

Compressing Semantic Information with Varying Priorities

Basak Guler Aylin Yener

Electrical Engineering Department
The Pennsylvania State University, University Park, PA, 16802
basak@psu.edu yener@ee.psu.edu

Abstract

Semantics of communicated data can lead to conclusions with varying degrees of priorities. Depending on the interests of the communicating parties, some facts lead to conclusions that carry a high risk when ignored, and others may not be worth the resources to share the facts leading to those uninteresting conclusions. This paper studies the worst-case semantic data compression problem for sharing facts that lead to conclusions with such varying priorities. We establish the performance bounds by utilizing the partial dependencies between the ideas and the priority distributions on the conclusions. We show that multiple term descriptions of the facts and conclusions improve the compression performance when combined with judicious partitioning of the fact space.

1 Introduction

Understanding the impact of the *meaning* of communicated messages, i.e., semantics of data, on networked communication, has recently been re-emphasized thanks to the growing interaction of humans with computers, as well as the proliferation of smart devices and cyber-physical systems. As sources become aware of their environments, sharing *useful* information becomes more important than sharing *any* information. In effect, semantic communication needs classical information theory to be extended to quantify the impact of the meaning of information conveyed in a physical system [1, 2].

Our aim in this paper is to develop an understanding of the worst-case data compression performance of communicating with semantic data between two parties by utilizing multiple term descriptions. As importantly, we will consider conclusions of different priorities. To motivate the latter, consider a network in which the first person knows the facts x and x' , and the second person knows y and y' . Assume that the following conclusions can be derived from these facts. $x \wedge y \rightarrow c$, $x \wedge y' \rightarrow c'$ and $x' \wedge y' \rightarrow c''$, whereas $x' \wedge y$ does not lead to any useful conclusion. Let the normalized priorities of these conclusions be $p_c = 0.55$, $p_{c'} = 0.44$, $p_{c''} = 0.01$. Then, depending on the scarcity of the resources, sources may agree to ignore the conclusion c'' to achieve better compression.

Semantic relations are utilized in [3] for designing semantic codewords close to the source within a given distortion. The use of semantic ambiguity and redundancy in communication between semantically-aware nodes is inspected in [4]. The relation between the worst-case message length and the number of rounds of interaction in a semantic network is studied in [5] for single term transmissions. This work, on the other hand, exploits the

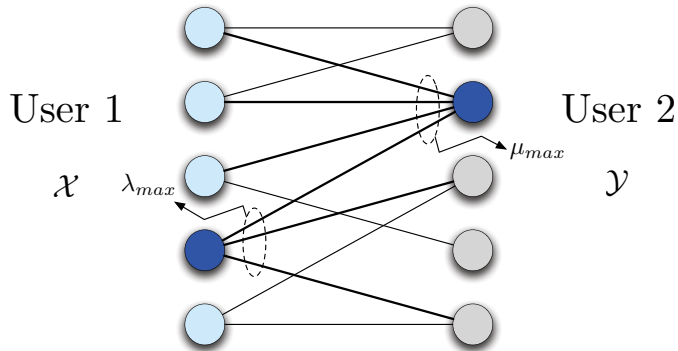


Figure 1: Semantic network with 2 users, 5 facts for each user, $\lambda_{max} = 3$ and $\mu_{max} = 4$.

benefits of multiple-term descriptions and priority assessments to achieve better compression bounds. We look into the problem of communicating with a sequence of facts that lead to *desired* conclusions. We next address the case when conclusions have different priorities and users want to recover the facts of the other person for *most* conclusions except a small low priority set, instead of merely decoding any fact. These facts might be in the form of RDF (Resource Description Framework) triples. We consider the facts as logical variables. These logical relations refer to the semantic aspects of information. We provide lower and upper bounds in terms of worst-case, zero-error message length to understand the impact of multiple term descriptions and semantic preferences on compression performance.

2 System Model

We consider a two-way communication network through which two persons share semantic information. The first person (user 1) observes facts from a set \mathcal{X} whereas the second person (user 2) knows facts from \mathcal{Y} . Some pairs of facts lead to conclusions that are of importance to the users. This set of *desired* conclusions represent the network interests and system requirements. We note that not all the combinations of facts lead to a desired conclusion. The relations between facts and conclusions are restricted to conjunctive expressions in propositional logic [4]. We assume each party shares a block of n facts, denoted by $x^n = (x_1, x_2 \dots, x_n)$ and $y^n = (y_1, y_2 \dots, y_n)$ for users 1 and 2, respectively. The indices $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$ for $i = 1, \dots, n$ refer to the i^{th} fact chosen from the sets \mathcal{X} and \mathcal{Y} . A subset of fact pairs $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$ lead to useful conclusions $c_i \in \mathcal{C}$. User 1 wants to learn y^n and user 2 wants to learn x^n , as long as the following condition is satisfied:

$$x_i \wedge y_i \rightarrow c_i, \quad c_i \in \mathcal{C}, \quad i = 1, \dots, n \quad (1)$$

That is, the two parties want to learn the facts of the other as long as each fact from one person results in a meaningful conclusion with the corresponding fact from the other user. We assume that each fact x_i from user 1 yields a desired conclusion with at most λ_{max} facts y_i from user 2, as illustrated in Fig. 1. Similarly, each fact y_i leads to a useful conclusion with at most μ_{max} facts x_i from user 1.

