent when we restrict coding schemes to those feasible ones, i.e., messages can be recovered at destinations reliably. Also, such definitions apply to node variables.

Theorem 1: In a general deterministic network, a subset of variables *A* informationally dominates another subset B if and only if *A* is a sufficient statistic for *B* for all feasible coding schemes.

Proof: By Definition 3, *A* informationally dominating *B* indicates *B* is a function of *A*. That is

$$H(B \mid A) = 0, \quad (6)$$

and $W - A - B$ forms a Markov chain.

On the other hand, if *A* is a sufficient statistic for *B*, i.e.,

$$W - A - B, \quad (7)$$

we have

$$H(B) = I(W:B) = I(A;B). \quad (8)$$

So, $H(B \mid A) = 0$, and *B* is a function of *A*.

□

In the special case $B \subseteq \{\hat{W}_k^{(i)}\}_{k,i}$, i.e., when we are interested in information dominance for a subset of message estimates, we can relax the condition a bit.

Theorem 2: In a general deterministic network, a subset of variables A informationally dominates another subset $B \subseteq \{\hat{W}_k^{(i)}\}_{k,i}$ if and only if

$$C - A - B \quad (9)$$

forms a Markov chain for all feasible coding schemes, where

$$\{W_k : \hat{W}_k^{(i)} \in B \text{ for some } i\} \subseteq C \subseteq W. \quad (10)$$

Proof: If $C - A - B$,

[1] With noise-free assumption $Y_e = X_e$.

$$H(B \mid A) = H(B \mid C, A) + I(C; B \mid A) \\ \leq H(B \mid C) + I(C; B \mid A) \to 0,$$

where $H(B \mid C) \to 0$ is due to a variant of Fano's inequality. If $C = \{W_k : \hat{W}_k^{(i)} \in B \text{ for some } i\}$, we have $H(C \mid A) \to 0$ as well. The other direction is straightforward.

□

Remark: As we allow general coding schemes at nodes, in general such a subset B depends on all messages. Restricting attentions to only feasible coding schemes as dictated by Definition 3 allows us to replace *W* with its subset *C* defined in (10).

One important application of sufficient statistics in information networks is that, with $H(B \mid A) = 0$, we have

$$H(A, B) \leq H(A) \leq NC_A, \tag{11}$$

Where $C_A = \sum_{e \in A} C(e)$. Such inequalities can help determine outer bounds of information networks.

IV. IDENTIFYING CONDITIONAL INDEPENDENCE IN GENERAL GRAPHS

The purpose of drawing connections between information dominance and sufficient statistics, or equivalently, conditional independence, is that the latter can be effectively read off a graph through formal procedures [10]-[17]. Such procedures only depend on the inherent graph structure, irrespective of coding schemes. In this section, we introduce relevant results in the area artificial intelligence.

The discussion is put in the context of a rather general type of graph studied in [15], in which there can be any subset of three types of edges between two vertices v_1 and v_2: undirected edge $\{v_1, v_2\}$, and arrows in either direction (v_1, v_2) and (v_2, v_1) [2]. This type of graph is a generalization of the reciprocal graph [14], which strictly includes directed cyclic graph (DCG) and chain graph [12]. The chain graph, allowing mixture of directed and undirected edges but prohibiting directed cycles, contains undirected graph (UG) and directed acyclic graph (DAG) [11] as special cases.

