

# Expert Graphs: Synthesizing New Expertise via Collaboration

Bijan Mazaheri<sup>†</sup>, Siddharth Jain<sup>\*</sup>, and Jehoshua Bruck<sup>\*</sup>

<sup>\*</sup>Electrical Engineering, California Institute of Technology, U.S.A. {sidjain,bruck}@caltech.edu

<sup>†</sup>Computing & Mathematical Sciences, California Institute of Technology, U.S.A. bmazaher@caltech.edu

**Abstract**—Consider multiple experts with overlapping expertise working on a classification problem under uncertain input. What constitutes a consistent set of opinions? How can we predict the opinions of experts on missing sub-domains? In this paper, we define a framework of to analyze this problem, termed “expert graphs.” In an expert graph, vertices represent classes and edges represent binary opinions on the topics of their vertices. We derive necessary conditions for expert graph validity and use them to create “synthetic experts” which describe opinions consistent with the observed opinions of other experts. We show this framework to be equivalent to the well-studied linear ordering polytope. We show our conditions are not sufficient for describing all expert graphs on cliques, but are sufficient for cycles.

## I. INTRODUCTION

The merits of specialization have distributed human knowledge across many overlapping silos. Accessing this network of human knowledge requires soliciting opinions from experts with different overlapping expertise - a daunting experience fraught with inconsistency and confusion. This paper will attempt to explain how this perceived inconsistency can emerge from uncertain inputs. From this understanding, we will establish new rules for “consistency,” which we can use to synthesize opinions in areas of expertise that we may not have access to.

We will illustrate our setting with an example from medicine. Consider a patient suffering from a cough who is deciding what kind of doctor to see. There may be a number of reasons for the cough including pollen allergies, a virus, or perhaps cancer. In addition, we will consider uncertainty in the form of random variable input. For example, our patient may not know whether he or she has been exposed to asbestos – a significant risk factor in lung cancer. Each state of this random variable can be associated with a *categorical distribution* given by a vector of probabilities for each disease. For example,

	Pollen	Virus	Cancer
Asbestos	.2	.1	.7
No Asbestos	.2	.5	.3

Different specialists exist for subsets of these categories: a general practitioner may be an expert at distinguishing allergies from viruses, an oncologist may have experience with viruses and cancer, and an expert on environmental exposure

TABLE I

A table of probabilities for three possible input states  $u^{(1)}, u^{(2)}, u^{(3)}$  and three possible classes  $y^{(1)}, y^{(2)}, y^{(3)}$  which yields nontransitive preferences.

	$y^{(1)}$	$y^{(2)}$	$y^{(3)}$
$u^{(1)}$	.01	.90	.09
$u^{(2)}$	.09	.01	.90
$u^{(3)}$	.90	.09	.01

may know the difference between pollen-induced allergies and toxin-induced cancers.

Since none of these experts have studied all three causes, the probabilities in the above table are invisible to them. Instead, our experts have formed *situational opinions*, consisting of conditional probabilities. For example, the oncologist may have a sense of the percentage of her asbestos-exposed patients who are ultimately diagnosed with cancer, which is  $\frac{7}{8}$ . Similarly, she understands that only  $\frac{3}{8}$  of her unexposed patients develop cancer. Together, she may report an average of these situational opinions based on the perceived probabilities of asbestos exposure. This calculation may be unconscious or inaccessible (as is the case with many machine learning algorithms). Hence, we may only have access to the average of these situational opinions.

To see how inconsistencies can develop, consider Table I. Each column consists of one small (.01), one medium (.09) and one large (.90) probability, each of which dominates the previous probability in this order. Notice that the  $y^{(1)}$  column dominates the  $y^{(2)}$  column at  $u^{(2)}$  and  $u^{(3)}$ . Similarly,  $y^{(2)}$  dominates the  $y^{(3)}$  at  $u^{(1)}, u^{(3)}$ , and  $y^{(3)}$  dominates  $y^{(1)}$  at  $u^{(1)}, u^{(2)}$ . Due to the cyclic rotation of probabilities, the average opinions of all three possible experts return the same value:

$$\frac{1}{3} \left( \frac{.01}{.91} + \frac{.09}{.10} + \frac{.90}{.99} \right) > .6$$

Hence, this table yields a cycle of preference ( $y^{(1)} \leftarrow y^{(2)} \leftarrow y^{(3)} \leftarrow y^{(1)}$ ) despite having perfectly knowledgeable experts. No matter which class is chosen, one expert will vehemently assert that there is at least a 60% chance we are wrong! This is an example of what is known in voting theory as the Condorcet Paradox [14], which arises when voter preferences are arranged in a similarly cyclic fashion.

The unintuitive nature of combining expertise in the presence of uncertainty calls for a careful treatment of this problem. This paper will develop a framework for understanding networks of experts with and overlapping knowledge, called *expert graphs*.

### A. Related work

In voting theory, there is significant work on the ‘linear ordering polytope’ which corresponds to the set of possible networks of hypothetical pairwise elections given a population of rankings [7]. This field has separately developed what we call the ‘curl condition’ and shown that it is necessary and sufficient for  $n \leq 5$  classes, and just necessary for  $n \geq 6$  classes. In this setting, each voter can be thought of as a ‘state’ of uncertainty with an absolute ranking. Our problem introduces aleatoric uncertainty to this setting, resulting in soft preferences instead of binary rankings.

The task of synthesizing our experts’ knowledge is essentially that of a classifier built from already-trained pairwise classifiers. There are multiple frameworks that exist to use the outcome of pairwise comparisons to obtain multiclassifiers (see [1], [6]). The focus of this literature is on the *design* of classifiers for aggregation. We will instead consider the setting of *unengineered* binary classifiers, referred to as experts.

Related literature also exists in ensemble methods [4], [13], [15], which generally focus on combining experts trained on the *same* task, and whose knowledge is *imperfect* due to differing training datasets. In contrast, our experts are perfect, but trained on *different* tasks.

### B. Paper outline

- In Section II we establish notation and develop the *expert graph* framework.
- In Section III we define the *curl* and *curl consistent weighted digraphs* (CCWDs). We prove this condition is necessary for the set of feasible expert graphs.
- In Section IV we use the curl condition to give curl consistency bounds on missing opinions. We call these edges *synthetic experts*.
- In Section V we discuss a similar framework, linear ordering graphs (LOGs). We prove this framework is equivalent to the expert graph framework. We use this equivalence to prove when the curl condition is sufficient (or insufficient) for describing all possible expert graphs in various settings.
- In Section VI we conclude the paper and point to exciting future directions and applications.

## II. THE EXPERT GRAPH FRAMEWORK

### A. Notation

We will first provide notation used throughout the paper.

- $[\ell]$  is used to denote the set  $\{1, 2, \dots, \ell\}$  for any  $\ell \in \mathbb{N}$ .
- $|A|$  denotes the cardinality of set  $A$ .
- $\mathbb{1}[c]$  will be used for an indicator function which is 1 if condition  $c$  is met and 0 otherwise.
- Any bold symbol unless otherwise stated will be used to denote a vector.  $\mathbf{v} = (v_1, \dots, v_\ell)^T$  denotes a  $\ell$ -length vector. To keep track of indexing more easily, we will generally use superscripts to index class and subscripts

to index state. For example  $p_u^{(i)} = \Pr(y^{(i)} | u)$ ,  $\mathbf{p}_u = (p_u^{(1)}, \dots, p_u^{(n)})^T$  and  $\mathbf{p}^{(i)} = (p_{u^{(1)}}^{(i)}, \dots, p_{u^{(k)}}^{(i)})^T$ .

- $\mathbf{1}_\ell$  denotes an all 1 vector of size  $\ell$ .
- $\Delta^\ell$  will be used to denote vectors of size  $\ell$  which can be probability distributions. That is,  $\lambda \in \Delta^\ell$  iff  $\lambda \in [0, 1]^\ell$  and  $\mathbf{1}^T \lambda = 1$ .
- We use  $\prec, \succ, \preceq, \succeq$  to denote element-wise inequality. For example, we say  $\mathbf{w} \succeq \mathbf{v}$  if  $w_i \geq v_i \forall i \in [\ell]$ .
- The L1 norm of a vector  $\mathbf{v}$  is given by  $\|\mathbf{v}\|_1 = \sum_{i=1}^\ell |v_i|$  and the L2 norm is given by  $\|\mathbf{v}\|_2 = \sqrt{\sum_{i=1}^\ell |v_i|^2}$ .
- We will use  $\text{Co}(S)$  to denote the open convex hull of  $S$ ,  $\overline{\text{Co}}(S)$  to denote the closed convex hull, and  $\text{Bo}(\cdot)$  to denote the boundary.
- We will use the convention of lowercase letters denoting specific assignments to random variables. For example,  $\Pr(A | u) = \Pr(A | U = u)$ .

### B. Expert graphs

We consider a prediction setting with:

- $X$ : Known and understood input.
- $U$ : Unknown, but understood input.
- $Y$ : Labels of a classification problem.

This setting will contain two types of uncertainty. In our introductory example, we were uncertain of exposure to asbestos, which was relevant to our prediction task. We represent uncertain input by a random variable  $U$  that takes values in alphabet  $\mathcal{U} = [k]$  with probability distribution  $d(\cdot) : \mathcal{U} \mapsto \Delta^k$ .