We center our discussion on the source coding problem and assume that a noiseless channel exists between the two parties. Let us define the support set of the facts that lead to useful conclusions by:

$$\mathcal{S}^n = \{(x^n, y^n) : x_i \wedge y_i \rightarrow c_i, x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n, c_i \in \mathcal{C}, i = 1, \dots, n\} \quad (2)$$

where \mathcal{S}^n is a subset of $\mathcal{X}^n \times \mathcal{Y}^n$ in which every fact-tuple leads to a sequence of desired conclusions, and \mathcal{C} is the finite discrete set of conclusions. Define the ambiguity set $\mathcal{I}_X(x^n)$ of sequence x^n for user 1:

$$\mathcal{I}_X(x^n) = \{y^n : x_i \wedge y_i \rightarrow c_i, y^n \in \mathcal{Y}^n, c_i \in \mathcal{C}, i = 1, \dots, n\} \quad (3)$$

which is the set of n -tuples for which each y_i leads to a conclusion with the corresponding x_i . The number of sequences $y^n \in \mathcal{Y}^n$ from user 2 that leads to a conclusion with a given x^n is $\lambda(x^n)$, or $|\mathcal{I}_X(x^n)| = \lambda(x^n)$. The ambiguity set of a given y^n for user 2 is:

$$\mathcal{I}_Y(y^n) = \{x^n : x_i \wedge y_i \rightarrow c_i, x^n \in \mathcal{X}^n, c_i \in \mathcal{C}, i = 1, \dots, n\} \quad (4)$$

where $y^n \in \mathcal{Y}^n$ and $|\mathcal{I}_Y(y^n)| = \mu(y^n)$.

Consider a mapping ϕ for encoding the fact pairs from \mathcal{S}^n to codewords of binary bit streams. A sequence of codewords $\phi(x^n, y^n) = [\phi_i(x^n, y^n)]_{i=1}^r$ is exchanged during an r -round communication with $\phi_i(x^n, y^n)$ denoting the codeword from round i :

$$\phi_i(x^n, y^n) = [\phi_i^X(x^n, y^n), \phi_i^Y(x^n, y^n)] \quad (5)$$

where $\phi_i^X(x^n, y^n)$ and $\phi_i^Y(x^n, y^n)$ are the codewords transmitted from users 1 and 2 in round i , respectively. The sequences $\phi^X(x^n, y^n) = [\phi_i^X(x^n, y^n)]_{i=1}^r$ and $\phi^Y(x^n, y^n) = [\phi_i^Y(x^n, y^n)]_{i=1}^r$ denote the codewords transmitted in r rounds. Sources can transmit arbitrary length codewords simultaneously. Null transmissions are allowed at any round. We define the worst-case codeword length for mapping ϕ given a sequence of n facts as:

$$l_\phi = \max_{(x^n, y^n) \in \mathcal{S}^n} \frac{1}{n} |\phi(x^n, y^n)|, \quad \text{bits/fact} \quad (6)$$

The best encoding strategy in terms of the maximal length codeword is then obtained by:

$$l = \min_{\phi} l_\phi \quad (7)$$

Lemma 1. [6] *Let $(x^n, y^n), (\bar{x}^n, y^n), (\bar{x}^n, \bar{y}^n) \in \mathcal{S}^n$ be fact pairs such that $x^n, \bar{x}^n \in \mathcal{X}^n$, $x^n \neq \bar{x}^n$ and $y^n, \bar{y}^n \in \mathcal{Y}^n$, $y^n \neq \bar{y}^n$. If $\phi(x^n, y^n)$ or $\phi(\bar{x}^n, \bar{y}^n)$ is a prefix of the other, then $\phi(x^n, y^n) = \phi(\bar{x}^n, y^n) = \phi(\bar{x}^n, \bar{y}^n)$.*

Proposition 1. *The set of codewords in the ambiguity set $\mathcal{I}_X(x^n)$ for each $x^n \in \mathcal{X}^n$ is prefix-free. Similarly for each $y^n \in \mathcal{Y}^n$, codewords in $\mathcal{I}_Y(y^n)$ are prefix-free.*

This proposition follows from Lemma 1 and the property that codewords corresponding to the facts in the ambiguity set have to be prefix-free for the receiver to interpret them correctly. That is, when $(x^n, y^n) \in \mathcal{S}^n$ and $(x^n, \bar{y}^n) \in \mathcal{S}^n$, then $[\phi_i^X(x^n, y^n)]_{i=1}^{k-1} = [\phi_i^X(x^n, \bar{y}^n)]_{i=1}^{k-1}$ implies that neither $\phi_i^Y(x^n, y^n)$ nor $\phi_i^Y(x^n, \bar{y}^n)$ is a prefix of the other in order for the first person to correctly distinguish between y^n and \bar{y}^n . Similarly, when $(x^n, y^n) \in \mathcal{S}^n$ and $(\bar{x}^n, y^n) \in \mathcal{S}^n$, then if $[\phi_i^Y(x^n, y^n)]_{i=1}^{k-1} = [\phi_i^Y(\bar{x}^n, y^n)]_{i=1}^{k-1}$, then neither $\phi_i^X(x^n, y^n)$ nor $\phi_i^X(\bar{x}^n, y^n)$ is a prefix of the other for the second person to distinguish between x^n and \bar{x}^n .

3 Lower Bound on Message Length for Multiple Facts

In this section, we present the lower bound on the worst-case message length.