Some ordering can be established in such a general graph. A (simple) pseudo-path of length *n* from v_0 to v_n is a sequence of *n* distinct edges $(e_1, \cdots, e_n)$ going through a sequence of $n+1$ distinct vertices $v_0, \cdots, v_n$, such that e_i is an edge between v_{i-1} and v_i. [3] It is a (simple) path if $e_i = \{v_{i-1}, v_i\}$ or $e_i = (v_{i-1}, v_i)$. A path is undirected if every edge is undirected, and directed otherwise. For two vertices v_i and v_j, a pre-order $v_i \leq v_j$ is defined if $v_i = v_j$ or there is a path from v_i to v_j. We say $v_i \approx v_j$ if $v_i \leq v_j$ and $v_j \leq v_i$, and $v_i \equiv v_j$ if $v_i = v_j$ or there is an undirected path from v_i to v_j. Both $\approx$ and $\equiv$ introduce equivalence relations: the $\approx$-equivalence class of vertex *v*, $c(v)$, is called the cycle component of *v*, while the $\equiv$-equivalence class of vertex *v*, $u(v)$, is called the path component of *v*. The induced equivalent classes are denoted as $C(G)$ and $U(G)$, respectively. The ancestor set of vertex *v* is defined as $\mathrm{an}(v) = \{u \in V : u \leq v\}$. The boundary set of vertex *v* is defined as $\mathrm{bd}(v) = \{u \in V : \{u, v\} \in E \text{ or } (u, v) \in E\}$, and the parent set of *v* is defined as $\mathrm{pa}(v) = \{u \in V : (u, v) \in E\}$. Clearly we have $\mathrm{pa}(v) \subseteq \mathrm{bd}(v) \subseteq \mathrm{an}(v)$ and $\mathrm{u}(v) \subseteq \mathrm{c}(v) \subseteq \mathrm{an}(v)$. For $A \subseteq V$, we define $\mathrm{an}(A) = \bigcup_{v \in A} \mathrm{an}(v)$, $\mathrm{bd}(A) = (\bigcup_{v \in A} \mathrm{bd}(v)) \setminus A$, and $\mathrm{pa}(A) = \bigcup_{v \in A} \mathrm{pa}(v)$. The closure of $A \subseteq V$ is defined as $\mathrm{cl}(A) = A \cup \mathrm{bd}(A)$. A subset *A* is called an anterior set if $A = \mathrm{an}(A)$.

For a general graph $G = (V, E)$, the class of anterior sets, $A(G)$, forms a finite distributive lattice, with set union and intersection as join and meet operations [18]. For $A \in A(G)$, define $\langle A \rangle = \bigcup \{B \in A(G) : B \subset A\}$, and $[A] = A \setminus \langle A \rangle$. The class of joint-irreducible elements of $A(G)$ is defined as $J(A(G)) = \{A \in A(G) : [A] \neq \varnothing\}$.

For a general set *V*, $\mathscr{P}_2(V) = \{\{u, v\} : u \neq v \in V\}$ is the set of all possible undirected edges, and $E^*(V) = \{(u, v) : u \neq v \in V\}$ the set of all possible directed edges. For a general graph $G = (V, E)$, let G_A be the subgraph of *G* induced by $A \subseteq V$, $E^u = \{\{u, v\} : \{u, v\} \text{ or } (u, v) \in E\}$, $E^m = E^u \cup \bigcup_{U \in U(G)} \mathscr{P}_2(\mathrm{pa}(U))$. The moral graph of *G* is an undirected graph $G^m = (V, E^m)$. Intuitively, a graph is moralized through "marrying parents" of all path components of *G* and subsequently removing all directions of edges.

Suppose a graph G is associated with a set of random variables $\{X_v\}_{v \in V}$ with a joint distribution P. In our study, we restrict our attention to the probability measures that have positive density functions with respect to (w.r.t.) some product measure on the Borel sets of $\Re^V$. For three subsets *A*, *B* and *C* of *V* we indicate the fact that X_A and X_B are conditionally independent given X_C under *P* by $X_A \perp X_B \mid X_C[P]$, and abbreviates it as $A \perp B \mid C$ if no ambiguity occurs. A distribution *P* is said *global G-Markov* if $A \perp B \mid C$ whenever *C* separates *A* and *B* in *G*. The set of distributions that satisfy the global Markov property w.r.t. *G* is denoted by M(*G*). When $P \in \mathrm{M}(G)$, it is said *G* is an independence map of *P*.

Before we proceed further, some discussion of graph separation is in order. For UG, the definition is clear: *C* separates *A*

[2] In this type of graph, undirected edge and two arrows of opposite directions are considered different, and can coexist between two nodes.

[3] Two vertices are said to be weakly connected if there is a pseudo-path between them.

and B in G if every path from A to B contains a vertex of C. For DAG, a well known criterion is Pearl's d-separation [11].