Even given full knowledge of asbestos exposure and other factors ( $U$  and  $X$ ), we were not completely sure of the disease. Hence, our labels  $\mathcal{Y}$  are also represented by a random variable partially determined by  $U$ . Specifically, each label  $Y$  will have a categorical distribution over alphabet  $\mathcal{Y} = \{y^{(1)}, y^{(2)}, \dots, y^{(n)}\}$  given input  $u$ :

$$p_u^{(i)} = \Pr(Y = y^{(i)} | u, x) \quad (1)$$

**Assumption 1.** We assume for any  $u \in \mathcal{U}$ ,  $\mathbf{p}_u \succ 0$ .

We refer to these  $n \times k$  probabilities as the ‘probability table,’  $\mathbf{P}$  (see Table I for an example) and denote the vector of probabilities for a given input as  $\mathbf{p}_u \in \Delta^n$ . For the rest of the paper we will be considering the probabilities conditioned on the known input  $x$ , but we will omit the  $x$  from the conditioning for simplicity.

An oncologist’s perception of probabilities will be based on the patients they see and study daily - a population that is unlikely to have many patients only suffering from allergies. Hence, our experts do not have access to the probability table; they only understand the relative probabilities between the pair of classes in their *area of expertise*. For simplicity, we will consider a list of  $m$  binary experts  $E \subseteq \mathcal{Y} \times \mathcal{Y}$ .

**Definition 1.** Expert  $(y^{(i)}, y^{(j)}) \in E$  has access to **situational opinions**  $f_u(y^{(i)}, y^{(j)})$  for each  $u \in \mathcal{U}$  given by the following conditional probability:

$$\begin{aligned} f_u(y^{(i)}, y^{(j)}) &:= \Pr(Y = y^{(i)} | Y \in \{y^{(i)}, y^{(j)}\}, u) \\ &= \frac{p_u^{(i)}}{p_u^{(i)} + p_u^{(j)}} \end{aligned} \quad (2)$$

	$y^{(1)}$	$y^{(2)}$	$y^{(3)}$
$u^{(1)}$	.01	.90	.09
$u^{(2)}$	.09	.01	.90
$u^{(3)}$	.90	.09	.01

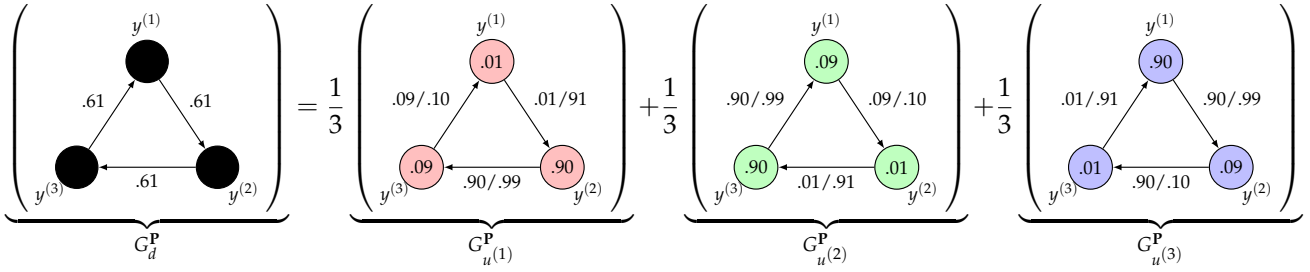


Fig. 1. The expert graph  $G_d^{\mathbf{P}}$  given by situational expert graphs with  $\mathbf{P}$  given by Table I (displayed again here) and  $d(\cdot)$  the uniform distribution. Probabilities from Table I are given inside vertices of situational expert graphs, and expert opinions  $f_u(y^{(i)}, y^{(j)})$  are given as weights on edge  $y^{(i)} \rightarrow y^{(j)}$ . The expert graph  $G_d^{\mathbf{P}}$  has edge-weights averaged over the situational expert graphs  $G_{u^{(1)}}^{\mathbf{P}}, G_{u^{(2)}}^{\mathbf{P}}, G_{u^{(3)}}^{\mathbf{P}}$ .

**Observation 2.** From assumption 1,  $f_u(e) \in (0, 1) \forall e \in E$ .

**Observation 3.**  $f_u(y^{(i)}, y^{(j)}) = 1 - f_u(y^{(j)}, y^{(i)})$ .

With this we can define useful building blocks, called *situational expert graphs*.

**Definition 2.** A **situational expert graph**  $G_u^{\mathbf{P}} = (\mathcal{Y}, E, f_u(\cdot))$  encodes experts' pairwise situational opinions  $f_u(e)$  as weights on directed edges  $e = (y^{(i)}, y^{(j)}) \in E$ .

In our introductory example, our experts reported average opinions over likely values of  $u$ . To make this precise, we will study the *overall opinion* of experts over distribution  $d(\cdot)$ :

**Definition 3.** Expert  $e = (y^{(i)}, y^{(j)})$  will report their **overall opinion**, the expected value<sup>1</sup> of situational opinions:

$$\mathbb{E}_d[f_u(e)] = \sum_{u \in \mathcal{U}} d(u) f_u(e).$$

Expert graphs will be the analog of situational expert graphs with weights given by *overall opinions*:

**Definition 4.** An **expert graph**  $G_d^{\mathbf{P}} = (\mathcal{Y}, E, f_d(\cdot))$  encodes experts' pairwise overall opinions  $f_d(e) = \mathbb{E}_d(f_u(e))$  as weights on directed edges  $e = (y^{(i)}, y^{(j)}) \in E$ . An expert graph where  $(\mathcal{Y}, E)$  is a cycle will be referred to as an **expert cycle**.

We emphasize that input distribution  $d(\cdot)$  and probability table  $\mathbf{P}$  generate expert graph  $G_d^{\mathbf{P}}$ , but they may not be known to the observer of the graph due to the sometimes inaccessible reasoning of both human and machine classifiers.

**Observation 4.** As in Observation 3,  $f_d(y^{(i)}, y^{(j)}) = 1 - f_d(y^{(j)}, y^{(i)})$ .

Figure 2 shows an example of an expert graph with  $n = 4$  classes and  $m = 5$  experts and demonstrates the equivalence given in Observation 4.

<sup>1</sup>In this paper we focus on discrete distributions for which the expectation is a sum. An extension to non-discrete distributions is naturally given by using an integral for the expectation.

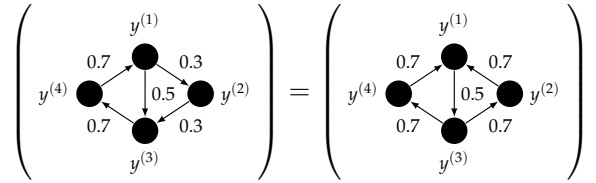


Fig. 2. Two equivalent expert graphs with  $\mathcal{Y} = \{y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}\}$  and  $f_d(y^{(1)}, y^{(2)}) = 0.3$ ,  $f_d(y^{(2)}, y^{(3)}) = 0.3$ ,  $f_d(y^{(3)}, y^{(4)}) = 0.7$ ,  $f_d(y^{(4)}, y^{(1)}) = 0.7$ ,  $f_d(y^{(1)}, y^{(3)}) = 0.5$ .

It will be useful to think of our expert graphs as decompositions of situational expert graphs due to the following observation:

**Observation 5.** Let  $\mathcal{G}_{\mathcal{U}}$  denote the set of all possible situational expert graphs and let  $\mathcal{G}$  denote the set of all possible expert graphs. We have that  $\mathcal{G} = \overline{\text{Co}}(\mathcal{G}_{\mathcal{U}})$ .

In this perspective, each situational expert graph can be thought of as representing a specific row of the probability table. An example of decomposing an expert into situational expert graphs is given in Figure 1.

### III. THE CURL CONDITION

#### A. Definition of curl

We now give a property on expert graphs called the *curl*, which is motivated by vector calculus and physics.

**Definition 5.** Given an expert graph  $G_d^{\mathbf{P}} = (\mathcal{Y}, E, f_d)$  and a cycle of indices  $\mathcal{C} = (c_1, c_2, \dots, c_\ell)$ , we define the **curl** to be:

$$\text{Curl}(G_d^{\mathbf{P}}, \mathcal{C}) = f_d(y^{(c_\ell)}, y^{(c_1)}) + \sum_{i=1}^{\ell-1} f_d(y^{(c_i)}, y^{(c_{i+1})}). \quad (3)$$

**Example 1.** In Figure 2,  $\text{Curl}(G, (1, 2, 3)) = .3 + .3 + .5 = 1.1$ .

It will be useful to also define a similar notion for situational expert graphs, called the *situational curl*.

**Definition 6.** Given a situational expert graph  $G_u^{\mathbf{P}} = (\mathcal{Y}, E, f_u)$  and a cycle of indices  $\mathcal{C} = (c_1, c_2, \dots, c_\ell)$ , we define the **situational curl** to be:

$$\text{Curl}(G_u^{\mathbf{P}}, \mathcal{C}) = f_u(y^{(c_\ell)}, y^{(c_1)}) + \sum_{i=1}^{\ell-1} f_u(y^{(c_i)}, y^{(c_{i+1})}). \quad (4)$$

**Observation 6.** If  $\bar{\mathcal{C}} = (c_\ell, c_{\ell-1}, \dots, c_2, c_1)$  is the reversed direction of the cycle  $\mathcal{C}$ , applying Observation 3 and Observation 4 to every edge gives:

$$\begin{aligned} \text{Curl}(G_u^{\mathbf{P}}, \bar{\mathcal{C}}) &= \ell - \text{Curl}(G_u, \mathcal{C}), \\ \text{Curl}(G_d^{\mathbf{P}}, \bar{\mathcal{C}}) &= \ell - \text{Curl}(G_d, \mathcal{C}). \end{aligned}$$

### B. The curl condition

We will now give upper and lower bounds on the curl. We will first give this condition for *situational expert graphs* which will then extend to regular expert graphs by linearity of expectation.