Theorem 1. *The worst-case codeword length for the mapping ϕ is given as follows:*

$$l \geq \max_{\substack{(\bar{x}^n, y^n) \in \mathcal{S}^n \\ (x^n, \bar{y}^n) \in \mathcal{S}^n}} \left\{ \log \lambda_{max} + \frac{1}{n} \log \mu(y^n), \frac{1}{n} \log \lambda(x^n) + \log \mu_{max} \right\} \text{ bits/fact} \quad (8)$$

Proof. Form a bipartite graph $G = (V, U, E)$ with vertices in V and U being the n -tuples $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, respectively. Define an edge (x^n, y^n) between x^n and y^n if:

$$x_i \wedge y_i \rightarrow c_i, \quad c_i \in \mathcal{C}, \quad \forall i \in \{1, \dots, n\}. \quad (9)$$

It follows from (9) that the maximum degree for each vertex x^n (y^n) is at most λ_{max}^n (μ_{max}^n). From Proposition 1, for any given ϕ , we have for all $(x^n, y^n) \in \mathcal{S}^n$ that $|\phi^X(x^n, y^n)| \geq \lceil \log \lambda(x^n) \rceil$ and $|\phi^Y(x^n, y^n)| \geq \lceil \log \mu(y^n) \rceil$. Hence, for any given ϕ , for all $(x^n, y^n) \in \mathcal{S}^n$:

$$|\phi^X(x^n, y^n)| + |\phi^Y(x^n, y^n)| \geq \lceil \log \lambda(x^n) \rceil + \lceil \log \mu(y^n) \rceil \quad (10)$$

We can choose a ϕ that minimizes (10). The worst-case codeword length is then given by:

$$l = \min_{\phi} \max_{(x^n, y^n) \in \mathcal{S}^n} \frac{1}{n} |\phi(x^n, y^n)| \quad (11)$$

$$\geq \max_{(x^n, y^n) \in \mathcal{S}^n} \min_{\phi} \frac{1}{n} |\phi(x^n, y^n)| \quad (12)$$

$$= \max_{(x^n, y^n) \in \mathcal{S}^n} \min_{\phi} \frac{1}{n} (|\phi^X(x^n, y^n)| + |\phi^Y(x^n, y^n)|) \quad (13)$$

$$\geq \max_{(x^n, y^n) \in \mathcal{S}^n} \frac{1}{n} (\lceil \log \lambda(x^n) \rceil + \lceil \log \mu(y^n) \rceil) \quad (14)$$

$$\geq \max_{\substack{(\bar{x}^n, y^n) \in \mathcal{S}^n \\ (x^n, \bar{y}^n) \in \mathcal{S}^n}} \frac{1}{n} \left\{ \lceil \log \lambda_{max}^n \rceil + \lceil \log \mu(y^n) \rceil, \lceil \log \lambda(x^n) \rceil + \lceil \log \mu_{max}^n \rceil \right\} \quad (15)$$

$$\geq \max_{\substack{(\bar{x}^n, y^n) \in \mathcal{S}^n \\ (x^n, \bar{y}^n) \in \mathcal{S}^n}} \left\{ \log \lambda_{max} + \frac{1}{n} \log \mu(y^n), \frac{1}{n} \log \lambda(x^n) + \log \mu_{max} \right\} \quad (16)$$

where \bar{x}^n denotes a vertex with degree λ_{max}^n and \bar{y}^n is a vertex with degree μ_{max}^n . We define $\mu(y^n)$ and $\lambda(x^n)$ as the degrees of y^n or x^n that are adjacent to \bar{x}^n and \bar{y}^n , respectively. Then (12) follows from the min-max inequality, (13) from the definition of $|\phi(x^n, y^n)|$, and (14) is from (10). \square

4 Upper Bound on Message Length for Multiple Facts with Semantic Relations

In this section, we provide an upper bound on the worst case message length. To do so, we first briefly review some notable results on hypergraph partitioning and product graphs as applied to our problem.

Lemma 2. [7] *Assume that for all $x^n \in \mathcal{X}^n$, $|\mathcal{I}_X(x^n)| \leq \alpha$ and for all $y^n \in \mathcal{Y}^n$, $|\mathcal{I}_Y(y^n)| \leq \beta$. The worst-case message length is then bounded by:*

$$l \leq \lceil \log(\alpha \cdot \beta) \rceil + \lceil \log \min(\alpha, \beta) \rceil \quad (17)$$

Lemma 3. [8] Define $\Gamma = (V, E)$ to be a hypergraph with a vertex set of size $|V|$, and the hyperedges $E_m \subseteq V$ with $m = 1, \dots, |E|$. Assume that each hyperedge has at most κ elements, i.e., $|E_i| \leq \kappa$. Then for a given $\epsilon > 0$, there exists a constant $c(\epsilon)$ such that $\forall s \geq (\ln \sqrt{|V||E|})^{1+\epsilon}$ and $s > 1$, a partition $V_1, V_2, \dots, V_{\lceil \frac{\kappa}{s} c(\epsilon) \rceil}$ of the vertex set V can be found with $|V_k \cap E_m| < s$ for $m = 1, \dots, |E|$ and $k = 1, \dots, \lceil \frac{\kappa}{s} c(\epsilon) \rceil$.

Definition 1. (Strong Product) [9] Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs. The strong product $G_1 \boxtimes G_2$ of G_1 and G_2 is represented by the graph $\Gamma = (V, E)$. Its vertex set is given by $V = V_1 \times V_2$. Let $v = (x, y)$ and $v' = (x', y')$ be two distinct vertices $v, v' \in V$. An edge (v, v') exists between v and v' if and only if $(x, x') \in E_1$ and $y = y'$, or $x = x'$ and $(y, y') \in E_2$, or $(x, x') \in E_1$ and $(y, y') \in E_2$.