Definition 4: In a DAG, a simple pseudo-path p is said to be d-separated (or blocked) by a subset of nodes C if and only if either of the two conditions is satisfied: (1) p contains one collider v (i.e., an intermediate vertex which has two incoming edges) such that $v \notin \mathrm{an}(C)$; or (2) p contains one non-collider v (i.e., an intermediate vertex which has at least one outgoing edge) such that $v \in C$. C is said to d-separate subsets A and B in a DAG if and only if every simple pseudo-path from A to B is d-separated by C.

In our study, we adopt a more general definition of graph separation as follows.

Definition 5: For three subsets of vertices A, B and C in a general graph G, C separates A and B in G if C separates A and B in the undirected graph $\left(G_{\mathrm{an}(A\cup B\cup C)}\right)^m$.

It has been shown [12][13] that for chain graphs and directed cyclic graphs, Definition 5 coincides with Definition 4.

Back to our purpose, for all global Markov distributions, the underlying conditional independence information can be read off their associated graph by the graph separation criterion. The problem remains how to verify that a given distribution is global Markov. In literature, the "Gibbs = Markov" approach is often invoked for this purpose.

Definition 6: A distribution P is said to have a Gibbs-factorization (or factorize) according to an undirected graph G if the density function of P admits

$$f(x_V) = \alpha \prod_{C\in\mathscr{K}} \psi^C(x_C), \tag{12}$$

where $\mathscr{K}$ denotes the set of cliques of G, α is a constant, and $\left\{\psi^C\right\}$ are non-negative functions depending on x_C only.

Definition 7: A distribution P is said to have a Gibbs-factorization (or factorize) according to a general graph G if for each $A \in A(G)$, $f(x_A)$ factorizes according to $(G_A)^m$.

Definition 7': A distribution P is said to have a Gibbs-factorization (or factorize) according to a general graph G if

(a) $f(x_V) = \prod_{A\in J(A(G))} f\left(x_{[A]} \mid x_{\langle A\rangle}\right)$;

(b) for each $A \in J(A(G))$, $f\left(x_{[A]} \mid x_{\langle A\rangle}\right) = f\left(x_{[A]} \mid x_{\mathrm{bd}([A])}\right)$;

(c) for each $A \in J(A(G))$, $f\left(x_{[A]} \mid x_{\mathrm{bd}([A])}\right)$ factorizes according to $(G_A)^m$.

Lemma 1 [15]: Given a general graph G, a probability distribution with a positive density function is global G-Markov if and only if it factorizes according to G.

V. APPLICATION ON INFORMATION NETWORKS

In this section, we apply the results in Section IV to our network model discussed in Section II. A first step is to show that the induced distribution there factorizes w.r.t. the network graph. Then by Lemma 1 we can readily identify conditional independence relations through checking graph separation. As our network model only concerns directed graphs (i.e., only contains directed edges), procedures in Section IV can be simplified; this will be illustrated through examples, which derive some results in literature in a simple way.

Our discussion differentiates acyclic and cyclic networks. The following result is useful for both scenarios.

Lemma 2: A distribution P factorizes according to a directed (cyclic or acyclic) graph G if for each $A \in A(G)$, $f(x_A)$ factorizes as $\prod_{v\in A} k^v(v, \mathrm{pa}(v))$ [4].

A. Acyclic Networks

We first consider a directed acyclic network. Without loss of generality, we also make the following assumption.

Assumption 1: Graphs are connected weakly, and all the vertices and edges are reachable from some source (i.e., there is no isolated or autonomous node).

Recall that a general network coding model admits

$$X_v = g_v\left(\mathrm{pa}(X_v), W^{(v)}\right). \tag{13}$$

Our result requires another assumption as follows.

Assumption 2: For each node v such that $W^{(v)} \neq \varnothing$,

$$W^{(v)} = g'_v\left(\mathrm{pa}(X_v), X_v\right). \tag{14}$$

We think this assumption is not critical, as a feasible network coding solution requires all messages be recovered reliably. This assumption is true for linear network coding schemes.

Theorem 3: For directed acyclic networks, the induced distribution on $\{X_v\}_{v\in V}$ (equivalently $\{X_e\}_{e\in E}$ for point-to-point networks) and $\{\hat{W}_k^{(i)}\}_{k,i}$ is global Markov w.r.t. the underlying graph.