**Lemma 7.** Given a situational expert graph  $G_u^{\mathbf{P}}$  cycle  $\mathcal{C}$  of length  $\ell \geq 3$ , the situational curl must follow:

$$1 < \text{Curl}(G_u^{\mathbf{P}}, \mathcal{C}) < \ell - 1. \quad (5)$$

*Proof:* We begin by showing the lower bound. Recall that

$$\text{Curl}(G_u^{\mathbf{P}}, \mathcal{C}) = \frac{p_u^{c_\ell}}{p_u^{c_\ell} + p_u^{c_1}} + \sum_{i=1}^{\ell-1} \frac{p_u^{(c_i)}}{p_u^{(c_i)} + p_u^{(c_{i+1})}}. \quad (6)$$

The denominators of every term are strictly upper bounded by 1. This gives

$$\text{Curl}(G_u^{\mathbf{P}}, \mathcal{C}) > \sum_{i=1}^{\ell} p_u^{(c_i)} = 1. \quad (7)$$

Now, we can use this result to prove the upper bound. By Observation 6 we have

$$\text{Curl}(G_u^{\mathbf{P}}, \bar{\mathcal{C}}) = \ell - \text{Curl}(G_u^{\mathbf{P}}, \mathcal{C}) \stackrel{(a)}{<} \ell - 1,$$

where (a) is given by invoking the lower bound proved above. ■

**Corollary 8.** Given expert graph  $G_d^{\mathbf{P}}$  and a cycle  $\mathcal{C}$  on  $\ell \geq 3$  classes, then the curl follows<sup>2</sup>

$$1 < \text{Curl}(G_d^{\mathbf{P}}, \mathcal{C}) < \ell - 1. \quad (8)$$

*Proof:* This result follows from Lemma 7 and linearity of expectation on the definition of  $f_d(\cdot)$ . ■

**Definition 7.** We call a weighted digraph  $G = (\mathcal{Y}, E, f(\cdot))$  obeying  $f(y^{(i)}, y^{(j)}) = 1 - f(y^{(j)}, y^{(i)})$  **curl consistent** if all of its cycles obey the curl condition given by Lemma 7 and Corollary 8. The set of such digraphs is referred to as **curl consistent weighted digraphs (CCWDs)**.

It is helpful to think of the class of CCWDs as different from the set of expert graphs, which must be generated by some probability table. Corollary 8 gives us a *necessary* condition

<sup>2</sup>The notion of curl can be extended to one vs multiple class experts as well, in which case we can extend the upper bound of corollary 8 to  $\ell - q + 1$ , where  $q$  represents the total number of classes used by the expert.

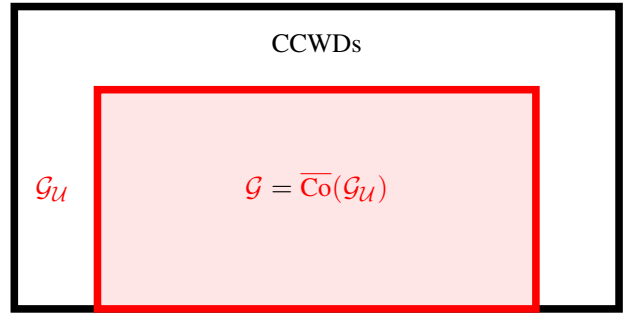


Fig. 3.  $G_u$  represents the set of possible situational expert graphs given by some probability vector, and  $\mathcal{G}$  is their closed convex hull, representing the possible overall expert graphs. The set of curl consistent weighted digraphs (CCWDs) is given in black, fully containing  $\mathcal{G}$  as shown by Lemma 7 and Corollary 8.

for expert graphs that is a result of them being generated by a probability table  $\mathbf{P}$  and input distribution  $d(\cdot)$ . This section has shown CCWDs must contain the set of expert graphs  $\mathcal{G}$ , as illustrated in Figure 3.

## IV. SYNTHETIC EXPERTS

### A. Bounding missing edges

The definition of expert graphs is given on digraphs that are not necessarily complete. If we are given an expert graph with missing edge-weights, many possible probability table and distribution pairs ( $\mathbf{P}$  and  $d(\cdot)$ ) may generate the *given* edge-weights. Each of these possible models gives different values on the *missing* edge-weights. Hence, in the absence of a relevant expert, we would like to give ranges which are consistent with *some* model given by  $\mathbf{P}$  and  $d(\cdot)$ , i.e. membership in the class of expert graphs  $\mathcal{G}$ . Unfortunately, this problem is very difficult, as we will see from the reduction of our system to the linear ordering polytope in Section V.

Alternatively, we know from the previous section that a violation of the curl condition signifies that no model can generate the weighted digraph as an expert graph. Hence, a more feasible step towards handling missing experts involves bounding edge-weights so that we maintain the graph's curl consistency.

We begin by giving the notion of path weights.

**Definition 8.** Let  $\mathcal{B} = (b_1, b_2, \dots, b_\ell)$  denote ordered indices of classes in expert graph  $G_d^{\mathbf{P}} = (\mathcal{Y}, E, f_d(\cdot))$  with  $(y^{(b_i)}, y^{(b_{i+1})}) \in E \forall i \in [\ell - 1]$ . The **weight** of a path is given by:

$$W(G_d^{\mathbf{P}}, \mathcal{B}) = \sum_{i=1}^{\ell-1} f_d(y^{(b_i)}, y^{(b_{i+1})}). \quad (9)$$

The path weight is very similar to the curl, but is defined on a path that does not loop back to its starting point.

**Lemma 9.** Consider CCWD  $G = (\mathcal{Y}, E, f(\cdot))$ , with edge  $(y^{(i)}, y^{(j)}) \notin E$ . Let  $\mathcal{B}^{(ij)}$  be the set of paths in  $G$  beginning at vertex  $y^{(i)}$  and ending at vertex  $y^{(j)}$ :

$$\mathcal{B}^{(ij)} = \{B : b_1 = i, b_{|B|} = j\}.$$

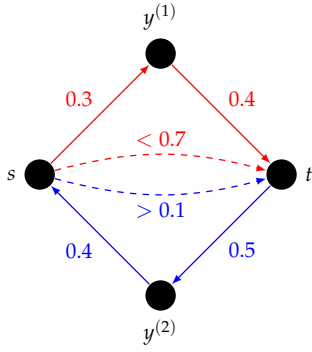


Fig. 4. The shortest path from  $s \rightarrow t$ , given by  $s \rightarrow y^{(1)} \rightarrow t$  gives  $f_d(s, t) < 0.7$ . The shortest path from  $t$  to  $s$ , given by  $t \rightarrow y^{(2)} \rightarrow s$  gives  $f_d(s, t) > 1 - .9 = .1$ . Together, we get  $f_d(s, t) \in (0.1, 0.7)$ .

Then,

$$1 - \min_{\bar{B} \in \mathcal{B}^{(ij)}} W(G, \bar{B}) < f(y^{(i)}, y^{(j)}) < \min_{B \in \mathcal{B}^{(ij)}} W(G, B). \quad (10)$$

*Proof:* Consider the  $\ell$ -length cycle  $B = (i, \dots, j)$  and its reverse  $\bar{B} = (j, \dots, i)$ . Corollary 8 gives

$$1 < \text{Curl}(G, \bar{B}) < \ell - 1. \quad (11)$$

Notice that

$$\text{Curl}(G, \bar{B}) = W(G, \bar{B}) + f(y^{(i)}, y^{(j)}).$$

Substituting this into Equation 11 gives

$$1 - W(G, \bar{B}) < f(y^{(i)}, y^{(j)}) < \ell - 1 - W(G, \bar{B}).$$

Now, by Observation 6, we can simplify the upper bound to get

$$1 - W(G, \bar{B}) < f(y^{(i)}, y^{(j)}) < W(G, B). \quad (12)$$

Recall that  $B$  and  $\bar{B}$  were chosen without loss of generality. Obtaining the optimal bounds over all paths yields the desired result. ■

A number of shortest-path algorithms for weighted directed graphs, such as Floyd-Warshall [5], can be applied to find the best bounds given by Lemma 9. An example of this process is illustrated in Figure 4.

We will later show that the curl condition is sufficient for describing all possible expert graphs whose edges form a single cycle, giving a case for which these bounds are tight for both CCWDs and expert graphs. There are, however, counter-examples of CCWDs that are not achievable by some probability table and input distribution. Thus, in general, these bounds can be improved.

### B. $\zeta$ -accurate synthetic experts

We can restrict the possible edge-weights of a synthetic expert to arbitrary precision.

**Definition 9.** For  $\zeta > 0$  call the synthetic expert between  $y^{(i)}$  and  $y^{(j)}$  a  $\zeta$ -accurate synthetic expert if for some lower bound  $L$ ,  $f(y^{(i)}, y^{(j)}) \in (L, L + \zeta)$ .