The strong product of graphs represents a system in which users are interested in sharing any block of facts so long as each fact in the block leads to a conclusion with the corresponding fact from the other user. Let the chromatic numbers of G_1 and G_2 be $\chi(G_1)$ and $\chi(G_2)$, respectively. Then, the chromatic number $\chi(\Gamma)$ of $\Gamma = G_1 \boxtimes G_2$, satisfies the following inequality [9]:

$$\max(\chi(G_1), \chi(G_2)) \leq \chi(\Gamma) \leq \chi(G_1) \cdot \chi(G_2) \quad (18)$$

We now present our main result of this section.

Theorem 2. The upper bound on the maximum code length for the two-way semantic network when each user shares a block of n facts, with the condition that the facts are to be recovered only if each of the n pairs leads to a conclusion is:

$$l \leq \lceil \log \lambda_{max} \rceil + \lceil \log \mu_{max} \rceil + \frac{(1+\epsilon)}{n} \log \log \sqrt{\chi(\mathcal{G}_X)\chi(\mathcal{G}_Y)} + (1+\epsilon) \frac{\log n}{n} + \frac{2}{n} \log c(\epsilon) + \frac{5}{n} \text{ bits/fact} \quad (19)$$

Proof. Define a characteristic graph $\mathcal{G}_X = (V_X, E_X)$ for user 1 with vertices $V_X = \{x_i : x_i \wedge y_i \rightarrow c_i, x_i \in \mathcal{X}, y_i \in \mathcal{Y}, c_i \in \mathcal{C}\}$ and an edge (x_i, \bar{x}_i) if there exists a fact y_i such that $x_i \wedge y_i \rightarrow c_i$, $\bar{x}_i \wedge y_i \rightarrow \bar{c}_i$ and $c_i \neq \bar{c}_i$. The chromatic number of \mathcal{G}_X is $\chi(\mathcal{G}_X)$. Similarly graph $\mathcal{G}_Y = (V_Y, E_Y)$ for user 2 has vertex set $V_Y = \{y_i : x_i \wedge y_i \rightarrow c_i, x_i \in \mathcal{X}, y_i \in \mathcal{Y}, c_i \in \mathcal{C}\}$ and an edge (y_i, \bar{y}_i) if there exists x_i such that $x_i \wedge y_i \rightarrow c_i$, $x_i \wedge \bar{y}_i \rightarrow \bar{c}_i$ and $c_i \neq \bar{c}_i$. An edge refers to two nodes that produce different conclusions when paired with some identical fact from the other user, and therefore need to be distinguished. The independent sets of the characteristic graph, on the other hand, represent the facts that produce the same conclusions with any fact from the other user. We assume that the two users want to avoid using extra resources for sharing the facts that lead to the same conclusions. Next, we define the following mapping:

$$g(x_i \wedge y_i) = \begin{cases} 1 & \text{if } x_i \wedge y_i \rightarrow c_i, x_i \in \mathcal{X}, y_i \in \mathcal{Y}, c_i \in \mathcal{C} \\ 0 & \text{o.w.} \end{cases} \quad (20)$$

for $i = 1, \dots, n$. Users 1 and 2 want to learn y^n and x^n , respectively, if the following condition is satisfied:

$$g(x_1 \wedge y_1) \wedge g(x_2 \wedge y_2) \dots \wedge g(x_n \wedge y_n) = 1 \quad (21)$$

Consider a bipartite graph $G = (V, U, E)$ as in Section 3 with vertices $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, and an edge (x^n, y^n) if (21) is satisfied. We know that the maximum degree of each x^n and y^n is no greater than λ_{max}^n and μ_{max}^n , respectively. Next, we define the n -fold characteristic graphs \mathcal{G}_X^n and \mathcal{G}_Y^n for the first and the second users. The vertices for \mathcal{G}_X^n are the n -tuples $x^n \in \mathcal{X}^n$, and edge (x^n, \bar{x}^n) exists whenever $x_i \wedge y_i \rightarrow c_i$, $\bar{x}_i \wedge y_i \rightarrow \bar{c}_i$, and $c_i \neq \bar{c}_i$ for $i = 1, \dots, n$ for some $y^n \in \mathcal{Y}^n$. For the second person, define \mathcal{G}_Y^n with vertices $y^n \in \mathcal{Y}^n$. The edge (y^n, \bar{y}^n) exists if and only if $x_i \wedge y_i \rightarrow c_i$, $x_i \wedge \bar{y}_i \rightarrow \bar{c}_i$, and $c_i \neq \bar{c}_i$ for $i = 1, \dots, n$ for some $x^n \in \mathcal{X}^n$. We see from condition (21) and Definition 1 that \mathcal{G}_X^n can be defined as the n -fold strong product of \mathcal{G}_X , whereas \mathcal{G}_Y^n can be defined as the n -fold strong product of \mathcal{G}_Y .