Sketch of Proof: First note the joint distribution of messages takes a product form due to independence assumption. By change of variables, the induced joint distribution on the node variables also takes a similar product form, further multiplied by the absolute value of the Jacobean determinant. For the product form, each term assumes the form of (14). The Jacobean determinant also takes the same product form, as a DAG can be ordered so that the transformation matrix is lower triangular [3][13]. For nodes with no source nodes attached, the corresponding terms can be expressed by indicator functions. The theorem follows by application of Lemma 2 and Lemma 1. □

Therefore, indentifying sufficient statistics in acyclic information networks can be transform into a problem of checking separation in the graph. For directed graphs, both Definition 4 and 5 will do, and admit a similar procedure described as follows.

[4] $\left\{k^v\right\}$ are non-negative functions referred to as kernels.

(1) Only keep the subgraph induced by vertices *A*, *B* and *C*, and their ancestors. This is equivalent to construct $G_{\mathrm{an}(A\cup B\cup C)}$ in Definition 5, and cut off all pseudo-paths of the first type in Definition 4.

(2) There are two choices: either *connect* by an edge every pair of vertices that share a common child according to Definition 5, or *remove* all *outgoing* edges of *C* according to Definition 4 (equivalent to cut off all pseudo-paths of the second type).

Separation (by *C*) is declared if there is no pseudo-path between *A* and *B* that bypasses *C*.

In the following, we illustrate this procedure through some examples, and combine them with our results in Section III to re-derive some results in literature.

Our approach directly applies to node variables $\{X_v\}_{v\in V}$. For point-to-point networks, it may be convenient to work with $\{X_e\}_{e\in E}$, and one approach is to transform the original network into a function dependence graph (FDG) for edge variables, resembling the line graph [18] of the original one, with the addition of vertices representing sources and sinks and associated edges [9]. However, in many cases we can also directly look into the original graph for separation (and conditional independence) of edge variables, due to the following observations.

(1) Only directed paths from the source nodes to relevant subsets of nodes (standing for edge variables) in the FDG are needed for separation check.

(2) A directed path in the FDG corresponds to a directed path in the original network, and vice versa.

(3) Removing an edge in the original network implies removing all outgoing edges of the corresponding node in the FDG.

Example 1: Downstreamness

Definition 8: A subset of nodes (or edges) *B* is said to be downstream of another subset *A* if every directed path from the source nodes to *B* intersects *A*.

This definition resembles that in [8], so does the following result.

Lemma 3: If subset *B* is downstream of subset *A*, X_A is a sufficient statistic for X_B .

Proof: It is easy to see that only directed paths from all source nodes to *B* (which pass through *A*) remain in $G_{\mathrm{an}(A\cup B\cup W)}$. Marriage edges may be added between these paths, but not along any of them. Alternatively, all these directed paths are cut after removing outgoing edges of *A*. So surely *A* separates *W* and *B* in *G*. By Theorem 3, $W \perp B \mid A$.

□

Remark: By Theorem 1, *A* informationally dominates *B*. This argument provides a simple proof (for the case of edge sets) to Lemma 7 and Lemma 11 in [8], which includes the input-output inequality in [7] as a special case.

Example 2: Cryptanalysis

Another interesting case is to determine whether a set of source-sink pairs *C* and *B*[5] are informationally dominated by a subset of nodes or edges *A*. If it is true, in the Shannon sense, we can figure out the intended information exchange through observing the transmission of these nodes or edges. By Theorem 2, it is equivalent to verify $C \perp B \mid A$, which can be done through the following procedure:

(1) Keep all the directed paths from all sources to *A* and *B*, and remove all other edges and vertices in the information network.

(2) For node set *A*, remove all its outgoing edges; for edge set *A*, simply remove it.

If the *C* and *B* are strongly separated in the remaining network, the conclusion holds. This argument significantly simplifies the proof of Lemmas 9, 13 and 16 in [8].

Clearly, if an edge set *A* strongly separates *C* and *B* in the original network, we reach the same conclusion. This argument serves as a simple alternative to the crypto inequality in [7].

B. Cyclic Networks

In network coding literature, cyclic information networks can be transformed into acyclic ones through time expanded representation [3][5][8]. It would be desired if inference could be directly made on the original network, rather than the (potentially) infinite time-parameterized graph.