**Lemma 10.** Given CCWD  $G = (\mathcal{Y}, E, f(\cdot))$ , the synthetic expert between  $y^{(i)}$  and  $y^{(j)}$  is  $\zeta$ -accurate if and only if there

exists a cycle of indices  $\mathcal{C} = (c_1, \dots, c_\ell)$  with  $i, j \in \mathcal{C}$  and  $\text{Curl}(G, \mathcal{C}) \leq 1 + \zeta$ .

*Proof:* Choose  $y^{(j)}$  and  $y^{(i)}$  wlog and denote the shortest path:

$$S^{(ij)} = \arg \min_{B \in \mathcal{B}^{(ij)}} W(G, B). \quad (13)$$

Recall from Lemma 9 that

$$1 - W(G, S^{(ji)}) < f(y^{(i)}, y^{(j)}) < W(G, S^{(ij)}).$$

Let  $L = 1 - W(G, S^{(ji)})$ . The desired gap between bounds is  $W(G, S^{(ij)}) - (1 - W(G, S^{(ji)})) \leq \zeta$ , which is true if and only if

$$W(G, S^{(ji)}) + W(G, S^{(ij)}) \leq 1 + \zeta.$$

Now consider the cycle  $\mathcal{C}^{(iji)} = S^{(ij)}S^{(ji)}$ , which is the shortest cycle through  $y^{(i)}$  and  $y^{(j)}$ . We have

$$\text{Curl}(G, \mathcal{C}^{(iji)}) = W(G, S^{(ji)}) + W(G, S^{(ij)}) \leq 1 + \zeta. \quad \blacksquare$$

$\zeta$ -accurate synthetic experts describe convergence to fully determining a missing opinion without access to the relevant expert. These synthetic experts can form subgraphs of any CCWD.

**Lemma 11.** Given a CCWD  $G = (\mathcal{Y}, E, f(\cdot))$ , there exists another weighted digraph  $G' = (\mathcal{Y} \cup \mathcal{Y}', E', f'(\cdot))$  with  $\mathcal{Y}' \cap \mathcal{Y} = \emptyset$  and  $E \cap E' = \emptyset$  such that

- 1) All synthetic experts  $e \in E$  in  $G'$  are  $\zeta$ -accurate. That is, we have  $f(e) \in (L^{(e)}, L^{(e)} + \zeta)$  for some lower bound  $L^{(e)}$ .
- 2)  $G'$  is also curl consistent.

*Proof:* We give a constructive proof, for which an example is given in Figure 5. Begin with the set of classes  $\mathcal{Y}$  and  $E' = \emptyset$ . Now, for each  $e = (y^{(i)}, y^{(j)}) \in E$ , do the following:

- 1) Add a vertex  $y^{(ij)}$  to  $\mathcal{Y}'$  and assign values to  $f'(y^{(i)}, y^{(ij)})$  and  $f'(y^{(ij)}, y^{(j)})$  to create a path from  $y^{(i)} \rightarrow y^{(ij)} \rightarrow y^{(j)}$  of weight  $f(e) + \frac{\zeta}{2}$ .
- 2) Similarly add a vertex  $y^{(ji)}$  to create a path from  $y^{(j)} \rightarrow y^{(ji)} \rightarrow y^{(i)}$  of weight  $1 - f(e) + \frac{\zeta}{2}$ .

By Lemma 10,  $e$  is now a  $\zeta$ -accurate synthetic classifier in  $G'$  that includes value  $f(e)$ .

If this process created a cycle  $\mathcal{C}'$  in  $G'$  with  $\text{Curl}(G', \mathcal{C}') \leq 1$ , then there must also be a cycle through  $G$  with  $\text{Curl}(G, \mathcal{C}) \leq 1$ , which is a contradiction. Thus, we know  $G'$  as constructed is curl consistent. ■

### C. Feasibility of networks of synthetic experts

The bounds provided in Lemma 9 interact with each other when many experts are being synthesized on the same graph. For example, the bounds obtained for each edge may not be attainable simultaneously while still guaranteeing membership in the class of CCWDs (see Figure 6).

While the bounds for synthetic experts may not be simultaneously attainable, each value within the the bounds does correspond to a valid CCWD.

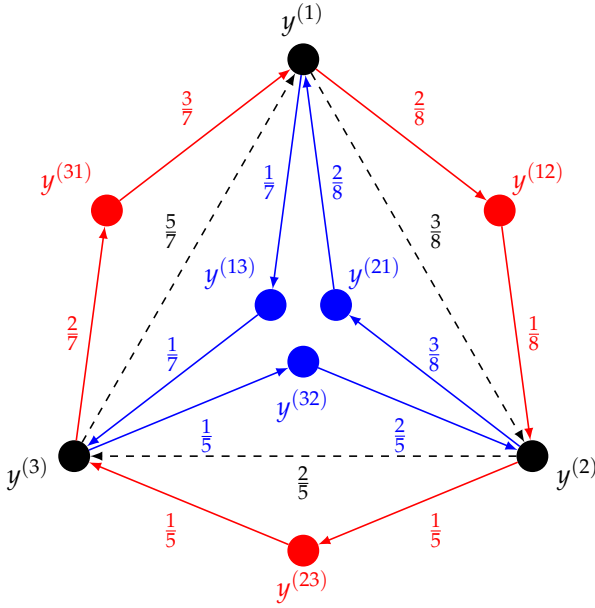


Fig. 5. An example of how to create  $\zeta$ -accurate synthetic experts that make up any curl consistent graph by adding additional paths. Assume all edge weights given in color have a  $+\frac{1}{4}$ . Here, the cycle on  $y^{(1)}, y^{(2)}, y^{(3)}$  is created by: (1) Adding shortest paths (shown in red on the other part of the cycle) with total weight equal to the desired edge weight  $f(e) + \frac{1}{2}$ . (2) Adding shortest reverse paths (shown in blue on the inner part of the cycle) with total weight equal to  $1 - f(e) + \frac{1}{2}$ .

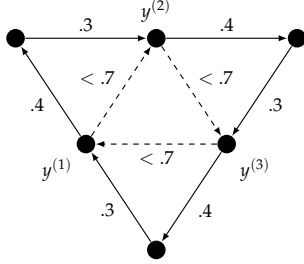


Fig. 6. An example of how we cannot always achieve any combination of synthetic expert bounds. Here, a choice of just under .7 for all the synthetic experts (given in dashed lines) would violate the curl condition on cycle  $y^{(1)} \rightarrow y^{(2)} \rightarrow y^{(3)} \rightarrow y^{(1)}$ .

**Lemma 12.** Let  $S^{(ij)}$  denote the shortest path from  $y^{(i)}$  to  $y^{(j)}$  as given in Equation 13. Consider a CCWD  $G = (\mathcal{Y}, E, f(\cdot))$  and synthetic experts bounds on  $E'$  given by Lemma 9 and denoted by  $f(y^{(i)}, y^{(j)}) \in (1 - S^{(ji)}, S^{(ij)})$ . For each  $e' = (y^{(i)}, y^{(j)}) \in E'$  and assignment  $v \in (1 - S^{(ji)}, S^{(ij)})$ , there exists at least one  $G^* = (\mathcal{Y}, E^*, f^*(\cdot))$  with  $E^* = E \cup E'$  and  $f^*(e') = v$  that is curl consistent.

*Proof:* We will prove this by inducting on the cardinality of  $E'$ . First consider the base case of  $|E'| = 0$ , which is trivially solved by setting  $G^* = G$ .

Now, assume for induction that the statement holds for all  $E'$  with  $|E'| \leq q - 1$ . Choose  $e' = (y^{(i)}, y^{(j)}) \in E'$  and  $v \in (1 - S^{(ji)}, S^{(ij)})$  wlog and use this to define  $\tilde{G} = (\mathcal{Y}, \tilde{E}, \tilde{f}(\cdot))$  where  $\tilde{E} = E \cup \{e'\}$ ,  $\tilde{f}(e) = f(e) \forall e \in E$  and  $\tilde{f}(e') = v$ .

First, we show that  $\tilde{G}$  is curl consistent. All cycles  $\mathcal{C}$  which do not include  $e'$  have  $\text{Curl}(G, \mathcal{C}) = \text{Curl}(\tilde{G}, \mathcal{C})$ . Now

consider wlog some cycle  $\mathcal{C}^{(ij)} = (c_1, \dots, i, j, \dots, c_{\ell-2}) = (i, j)B^{(ji)}$  containing edge  $e'$ .

Recall from Lemma 9 that

$$1 - W(G, S^{(ji)}) = 1 - \min_{\bar{B} \in \mathcal{B}^{(ji)}} W(G, \bar{B}) < \tilde{f}(y^{(i)}, y^{(j)}). \quad (14)$$

Now, we can decompose the curl of  $\mathcal{C}^{(ij)}$  and apply Equation 14 to get

$$\begin{aligned} \text{Curl}(\tilde{G}, \mathcal{C}^{(ij)}) &= \tilde{f}(y^{(i)}, y^{(j)}) + W(\tilde{G}, B^{(ji)}) \\ &> 1 - \min_{\bar{B} \in \mathcal{B}^{(ji)}} W(G, \bar{B}) + W(\tilde{G}, B^{(ji)}). \end{aligned}$$

Because  $B^{(ji)} \in \mathcal{B}^{(ji)}$ , we have

$$\text{Curl}(\tilde{G}, \mathcal{C}^{(ij)}) > 1.$$

$\mathcal{C}^{(ij)}$  was chosen wlog, so we can conclude that all cycles have curl  $> 1$ . As before, all these cycles can be reversed which means they also have curl upper bounded by  $\ell - 1$ .