Define a minimum coloring of the characteristic graph \mathcal{G}_X^n by the set of colors $q_X \in \{1, \dots, \chi(\mathcal{G}_X^n)\}$. Similarly, for \mathcal{G}_Y^n define a coloring by $q_Y \in \{1, \dots, \chi(\mathcal{G}_Y^n)\}$. Let $q(x^n)$ and $q(y^n)$ denote the colors assigned to x^n and y^n . Let the ambiguity set for color q_X be:

$$\mathcal{J}_X(q_X) = \{q_Y : x_i \wedge y_i \rightarrow c_i, q(x^n) = q_X, q(y^n) = q_Y \in \{1, \dots, \chi(\mathcal{G}_Y^n)\}, x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n, c^n \in \mathcal{C}^n\} \quad (22)$$

Similarly, denote the ambiguity set $\mathcal{J}_Y(q_Y)$ for each color q_Y as:

$$\mathcal{J}_Y(q_Y) = \{q_X : x_i \wedge y_i \rightarrow c_i, q(x^n) = q_X \in \{1, \dots, \chi(\mathcal{G}_X^n)\}, q(y^n) = q_Y, x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n, c^n \in \mathcal{C}^n\} \quad (23)$$

By definition of the strong product and the construction of the characteristic graph, the cardinality of the ambiguity sets is bounded by $|\mathcal{J}_X(q_X)| \leq \lambda_{max}^n$, and $|\mathcal{J}_Y(q_Y)| \leq \mu_{max}^n$. Another way to see this is the fact that the number of elements in each ambiguity set cannot be greater than the maximum degree of each vertex of the bipartite graph $G = (V, U, E)$.

The theorem follows from Lemmas 2 and 3 by utilizing hypergraph partitioning [8]. Let $s = (\ln \sqrt{\chi(\mathcal{G}_X^n)\chi(\mathcal{G}_Y^n)})^{1+\epsilon}$. Partition the color set of each n -product graph \mathcal{G}_X^n and \mathcal{G}_Y^n by using Lemma 3 with the condition that in each partition, the number of colors from the ambiguity set is no greater than s . User 1 then sends the index of the partition the color of her fact lies in. This requires at most $\lceil \log(\frac{\mu_{max}^n}{s} c(\epsilon)) \rceil$ bits. User 2 also sends the index of the partition that the color of his fact is in by using no more than $\lceil \log(\frac{\lambda_{max}^n}{s} c(\epsilon)) \rceil$ bits. User 1 then uses the received index and her color to restrict the possible colors from user 2 in a s -dimensional subspace. Through a similar elimination, user 2 leaves at most s possible colors from user 1. This limits the communication to an at most $s \times s$ dimensional subspace. Then, from Lemma 2 with $\alpha = \beta = s$, no more than $3\lceil \log(s) \rceil$ bits are necessary for both parties to learn both sequences. The maximum number of bits required is then bounded by:

$$l \leq \frac{1}{n} (\lceil \log(\frac{\mu_{max}^n}{s} c(\epsilon)) \rceil + \lceil \log(\frac{\lambda_{max}^n}{s} c(\epsilon)) \rceil + 3\lceil \log(s) \rceil) \quad (24)$$

$$\leq \frac{1}{n} ((\log(\frac{\mu_{max}^n}{s} c(\epsilon)) + 1) + (\log(\frac{\lambda_{max}^n}{s} c(\epsilon)) + 1) + 3(\log(s) + 1)) \quad (25)$$

$$\leq \frac{1}{n} \left(\log \lambda_{max}^n + \log \mu_{max}^n + (1 + \epsilon) \log \ln \sqrt{\chi(\mathcal{G}_X^n)\chi(\mathcal{G}_Y^n)} + 2 \log c(\epsilon) + 5 \right) \quad (26)$$

$$\leq \lceil \log \lambda_{max} \rceil + \lceil \log \mu_{max} \rceil + \frac{1}{n} (1 + \epsilon) \log \log \sqrt{\chi(\mathcal{G}_X^n)\chi(\mathcal{G}_Y^n)} + \frac{2}{n} \log c(\epsilon) + \frac{5}{n} \quad (27)$$

Applying (18) recursively on the partitions $\mathcal{G}_X^i = \mathcal{G}_X \boxtimes \mathcal{G}_X^{i-1}$ for $i = 1, \dots, n$ bounds the chromatic number of \mathcal{G}_X^n by $\chi(\mathcal{G}_X^n) \leq \chi(\mathcal{G}_X)^n$. Similarly, we have $\chi(\mathcal{G}_Y^n) \leq \chi(\mathcal{G}_Y)^n$.

Thus, the worst-case message length is bounded above as:

$$l \leq \lceil \log \lambda_{max} \rceil + \lceil \log \mu_{max} \rceil + \frac{1}{n}(1 + \epsilon) \log \log \sqrt{\chi(\mathcal{G}_X)^n \chi(\mathcal{G}_Y)^n} + \frac{2}{n} \log c(\epsilon) + \frac{5}{n} \quad (28)$$

from which the theorem follows. We assume that the partitioning protocol is agreed upon by the two parties before the communication takes place. \square

Corollary 1. *For large block lengths, there exists graphs for which the upper and lower bounds are tight.*

$$l \rightarrow \lceil \log \lambda_{max} \rceil + \lceil \log \mu_{max} \rceil \quad (29)$$

Proof. From Theorem 2, the upper bound for large sequences satisfies:

$$\begin{aligned} \lim_{n \rightarrow \infty} l &\leq \lim_{n \rightarrow \infty} \left(\lceil \log \lambda_{max} \rceil + \lceil \log \mu_{max} \rceil + \frac{(1 + \epsilon)}{n} \log \log \sqrt{\chi(\mathcal{G}_X) \chi(\mathcal{G}_Y)} + (1 + \epsilon) \frac{\log n}{n} \right. \\ &\quad \left. + \frac{2}{n} \log c(\epsilon) + \frac{5}{n} \right) = \lceil \log \lambda_{max} \rceil + \lceil \log \mu_{max} \rceil \end{aligned}$$