For directed acyclic networks, a similar result as Theorem 3 can be obtained when linear network coding is explored.

Theorem 4: For directed cyclic networks, the induced distribution on $\{X_v\}_{v\in V}$ (equivalently $\{X_e\}_{e\in E}$ for point-to-point networks) and $\{\hat{W}_k^{(i)}\}_{k,i}$ is global Markov w.r.t. the underlying graph when coding schemes are confined to be linear.

Sketch of Proof: Similar to that of Theorem 3 – in this case the Jacobean determinant is a constant.

□

The non-linear coding scenario is subject to further study. In [13], separation is checked in the collapsed graph constructed from the original graph. An approach related to semi-Markov causal theories is proposed in [16] concerning discrete variables, but was found flawed [17].

VI. CONCLUSIONS

In this work, the connection between the concept of information dominance [8] and the procedure of identifying conditional independence [10]-[17] is drawn. It is hoped that progress in this direction can help determining the bottleneck of

[5] assuming $C = \{W_k : \hat{W}_k^{(i)} \in B \text{ for some } i\}$

information transportation and security in the information network.

In the paper, we consider the problem of checking whether $X - Z - Y$, given some subsets of *X*, *Y* and *Z*. Here we propose two other related problems.

1. Given *X* and *Y*, identify a critical set *Z* such that $X - Z - Y$ in an efficient way. We may further identify a minimum such set in some sense.
2. Given a set *Z*, identify maximum *X* and *Y* such that $X - Z - Y$ in some sense. The special case that $X = W$ has been addressed in [8].

Other possible extensions include consideration of correlation among sources and noise. A commonly used model for the noise process assumes it is white and independent across the network, and independent with the messages.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[2] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communications: Scaling laws and optimal operation," *IEEE Transactions on Information Theory*, vol. 50, no. 3, pp. 748–767, 2004.

[3] R. Ahlswede, N. Cai, S.-Y. R. Li and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204-1216, July 2000.

[4] N. Ratnakar and G. Kramer, "The multicast capacity of deterministic relay networks with no interference", *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2425-2432, June 2006.

[5] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow," *Proc. Allerton Conf. on Communication, Control, and Computing*, Sept. 2007.

[6] G. Kramer and S. A. Savari, "On networks of two-way channels," *DIMACS Workshop on Algebraic Coding Theory and Information Theory*, Piscataway, NJ, 2003.

[7] K. Jain, V. Vazirani, and G. Yuval, "On the capacity of multiple unicast sessions in undirected graphs," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2805-2809, June 2006.

[8] N J A. Harvey, R. Kleinberg, and A R. Lehman, "On the capacity of information networks", *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2445-2464, June 2006.

[9] G. Kramer and S. A. Savari, "Edge-cut bounds on network coding rates," *J. Network and System Management*, vol. 14, no. 1, Mar. 2006.

[10] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inform. Theory*, vol. 49, no. 1, pp. 4-21, Jan. 2003.

[11] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.

[12] S. L. Lauritzen, *Graphical Models*, Claredon Press: Oxford, 1996

[13] P. Spirtes, "Directed cyclic graphical representations of feedback models," *Proc. 11th Conf. Uncertainty in Artificial Intelligence,* San Mateo, CA: Morgan-Kaufmann, 1995.

[14] J. T. A. Koster, "Markov properties of nonrecursive causal models," *Annals of Statistics*, no. 24, no. 5, pp.2148-2177, 1996.

[15] J. T. A. Koster, "Gibbs and Markov properties of graphs," *Annals of Mathematics and Artificial Intelligence*, 21 (1997), pp. 13-26.

[16] J. Pearl and R. Dechter, "Identifying independencies in causal graphs with feedback," *Proc. 12th Conf. Uncertainty in Artificial Intelligence,* San Francisco, CA: Morgan-Kaufmann, 1996.

[17] R. M. Neal, "On deducing conditional independence from d-separation in causal graphs with feedback," *Journal of Artificial Intelligence Research*, 12 (2000) 87-91.

[18] B. A. Davey and H. A. Priestley, *Introduction to Lattices and Order*, second edition, Cambridge University Press, 2002.

[19] R. Diestel, *Graph Theory*, second edition, Springer, 2000.