So, every assignment to an edge in  $E'$  yields a  $\tilde{G}$  that is still curl consistent with one more edge. We can now let  $E'' = E' \setminus \{e'\}$  and use our inductive assumption on  $\tilde{G}$  with synthetic expert set  $E''$ . ■

Lemma 12 implies that, when referring to the set of CCWDs, the bounds given for synthetic experts are optimal. We will later see that not all curl consistent weighted digraphs can be expert graphs. Recall that Lemma 11 showed that our bounds could restrict us to being arbitrarily close to any CCWD in a network of synthetic experts. As a result, we will see that not all synthetic expert graphs are achievable as an expert graph.

## V. EQUIVALENCE TO THE LINEAR ORDERING POLYTOPE

Proving that the curl condition is sufficient for the set of possible expert graphs  $\mathcal{G}$  involves finding some probability table  $\mathbf{P}$ , and input distribution  $d(\cdot)$ , which generate  $G_d^{\mathbf{P}}$  for any possible assignment of edge-weights that satisfies the curl condition. In this section we will show that this task is not always possible.

To make the decomposition process easier, we will consider a related set: the “linear ordering polytope” [7]. In this setting, situational edges  $f_r(y^{(i)}, y^{(j)})$  are binary preferences instead of soft decisions given by conditional probabilities. We will call the analog to a situational expert graph will be called a “ranking graph.”

**Definition 10.** A ranking graph  $H_r = (\mathcal{Y}, E, f_r(\cdot))$  is a complete graph ( $E = \mathcal{Y} \times \mathcal{Y}$ ) with a binary edge function  $f_r(\cdot) : E \mapsto \{0, 1\}$  given by an implicit ordering of indices  $r = (r_1, \dots, r_n)$ :

$$f_r(y^{(r_i)}, y^{(r_j)}) = \mathbb{1}[i < j].$$

An example of a ranking graph on 4 classes is given in Figure 7.

To distinguish between our frameworks, the analogue of  $d(u)$  will be  $w(r)$ , giving a “ranking distribution.” This will yield a component of the linear ordering polytope, which we will call a “linear ordering graph.”

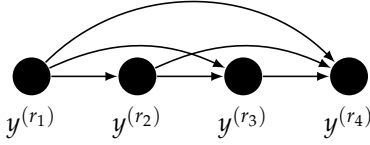


Fig. 7. An example of a ranking graph  $H_r(\mathcal{C}, E, f_r)$  with all edge weights 1. The classes have been arranged by rank from left to right, so that the direction with weight 1 always points right.

**Definition 11.** A **linear ordering graph (LOG)**  $H_w = (\mathcal{Y}, E, f_w(\cdot))$  encodes induced binary probabilities [7]

$$f_w(y^{(i)}, y^{(j)}) = \mathbb{E}_w[f_r(y^{(i)}, y^{(j)})] = \sum_{r \in \mathcal{R}} w(r) f_r(y^{(i)}, y^{(j)}), \quad (15)$$

where  $\mathcal{R}$  represents all  $n!$  possible orderings of indices.

**Observation 13.** Let  $\mathcal{H}_{\mathcal{R}}$  denote the set of all possible ranking graphs and  $\mathcal{H}$  denote the set of all possible LOGs (the linear ordering polytope). Similar to Observation 5, we have  $\mathcal{H} = \overline{\text{Co}(\mathcal{H}_{\mathcal{R}})}$ .

An example of a LOG as a composition of ranking graphs is given in Figure 8.

Ranking graphs and LOGs have a condition very close to the curl condition given in Lemma 7 and Corollary 8, with the only difference being a nonstrict inequality.

**Lemma 14.** Given a ranking graph  $H_r = (\mathcal{Y}, E, f_r(\cdot))$  and a cycle  $\mathcal{C}$  on  $\ell \geq 3$  classes, then

$$1 \leq \text{Curl}(H_r, \mathcal{C}) \leq \ell - 1. \quad (16)$$

*Proof:* Let  $\rho(c_i)$  give the ranking of  $c_i$  such that  $r_{\rho(c_i)} = c_i$ . First, note that with  $f_r(e) \in \{0, 1\}$  for all  $e \in E$ ,  $\text{Curl}(H_r, \mathcal{C})$  must be an integer. Now, assume for contradiction that  $\text{Curl}(H_r, \mathcal{C}) = 0$ , hence  $f_r(y^{(c_i)}, y^{(c_{i+1})}) = 0 \forall i \in [\ell - 1]$ . Recall that  $f_r(y^{(r_i)}, y^{(r_j)}) = \mathbb{1}[i < j]$ . This implies

$$r_{\rho(c_1)} > r_{\rho(c_2)} > \dots > r_{\rho(c_\ell)}.$$

A contradiction of transitivity now arises, because we also have  $f_r(y^{(c_\ell)}, y^{(c_1)}) = 0$ , so  $r_{\rho(c_1)} < r_{\rho(c_\ell)}$ .

So we have  $\text{Curl}(H_r, \mathcal{C}) > 0$ . This gives  $\text{Curl}(H_r, \mathcal{C}) \geq 1$  because we must have integer curl. As with expert graphs, the upper bound follows from considering the curl on the reverse cycle  $\bar{\mathcal{C}}$ , which must also follow this lower bound. ■

Linearity gives the corollary for the convex hull.

**Corollary 15.** Given LOG  $H_w = (\mathcal{Y}, E, f_w(\cdot))$  and a cycle  $\mathcal{C}$  on  $\ell \geq 3$  classes, then

$$1 \leq \text{Curl}(H_w, \mathcal{C}) \leq \ell - 1. \quad (17)$$

Somewhat surprisingly, these two frameworks are essentially<sup>3</sup> equivalent: any expert graph can also be achieved as a LOG, and any LOG in the interior of the linear ordering polytope can be achieved as an expert graph.

<sup>3</sup>LOGs may have  $\text{Curl}(H_w, \mathcal{C}) = 1$  or  $\text{Curl}(H_w, \mathcal{C}) = \ell - 1$  for cycle  $\mathcal{C}$  length  $\ell$ , which violates the strict inequalities in Corollary 8. To handle this detail, we consider the interior of the linear ordering polytope.

**Theorem 16.** The interior of the linear ordering polytope  $\text{Co}(\mathcal{H}_{\mathcal{R}})$  and the set of all possible expert graphs  $\mathcal{G}$  are equivalent. That is,  $G_d^{\mathbf{P}} = (\mathcal{Y}, E, f_d(\cdot)) \in \mathcal{G}$  if and only if there exists  $H_w = (\mathcal{Y}, E, f_w(\cdot)) \in \text{Co}(\mathcal{H}_{\mathcal{R}})$  with  $f_w(e) = f_d(e) \forall e \in E$ .

The next two subsections will focus on proving this equivalence. In Subsection V-A, we will give a reduction from any LOG to a probability table that generates an expert graph with the same pairwise edge weights. In Subsection V-B, we will give a reduction from any expert graph to a set of rankings  $r \in \mathcal{R}$  and weights  $w(r)$  that generate a LOG that matches the expert graph's pairwise weights. This equivalence will allow us to harness results from the linear ordering polytope framework to understand the power and limitations of the curl condition.

#### A. Reduction from LOGs to expert graphs

To show that any LOG can also be an expert graph, we will give a mapping from  $w(\cdot) : \mathcal{R} \mapsto [0, 1]$  to a set of states  $U$  with categorical distributions  $\mathbf{p}_u$  and probabilities  $d(\cdot) : \mathcal{U} \mapsto (0, 1)$  such that  $f_d(\cdot) = \mathbb{E}_d[f_u(\cdot)]$  matches  $f_w(\cdot) = \mathbb{E}_w[f_r(\cdot)]$ .

We will first show how to give a set of class probabilities  $\mathbf{p}_u$  for which the pairwise expert output gets arbitrarily close to that of a ranking  $r \in \mathcal{R}$ . Here it will be convenient to think of the edge-weights as a vector  $\mathbf{f}_u = (f_u(e_1), \dots, f_u(e_m))$  and  $\mathbf{f}_d = (f_d(e_1), \dots, f_d(e_m))$ .