Consider a graph with a color pair (q_X, q_Y) such that $q_X \in \mathcal{J}_Y(q_Y)$, $q_Y \in \mathcal{J}_X(q_X)$, and that $|\mathcal{J}_Y(q_Y)| = \mu_{max}^n$ and $|\mathcal{J}_X(q_X)| = \lambda_{max}^n$. Applying Theorem 1 gives the lower bound $l \geq \lceil \log \lambda_{max} \rceil + \lceil \log \mu_{max} \rceil$, hence the lower and upper bounds are tight. \square

5 Conclusions with Varying Priorities and the Role of Ignorance

We centered our discussion so far on the assumption that all conclusions are equally important, which is rarely the case in real life applications. In a semantic network, a more sensible approach is to assume that some conclusions have higher priorities than the others, and some are less critical. Consider the conclusions $c \leftrightarrow \text{Fire!}$ and $c' \leftrightarrow \text{Alice has a blue car}$. Although the exact priorities are system-dependent, in most situations the first conclusion carries a higher risk when ignored, hence its priority is higher than the latter. We here study the performance limits of the message length for lossless recovery of conclusions with *unequal priorities*. Define $p(c)$ to be the weight representing the *degree of priority* given to conclusion c , or the probability of the risk associated with ignoring c . Equally, $1 - p(c)$ can be interpreted as the tendency of ignoring c . The priority weights are normalized to yield a valid probability distribution which satisfy $\sum_{c \in \mathcal{C}} p(c) = 1$ and $p(c) \geq 0$ for all $c \in \mathcal{C}$. Then we can represent the conclusions by a random variable $C \sim p(c)$. The conclusions are assumed to be independent from each other, and the weight of a sequence c_1, \dots, c_n is:

$$p(c_1, \dots, c_n) = p(c^n) = \prod_{i=1}^n p(c_i), \quad c_i \in \mathcal{C} \quad (30)$$

User 1 produces a sequence x_1, \dots, x_n whereas user 2 has y_1, \dots, y_n , such that $x_i \wedge y_i \rightarrow c_i$, $c_i \in \mathcal{C}$ in which c_1, \dots, c_n represent the conclusions to be recovered by both users. Users are interested in sharing sequences for which each pair results in a useful conclusion. We assume in this section that fact pairs lead to distinct conclusions, so that no duplicates occur. We derive the upper and lower bounds in order for each user to recover the facts resulting to all but a small priority fraction of the conclusions.

Theorem 3. *The worst-case codeword length for the semantic network when conclusions have unequal priorities is:*

$$l \leq H(C|\bar{q}_Y) + H(C|\bar{q}_X) + \frac{(1+\epsilon)}{n} \log(H(C) + \epsilon) + \frac{(1+\epsilon)}{n} \log n + \frac{\sigma}{n} \quad (31)$$

where $\bar{q}_T = \arg \max_{q_T} H(C|q_T)$, $H(C|q_T)$ denotes the entropy of C given coloring q_T , and $\sigma = 5 + 2 \log c(\epsilon) + \epsilon' + \epsilon''$.

Proof. For any $\epsilon > 0$, we refer to the sequence c^n as ϵ -typical if:

$$\left| -\frac{1}{n} \log p(c_1, \dots, c_n) - H(C) \right| < \epsilon \quad (32)$$

where $H(C)$ is the entropy of C . Let A_ϵ^n be the set of typical sequences:

$$A_\epsilon^n(C) = \left\{ c^n : \left| -\frac{1}{n} \log p(c^n) - H(C) \right| < \epsilon \right\} \quad (33)$$

We introduce two ϵ -characteristic graphs, $\mathcal{G}_{X,\epsilon}^n$ and $\mathcal{G}_{Y,\epsilon}^n$. For user 1, let $\mathcal{G}_{X,\epsilon}^n = (V_X, E_X)$ be a graph with the vertex set $V_X = \mathcal{X}^n$. An edge (x^n, \bar{x}^n) exists if and only if $x_i \wedge y_i \rightarrow c_i$ and $\bar{x}_i \wedge y_i \rightarrow \bar{c}_i$, where $c_i \neq \bar{c}_i$ for $i = 1, \dots, n$ for some $y^n \in \mathcal{Y}^n$ and $c^n, \bar{c}^n \in A_\epsilon^n(C)$. Similarly, for user 2, $\mathcal{G}_{Y,\epsilon}^n = (V_Y, E_Y)$ has vertices $V_Y = \mathcal{Y}^n$. The edge (y^n, \bar{y}^n) exists if and only if $x_i \wedge y_i \rightarrow c_i$, $x_i \wedge \bar{y}_i \rightarrow \bar{c}_i$, and $c_i \neq \bar{c}_i$ for $i = 1, \dots, n$ for some $x^n \in \mathcal{X}^n$ and $c^n, \bar{c}^n \in A_\epsilon^n(C)$. The chromatic numbers of the characteristic graphs give the minimum number of colors required to represent the facts that lead to the conclusions in the typical set. The colors of user 1 are given by $q_X \in \{1, \dots, \chi(\mathcal{G}_{X,\epsilon}^n)\}$ and for user 2 by $q_Y \in \{1, \dots, \chi(\mathcal{G}_{Y,\epsilon}^n)\}$. Define $A_{\epsilon'}^n(C|q_X)$ to be the typical set of c^n sequences for a particular color q_X from user 1:

$$A_{\epsilon'}^n(C|q_X) = \left\{ c^n : \left| -\frac{1}{n} \log p(c^n|q_X) - H(C|q_X) \right| < \epsilon' \right\} \quad (34)$$