**Lemma 17.** Let  $H_r = (\mathcal{Y}, E, f_r(\cdot))$  be a ranking graph with  $r = (r_1, \dots, r_n)$ . For all  $\varepsilon > 0$ , we can construct a situational expert graph  $G_u^{\mathbf{P}} = (\mathcal{Y}, E, f_u(\cdot))$  with a categorical distribution given by  $p_u^{(1)}, \dots, p_u^{(n)} \in (0, 1)$  such that

$$\|\mathbf{f}_u - \mathbf{f}_r\|_2 = \sum_{e \in E} (f_r(e) - f_u(e))^2 < \varepsilon. \quad (18)$$

*Proof:* Let  $t \in \mathbb{R}$ . Let  $\alpha_i = \frac{1}{t^i}$  and  $z = \sum_{i=1}^n \alpha_i$ . If we set  $p_u^{(r_i)} = \frac{\alpha_i}{z}$ , then

$$f_u^{(t)}(y^{(r_i)}, y^{(r_j)}) : \begin{cases} \leq \frac{1}{t} & \text{if } i > j \\ = \frac{1}{2} & \text{if } i = j \\ > 1 - \frac{1}{t} & \text{if } i < j \end{cases}$$

Recall that  $f_r(y^{(r_i)}, y^{(r_j)}) = \mathbb{1}[i < j]$ . Thus, we have assigned successively larger probabilities to higher ranked classes such that the resulting experts have

$$|f_u^{(t)}(y^{(r_i)}, y^{(r_j)}) - f_r(y^{(r_i)}, y^{(r_j)})| \leq \frac{1}{t}. \quad (19)$$

There are  $m$  edges, so set  $\frac{1}{t} = \frac{\varepsilon}{m}$  to get  $\|\mathbf{f}_u^{(t)} - \mathbf{f}_r\|_2 \leq \varepsilon$  as desired. ■

We can use Lemma 17 to construct a state  $u_r$  with probability vector  $\mathbf{p}_u$  which achieves a  $f_u(\cdot)$  that is *very close* to  $f_r(\cdot)$  for each  $r \in \mathcal{R}$ .

The remainder of this reduction is given by showing the interior of the convex hull of our generated  $\mathbf{f}_u$  vectors is the same as the interior of the convex hull of  $\mathbf{f}_r$  vectors. To do this, we make use of the following more general result:

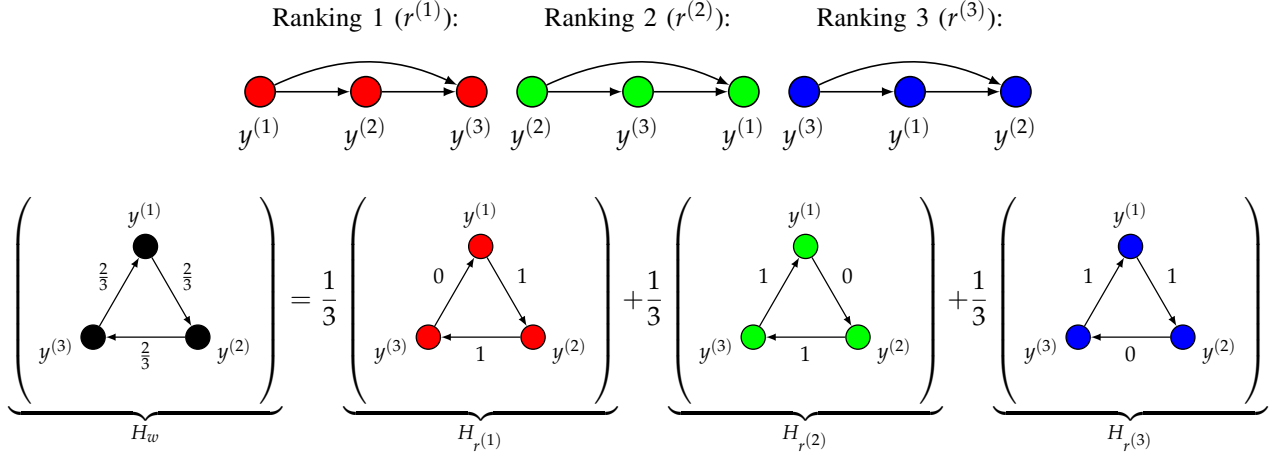


Fig. 8. The LOG  $H_w$  given by ranking graphs from rankings  $r^{(1)}$ ,  $r^{(2)}$  and  $r^{(3)}$  and uniform  $w(r) = \frac{1}{3}$ . Edge-weights in  $H_w$  are averages over the edge-weights in  $H_{r^{(1)}}$ ,  $H_{r^{(2)}}$ ,  $H_{r^{(3)}}$  determined by the rankings. Decompositions of this form can be used to prove that  $H_w \in \text{Co}(\mathcal{H}_{\mathcal{R}}) = \mathcal{H}$ .

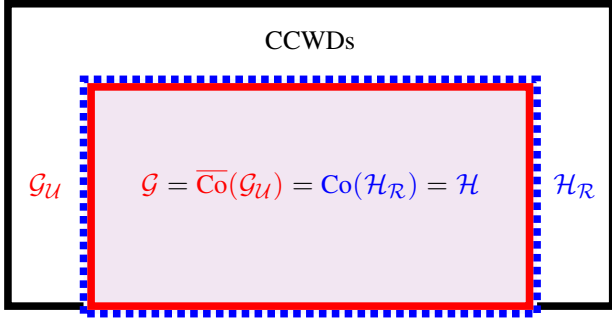


Fig. 9. An abstract diagram showing the membership of various sets defined in this section, building on Figure 3. The set of rankings  $\mathcal{H}_{\mathcal{R}}$  is shown as blue dots because they are discrete structures. These rankings can have cycles with curl = 1 or =  $\ell - 1$ , so they are shown to escape the curl condition set on the boundary. Theorem 16 shows that  $\mathcal{G} = \text{Co}(\mathcal{H}_{\mathcal{R}})$ . Finally, the gap between the CCWDs and  $\text{Co}(\mathcal{H}_{\mathcal{R}})$  comes from literature on the linear ordering polytope [7].

**Theorem 18.** Consider a set  $V = \{\mathbf{v}_1, \dots, \mathbf{v}_t\}$  and with  $\mathbf{v}_i \in \mathbb{R}^m$  for all  $i$ . If we have  $\tilde{V}$  such that for every  $\varepsilon$  and  $\mathbf{v} \in V$ , there exists  $\tilde{\mathbf{v}} \in \tilde{V}$  such that  $\|\tilde{\mathbf{v}} - \mathbf{v}\|_2 < \varepsilon$ , then we have

$$\text{Co}(V) \subseteq \overline{\text{Co}}(\tilde{V}). \quad (20)$$

*Proof:* The proof is the result of Lemma 24, which we prove in Appendix Section A. ■

This machinery allows us to prove the goal lemma of this subsection:

**Lemma 19.**  $\text{Co}(\mathcal{H}_{\mathcal{R}}) = \mathcal{G}$ .

*Proof:* Consider Theorem 18. Let  $V = \{\mathbf{f}_{\mathbf{r}} : H_{\mathbf{r}} \in \mathcal{H}_{\mathcal{R}}\}$  and  $\tilde{V} = \{\mathbf{f}_{\mathbf{u}} : (\mathcal{Y}, E, f_d(\cdot)) \in \mathcal{G}_U\}$ . Lemma 17 shows that for every  $\varepsilon$  and  $H_{\mathbf{r}} \in \mathcal{H}_{\mathcal{R}}$  there is some  $G_{\mathbf{u}}^{\mathbf{P}} \in \mathcal{G}_U$  for which the edge-weight vectors are within  $\varepsilon$ :  $\|\mathbf{f}_{\mathbf{u}} - \mathbf{f}_{\mathbf{r}}\|_2 < \varepsilon$ . This satisfies the requirements for us to apply Theorem 18 to say  $\text{Co}(\mathcal{H}_{\mathcal{R}}) = \overline{\text{Co}}(\mathcal{G}_U) = \mathcal{G}$ . ■

## B. Reduction from expert graphs to LOGs

We will now show that, given an expert graph  $G_d^{\mathbf{P}} = (\mathcal{C}, E, f_d(\cdot))$ , we can find a decomposition into ranking graphs. Following the usual strategy in this paper, we will begin by showing we can decompose a situational expert graph  $G_u^{\mathbf{P}}$  into LOGs, which will imply the overall result. Recall that any true expert graph *must* be generated by some probability table. Lemma 20 will sort the population into sub-populations for each *first* choice and determine their sizes from values in the probability table. A recursion on each sub-population will then determine the distribution of second choices, and so on. A sketch of this procedure is given in Figure 10.

In order to describe the expert graph on the sub-population with a set of choices, we will define “prefix expert graphs.”

**Definition 12.** Given an expert graph  $G_d^{\mathbf{P}} = (\mathcal{Y}, E, f_d(\cdot))$  and prefix  $a_1$  and non-specified leftover indices  $\{b_1, \dots, b_{n-1}\}$ , we define the **prefix expert graph**  $G_d^{\mathbf{P}}[a_1] = (\mathcal{Y}, E, f_{G_d^{\mathbf{P}}[a_1]}(\cdot))$  to be a graph with edge weights given by:

$$\begin{aligned} f_{G_d^{\mathbf{P}}[a_1]}(y^{(a_1)}, y^{(b_i)}) &= 1, \\ f_{G_d^{\mathbf{P}}[a_1]}(y^{(b_i)}, y^{(b_j)}) &= f_d(y^{(b_i)}, y^{(b_j)}). \end{aligned}$$

The prefix graph fixes a primary preference on  $C^{(a_1)}$  while maintaining relationships between  $C^{(b_1)}, \dots, C^{(b_{n-1})}$ .

**Lemma 20.** We can decompose any situational expert graph  $G_u^{\mathbf{P}} = (\mathcal{Y}, E, f_u(\cdot))$  with categorical distribution  $\mathbf{p}_u = (p_u^{(1)}, p_u^{(2)}, \dots, p_u^{(n)}) \in \Delta^n$  into prefix expert graphs with a distribution given by

$$G_u^{\mathbf{P}} = \sum_{a_1=1}^n p_u^{(a_1)} G_u^{\mathbf{P}}[a_1].$$

That is, the edge weights

$$f^*(e) = \sum_{a_1=1}^n p_u^{(a_1)} f_{G_u^{\mathbf{P}}[a_1]}(e).$$



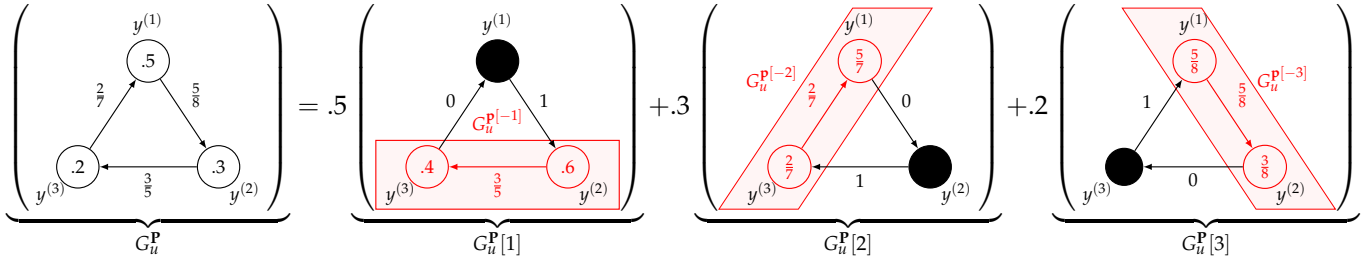


Fig. 10. A decomposition of an expert graph  $G_u^P$  into prefix expert graphs  $G_u^P[1], G_u^P[2], G_u^P[3]$  following Lemma 20. Probabilities from probability table  $\mathbf{p}$  are shown inside the nodes in  $G_u^P$  and probabilities for  $\mathbf{p}^{[-i]}$  are shown in red inside nodes in  $G_u^P[-i]$ . To get a decomposition into rankings  $H_r$  we apply the same procedure to the highlighted sub-graphs  $G_u^P[-1], G_u^P[-2], G_u^P[-3]$ . This determines the second choice distribution for each of these sub-populations, which will imply a third choice as well, giving a decomposition into rankings.

*Proof:* For this decomposition, the desired edge-weights  $f^*(e)$  for  $e = (y^{(i)}, y^{(j)})$  are achieved:

$$\begin{aligned}
f^*(e) &= p_u^{(i)} \underbrace{f_{G_u^P[i]}(e)}_{=1} + p_u^{(j)} \underbrace{f_{G_u^P[j]}(e)}_{=0} + \sum_{a_1 \neq i, j} p_u^{(a_1)} \underbrace{f_{G_u^P[a_1]}(e)}_{=f_u(e)} \\
&= p_u^{(i)} + \sum_{a_1 \neq i, j} p_u^{(a_1)} f_u(e) \\
&= (p_u^{(i)} + p_u^{(j)}) \frac{p_u^{(i)}}{p_u^{(i)} + p_u^{(j)}} + \sum_{a_1 \neq i, j} p_u^{(a_1)} f_u(e) \\
&= \sum_{a_1=1}^n p_u^{(b_1)} f_u(e) = f_u(e)
\end{aligned}$$

After Lemma 20 decides on the sub-population distribution for first choice candidates, we will recurse on each sub-population to determine the distribution on their second choice candidates. ■

**Definition 13.** We define  $G_u^P[-i]$  to be the the subgraph of  $G_u^P$  with  $y^{(i)}$  removed:

$$G_u^P[-i] = (\mathcal{Y} \setminus \{y^{(i)}\}, E \setminus \{e : y^{(i)} \in e\}, f_u^{(-i)}(\cdot)).$$

**Observation 21.** The subgraph  $G_u^P[-i]$  is generated by a new probability table with

$$p[-i]_u^{(j)} = \frac{p_u^{(j)}}{1 - p_u^{(i)}}. \quad (21)$$

**Lemma 22.** Given a situational expert graph  $G_u^P = (\mathcal{Y}, E, f_u(\cdot))$  with categorical distribution  $\mathbf{p}_u = (p_u^{(1)}, p_u^{(2)}, \dots, p_u^{(n)}) \in \Delta^n$ ,

$$w(r) = \prod_{i=1}^n \frac{p_u^{(r_i)}}{\prod_{j=1}^{i-1} (1 - p_u^{(r_j)})} = \frac{\prod_{i=1}^n p_u^{(r_i)}}{\prod_{i=0}^{n-1} (1 - p_u^{(r_i)})^{n-i}} \quad (22)$$

generates  $H_w = (\mathcal{Y}, E, f_w(\cdot))$  with  $f_w(e) = f_u(e) \forall e \in E$ .

*Proof:* Lemma 20 gives the decomposition

$$G_u^P = \sum_{i=1}^n p_u^{(i)} G_u^P[i].$$

Now, since the behavior of all edges with  $y^{(i)} \in e$  is the same for all components in  $G_u^P[i]$ , we can recurse on the sub-graph with  $y^{(i)}$  removed, with probabilities defined by Observation 21. The base case of  $\mathcal{Y} = \{y^{(j)}\}$  is trivial, with  $p_u^{(j)} = 1$ . Multiplication of the probabilities adjusted by Observation 21 gives Equation 22. ■

Finally, we can give the desired result for this section.

**Lemma 23.**  $\mathcal{G} \subseteq \text{Co}(\mathcal{R})$ .

*Proof:* Pick  $G_d^P \in \mathcal{G}$  without loss of generality. Decompose each state  $u$  using probability vector  $\mathbf{p}_u$  according to Lemma 22. Recall from Assumption 1 that all class probabilities  $p_u^{(i)} > 0$ , which yields  $w(r) > 0 \forall r \in \mathcal{R}$ . Hence,  $G_d^P \subseteq \text{Co}(\mathcal{R})$ . ■

### C. Implications for sufficiency of the curl condition

Theorem 18 allows us to harness results from the voting theory to show our condition is sufficient to describe all expert graphs with  $n \leq 5$  classes [7]. The curl condition is insufficient for the linear ordering polytope with  $n \geq 6$ .

Theorem 18 can also be used as a proof technique for sufficiency on more specific classes of expert graphs. The task now reduces to finding a decomposition of the weighted directed graph into a convex combination of rankings, for which we can find class probabilities via Lemma 17. Figure 8 gives an example of such a decomposition of a cycle. Importantly, these decompositions cannot depend on a probability table  $\mathbf{P}$  or distribution  $d(\cdot)$  as we did in Subsection V-B, since the goal here is to show that such a  $\mathbf{P}$  and distribution  $d(\cdot)$  exist.

We can show sufficiency of the curl condition for expert cycles via a decomposition algorithm of any curl-consistent expert cycle into ranking graphs that generate the edge-weights. The existence of such a decomposition is shown in Appendix Section B. A constructive decomposition is also given Appendix Section C, which is demonstrated in Figure 11.

## VI. CONCLUSION

We have defined expert graphs as a framework to understand the amalgamation of pairwise experts with overlapping training. To analyze these graphs, we introduced the curl and derived *necessary* lower and upper bounds. From this notion of curl we provided an algorithm to derive upper and lower bounds on missing edges of an expert graph. We have also shown our expert graph framework's characterization to be equivalent to that of the linear ordering polytope, which allows us to apply results from a rich literature [2], [7], [11].

Future work may expand the notion of the expert graph to an expert *hypergraph*, which includes experts with potentially *nonbinary* domains of expertise.

One of the most relevant applications of this framework is in machine learning, where a practitioner may use a number of classifiers trained on different overlapping datasets. After being given additional information, the user may wish to create a "synthetic classifier" for a new prior on the distribution of labels. Hence, our results point towards an exciting new way of handling label shift after training.

## REFERENCES

- [1] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: a unifying approach for margin classifiers," J. Mach. Learn. Res. 1, pp. 113–141, September 2001.
- [2] N. Alon, "Voting paradoxes and digraphs realizations," Advances in Applied Mathematics 29, no. 1, pp. 126–35. 2002.
- [3] D. Bertsimas and J. Tsitsiklis, "Introduction to Linear Optimization," (1st. ed.). Athena Scientific, 1997.
- [4] L. Breiman, "Bagging predictors," Machine learning 24, no. 2, pp. 123–140. 1996.
- [5] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, "Introduction to Algorithms," Third Edition (3rd. ed.). The MIT Press, 2009.
- [6] T. G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," J. Artif. Int. Res. 2 (1), pp. 263–286, August 1994.
- [7] P. C. Fishburn, "Induced binary probabilities and the linear ordering polytope: a status report," Mathematical Social Sciences 23, pp.67–80, 1992.
- [8] E. Frazzoli, and M. Dahleh, "6.241J Dynamic Systems and Control," Massachusetts Institute of Technology: MIT (Lecture 5), Spring 2011.
- [9] B. Grünbaum. *Convex polytopes*. Vol. 221. Springer Science & Business Media, 2013.
- [10] A. Indrayan, *Medical Biostatistics*, Second Edition, Chapman & Hall/CRC Press. 2008.
- [11] D. C. McGarvey, "A theorem on the construction of voting paradoxes," *Econometrica* 21, pp 608–610. 1953.
- [12] J. Moon, "Topics on tournaments in Graph Theory," Courier Dover Publications, 2015.
- [13] R.E. Schapire. "The strength of weak learnability," Machine learning 5, no. 2, pp. 197–227. 1990.
- [14] J. S. Weber, "An Elementary Proof of the Conditions for a Generalized Condorcet Paradox," *Public Choice* 77, no. 2, pp. 415–419. 1993.
- [15] Z. H. Zhou, "Ensemble methods: foundations and algorithms," CRC press. 2012.