Let $A_{\epsilon''}^n(C|q_Y)$ be the typical set of c^n sequences for a particular q_Y from user 2:

$$A_{\epsilon''}^n(C|q_Y) = \left\{ c^n : \left| -\frac{1}{n} \log p(c^n|q_Y) - H(C|q_Y) \right| < \epsilon'' \right\} \quad (35)$$

We choose n sufficiently large such that the total probability of the conclusions in the typical sets satisfy $P(A_\epsilon^n(C)) > 1 - \epsilon$, $P(A_{\epsilon'}^n(C|q_X)) > 1 - \epsilon'$ and $P(A_{\epsilon''}^n(C|q_Y)) > 1 - \epsilon''$ for arbitrary $\epsilon, \epsilon', \epsilon'' > 0$. We know that the number of elements in the typical set satisfies $|A_\epsilon^n(C|q_X)| \geq (1 - \epsilon')2^{n(H(C|q_X) - \epsilon')}$ and $|A_{\epsilon''}^n(C|q_Y)| \geq (1 - \epsilon'')2^{n(H(C|q_Y) - \epsilon'')}$. Moreover, $|A_{\epsilon'}^n(C|q_X)| \leq 2^{n(H(C|q_X) + \epsilon')}$ and $|A_{\epsilon''}^n(C|q_Y)| \leq 2^{n(H(C|q_Y) + \epsilon'')}$. Lastly, $|A_\epsilon^n(C)| \geq (1 - \epsilon)2^{n(H(C) - \epsilon)}$ and $|A_\epsilon^n(C)| \leq 2^{n(H(C) + \epsilon)}$. We then apply Lemma 3 on the hypergraph $H = (V, E)$ where $|V| = |\mathcal{G}_{X,\epsilon}^n|$, $|E| = |\mathcal{G}_{Y,\epsilon}^n|$ and $\kappa = \max_{q_X} 2^{n(H(C|q_X) + \epsilon')}$ for user 1. Similarly, for user 2 we define an hypergraph with $|V| = |\mathcal{G}_{Y,\epsilon}^n|$, $|E| = |\mathcal{G}_{X,\epsilon}^n|$

and $\kappa = \max_{q_Y} 2^{n(H(C|q_Y)+\varepsilon'')}$. Let $p = (\ln \sqrt{|\mathcal{G}_{X,\varepsilon}^n||\mathcal{G}_{Y,\varepsilon}^n|})^{1+\varepsilon}$. Applying the hypergraph partitioning protocol, the message length can be bounded by:

$$l \leq \frac{1}{n} \log \max_{q_Y} 2^{n(H(C|q_Y)+\varepsilon'')} + \frac{1}{n} \log \max_{q_X} 2^{n(H(C|q_X)+\varepsilon')} + \frac{(1+\varepsilon)}{n} \log \ln \sqrt{|\mathcal{G}_{X,\varepsilon}^n||\mathcal{G}_{Y,\varepsilon}^n|} + \frac{2}{n} \log c(\varepsilon) + \frac{5}{n} \quad (36)$$

$$\begin{aligned} &\leq H(C|\bar{q}_Y) + H(C|\bar{q}_X) + \frac{(1+\varepsilon)}{n} \log \ln \sqrt{|\mathcal{G}_{X,\varepsilon}^n||\mathcal{G}_{Y,\varepsilon}^n|} + \frac{(\varepsilon' + \varepsilon'')}{n} + \frac{2}{n} \log c(\varepsilon) + \frac{5}{n} \\ &\leq H(C|\bar{q}_Y) + H(C|\bar{q}_X) + \frac{(1+\varepsilon)}{n} \log \log 2^{n(H(C)+\varepsilon)} + \frac{(\varepsilon' + \varepsilon'')}{n} + \frac{2}{n} \log c(\varepsilon) + \frac{5}{n} \\ &\leq H(C|\bar{q}_Y) + H(C|\bar{q}_X) + \frac{(1+\varepsilon)}{n} \log (H(C) + \varepsilon) + \frac{(1+\varepsilon)}{n} \log n + \frac{\sigma}{n} \end{aligned} \quad (37)$$

when $\sigma = 5 + 2 \log c(\varepsilon) + \varepsilon' + \varepsilon''$. In other words, we partition the vertex set in such a way that for each sequence from a given user, the number of sequences that result in a sequence of conclusions in the typical set is bounded by the maximum number of elements in the typical set for a given color. \square

Corollary 2. *The upper bound on the worst-case message length for large blocks is:*

$$l \leq H(C|\bar{q}_Y) + H(C|\bar{q}_X) \quad (38)$$

Proof. The proof follows directly from (31) as $\lim_{n \rightarrow \infty} l \leq H(C|\bar{q}_Y) + H(C|\bar{q}_X)$. \square

This coding scheme can achieve an arbitrarily small error probability as the block length increases. In effect, a sequence of facts chosen by one user will not be recovered by the other either when it corresponds to a conclusion that is not in the latter user's typical set, or when the sequence pairs do not lead to the ε -typical conclusions. Define the event E_1 such that q_X and q_Y lead to a $c^n \notin A_\varepsilon^n(C)$. Let E_2 and E_3 refer to the pair q_X and q_Y for which $c^n \in A_\varepsilon^n(C)$, but $c^n \notin A_\varepsilon^n(C|q_X)$, or $c^n \notin A_\varepsilon^n(C|q_Y)$, respectively. Then the probability of error for not recovering the desired conclusions for a given pair of sequences can be bounded by $P_\varepsilon = P(E_1 \cup E_2 \cup E_3) \leq P(E_1) + P(E_2) + P(E_3) \leq \varepsilon + \varepsilon' + \varepsilon''$.

Theorem 4. *The lower bound on the message length with varying priorities is:*

$$l \geq \max_{\substack{(\bar{q}_X, q_Y) \in \mathcal{S} \\ (q_X, \bar{q}_Y) \in \mathcal{S}}} \left\{ H(C|\bar{q}_X) + \frac{\log \mu(q_Y)}{n} + \frac{\log(1-\varepsilon') - \varepsilon'}{n}, H(C|\bar{q}_Y) + \frac{\log \lambda(q_X)}{n} + \frac{\log(1-\varepsilon'') - \varepsilon''}{n} \right\} \quad (39)$$

where $\bar{q}_X = \arg \max_{q_X} H(C|q_X)$ and $\bar{q}_Y = \arg \max_{q_Y} H(C|q_Y)$.

Proof. For each person, define $\mu(q_Y)$ and $\lambda(q_X)$ as the number of colors from the other person that lead to a typical conclusion sequence with q_Y and q_X , respectively. By applying the lower bound on the number of elements of the typical sets $\mathcal{A}_\varepsilon^n(C|q_X)$ and $\mathcal{A}_{\varepsilon''}^n(C|q_Y)$, the message length can be bounded as:

$$l \geq \max_{\substack{(\bar{q}_X, q_Y) \in \mathcal{S} \\ (q_X, \bar{q}_Y) \in \mathcal{S}}} \frac{1}{n} \left\{ \lceil \log(1-\varepsilon') 2^{n(H(C|\bar{q}_X)-\varepsilon')} \rceil + \lceil \log \mu(q_Y) \rceil, \lceil \log \lambda(q_X) \rceil + \lceil \log(1-\varepsilon'') 2^{n(H(C|\bar{q}_Y)-\varepsilon'')} \rceil \right\} \quad (40)$$

$$\geq \max_{(\bar{q}_X, q_Y) \in \mathcal{S}, (q_X, \bar{q}_Y) \in \mathcal{S}} \left\{ H(C|\bar{q}_X) + \frac{\log \mu(q_Y)}{n} + \frac{\log(1 - \varepsilon') - \varepsilon'}{n}, \right. \\ \left. H(C|\bar{q}_Y) + \frac{\log \lambda(q_X)}{n} + \frac{\log(1 - \varepsilon'') - \varepsilon''}{n} \right\} \quad (41)$$

where $\bar{q}_X = \arg \max_{q_X} H(C|q_X)$ and $\bar{q}_Y = \arg \max_{q_Y} H(C|q_Y)$ and \mathcal{S} denotes the support set of the color pairs that lead to the ε -typical conclusion sequences. \square

Corollary 3. *For large block lengths, the message length is bounded below by:*

$$l \geq \max_{(\bar{q}_X, q_Y) \in \mathcal{S}, (q_X, \bar{q}_Y) \in \mathcal{S}} \left\{ H(C|\bar{q}_X) + \frac{\log \mu(q_Y)}{n}, H(C|\bar{q}_Y) + \frac{\log \lambda(q_X)}{n} \right\} \quad (42)$$

Theorem 5. *There exists graphs such that the upper and lower bounds are tight, and the message length is:*

$$l \rightarrow H(C|\bar{q}_X) + H(C|\bar{q}_Y) \quad (43)$$

Proof. Consider a graph with $(\bar{q}_X, \bar{q}_Y) \in \mathcal{S}$. Then the lower bound on the message length satisfies $l \geq H(C|\bar{q}_X) + H(C|\bar{q}_Y)$ from which, with Corollary 2, the statement follows. \square

6 Conclusion

In this paper, we have studied a two-way interactive network through which parties share semantic data. According to the network interests, some facts yield desired conclusions. The two parties are interested in sharing only the sequences that lead to a high probability set of desired conclusions. We provide the compression limits of the worst-case message length by utilizing logical relationships between facts and conclusions with varying priorities. Future directions include semantic source coding with multiple sources and more complex logical relations.

References

- [1] R. Carnap, Y. Bar-Hillel, "An outline of a theory of semantic information," *RLE Technical Reports*, vol. 247, MIT, Cambridge, MA, Oct. 1952.
- [2] L. Floridi, "Outline of a theory of strongly semantic information," *Minds Mach.*, vol. 14, no. 2, pp. 197-221, 2004.
- [3] F. M. Willems, T. Kalker, "Semantic compaction, transmission, and compression codes," *Proc. of Int. Symp. on Inf. Theory (ISIT)*, 2005, pp. 214-218.
- [4] P. Basu, J. Bao, M. Dean, J. A. Hendler, "Preserving quality of information by using semantic relationships," *PerCom Workshops*, 58-63, 2012.
- [5] B. Guler, A. Yener and P. Basu, "A Study of Semantic Data Compression," *Proc. of the IEEE GlobalSIP Symp. on Network Theory*, Austin, TX, Dec. 2013.
- [6] A. Orłitsky, "Worst-case interactive communication I: Two messages are almost optimal," *IEEE Trans. on Inf. Theory*, vol. 36, no. 5, Sept. 1990.
- [7] R. Ahlswede, "Coloring hypergraphs: A new approach to multi-user coding," *Journal of Combinatorics, Information & System Sciences*, vol. 4, no. 1, pp. 76-115.
- [8] A. El Gamal, A. Orłitsky, "Interactive data compression," *25th Annual Symposium on Foundations of Computer Science*, pp. 100-108, 24-26 Oct. 1984.
- [9] C. Berge, "Graphs and hypergraphs," North Holland, 1973.