## A. Proof of Theorem 18

Convex hulls of finite sets in  $\mathbb{R}^\ell$  are *convex* polytopes, which can be expressed as an intersection of  $h$  halfspaces indexed by  $f$  with  $\{\mathbf{x} : \mathbf{a}^{(f)\top} \mathbf{x} < b^{(f)}\}$  [9]. The perpendicular vectors  $\mathbf{a}^{(f)\top}$  can be combined as row-vectors of the matrix  $A$  so that any convex polytope can be expressed as

$$\{\mathbf{x} : A\mathbf{x} < \mathbf{b}\} = \left\{ \mathbf{x} : \begin{pmatrix} (\mathbf{a}^{(1)})^\top \\ \vdots \\ (\mathbf{a}^{(h)})^\top \end{pmatrix} \mathbf{x} < \begin{pmatrix} b^{(1)} \\ \vdots \\ b^{(h)} \end{pmatrix} \right\} \quad (23)$$

For convenience, the vectors  $\mathbf{a}^{(f)}, \tilde{\mathbf{a}}^{(f)}$  are assumed to be unit vectors throughout.

**Lemma 24.** *Given  $V = \{v_1, \dots, v_m\}$  and “perturbed points”  $\tilde{V} = \{\tilde{v}_1, \dots, \tilde{v}_m\}$  with  $v_i \in \mathbb{R}^\ell$  and  $\tilde{v}_j \in \mathbb{R}^\ell$  for all  $i, j$ . We have that if  $\mathbf{x} \in \text{Co}(V)$  and is  $\varepsilon > 0$  from the boundary  $\text{Bo}(\text{Co}(V))$ , then if we can find perturbed points  $\tilde{V}$  such that they are within  $\varepsilon$  from the desired  $V$ , then  $\mathbf{x} \in \overline{\text{Co}}(\tilde{V})$ .*

More precisely, let

$$\begin{aligned} \text{Co}(V) &= \{\mathbf{x} : A\mathbf{x} < \mathbf{b}\} \\ \text{Co}(\tilde{V}) &= \{\mathbf{x} : \tilde{A}\mathbf{x} < \tilde{\mathbf{b}}\} \end{aligned}$$

as given by Equation 23. If  $A\mathbf{x} < \mathbf{b} - \varepsilon \mathbf{1}_\ell$  and  $\|\mathbf{v}_i - \tilde{\mathbf{v}}_i\|_2 < \varepsilon \forall i$ , then  $\tilde{A}\mathbf{x} < \tilde{\mathbf{b}}$ .

To prove this theorem, we will need to show that the boundaries of the polytopes do not move too far. We will do this using Lemma 25, which bounds how far  $\text{Bo}(\text{Co}(V))$  can be from  $\text{Bo}(\text{Co}(\tilde{V}))$  along a single “face.”

**Definition 14.** Choose  $f \in [h]$ . Define:

$$\begin{aligned} W^{(f)} &= \{\mathbf{w} : (\mathbf{a}^{(f)})^\top \mathbf{w} = b^{(f)}, \mathbf{w} \in V\} \\ \tilde{W}^{(f)} &= \{\tilde{\mathbf{w}}_i : \mathbf{v}_i \in W^{(f)}\} \end{aligned}$$

We restrict the size of  $|W^{(f)}| = \ell$ , which is the number of points needed to define a halfspace in  $\mathbb{R}^\ell$ . This can be done by allowing for multiple identical  $\mathbf{a}_f, b_f$  combinations corresponding to all size  $\ell$  subsets of the  $v_i$  along the boundary.

Note that  $\text{Co}(W^{(f)})$  describes a “face” of the polytope  $\text{Co}(V)$  indexed by  $f$  which is perpendicular to  $\mathbf{a}^{(f)}$ .  $\text{Co}(\tilde{W}^{(f)})$  describes the perturbed face.

**Lemma 25.** *Choose  $f, g \in [h]$  arbitrarily and let  $W^{(f)} = \{\mathbf{w}_1^{(f)}, \dots, \mathbf{w}_\ell^{(f)}\}$  and  $\tilde{W}^{(f)} = \{\tilde{\mathbf{w}}_1^{(f)}, \dots, \tilde{\mathbf{w}}_\ell^{(f)}\}$ . For every  $\mathbf{m}^{(f)} \in \overline{\text{Co}}(W^{(f)})$ , we have  $(\tilde{\mathbf{a}}^{(g)})^\top \mathbf{m}^{(f)} < \tilde{b}^{(g)} + \varepsilon$ .*

*Proof:* Because  $m \in \overline{\text{Co}}(W^{(f)})$ , there is some  $\lambda \in \Delta_\ell$  with

$$\mathbf{m}^{(f)} = \sum_{i=1}^{\ell} \lambda_i \mathbf{w}_i^{(f)} \in \overline{\text{Co}}(W^{(f)}) \quad (24)$$

Consider also

$$\tilde{\mathbf{m}}^{(f)} = \sum_{i=1}^{\ell} \lambda_i \tilde{\mathbf{w}}_i^{(f)} \in \overline{\text{Co}}(\tilde{W}^{(f)}) \quad (25)$$

Note that the norm of the difference between these two vectors is bounded:

$$\begin{aligned} \|\mathbf{m}^{(f)} - \tilde{\mathbf{m}}^{(f)}\|_2 &= \left\| \sum_{i=1}^{\ell} \lambda_i (\mathbf{w}_i^{(f)} - \tilde{\mathbf{w}}_i^{(f)}) \right\|_2 \\ &\leq \sum_{i=1}^{\ell} \lambda_i \underbrace{\|\mathbf{w}_i^{(f)} - \tilde{\mathbf{w}}_i^{(f)}\|_2}_{< \varepsilon} < \varepsilon \end{aligned} \quad (26)$$

Also note that because  $\tilde{\mathbf{m}}^{(f)} \in \overline{\text{Co}}(\tilde{W}^{(f)}) \subseteq \overline{\text{Co}}(\tilde{V})$ , we have that  $(\tilde{\mathbf{a}}^{(g)})^\top \tilde{\mathbf{m}}^{(f)} \leq \tilde{b}^{(g)}$ . Now, a simple application of Cauchy-Schwartz gives:

$$\begin{aligned} (\tilde{\mathbf{a}}^{(g)})^\top \mathbf{m}^{(f)} &= (\tilde{\mathbf{a}}^{(g)})^\top (\tilde{\mathbf{m}}^{(f)} + (\mathbf{m}^{(f)} - \tilde{\mathbf{m}}^{(f)})) \\ &= \underbrace{(\tilde{\mathbf{a}}^{(g)})^\top \tilde{\mathbf{m}}^{(f)}}_{\leq \tilde{b}^{(g)}} + (\tilde{\mathbf{a}}^{(g)})^\top (\mathbf{m}^{(f)} - \tilde{\mathbf{m}}^{(f)}) \\ &\leq \tilde{b}^{(g)} + \|\tilde{\mathbf{a}}^{(g)}\|_2 \|\mathbf{m}^{(f)} - \tilde{\mathbf{m}}^{(f)}\|_2 \\ &< \tilde{b}^{(g)} + \varepsilon \end{aligned} \quad (27)$$

With this, we are now ready to prove Lemma 24. ■

*Proof:* Choose an arbitrary face  $g \in [h]$ . Recall we have  $\mathbf{x} \in \text{Co}(V)$  with  $(\mathbf{a}^{(g)})^\top \mathbf{x} < b - \varepsilon$  and we wish to show  $(\tilde{\mathbf{a}}^{(g)})^\top \mathbf{x} < \tilde{b}^{(g)}$ .

Let  $\mathbf{m}_x^{(f)}$  be the result of extending  $\tilde{\mathbf{a}}^{(g)}$  from  $\mathbf{x}$  to  $\text{Bo}(V)$ . This must hit some face with  $(\mathbf{a}^{(f)})^\top \mathbf{m}_x^{(f)} = b^{(f)}$ , so  $\mathbf{m}_x^{(f)} \in \text{Co}(W^{(f)})$ . That is, find  $\beta$  such that

$$\mathbf{m}_x^{(f)} = \beta \tilde{\mathbf{a}}^{(g)} + \mathbf{x} \in \text{Co}(W^{(f)}) \quad (28)$$

First, lets bound  $\beta$ . Notice that because  $\mathbf{m}_x^{(f)} \in \text{Co}(W^{(f)})$ , we have

$$\begin{aligned} (\mathbf{a}^{(f)})^\top \mathbf{m}_x^{(f)} &= (\mathbf{a}^{(f)})^\top \left( \sum_{i=1}^{\ell} \lambda_i \mathbf{w}_i^{(f)} \right) \\ &= \sum_{i=1}^{\ell} \lambda_i (\mathbf{a}^{(f)})^\top \mathbf{w}_i^{(f)} = b^{(f)} \end{aligned} \quad (29)$$

So, we have

$$b^{(f)} = (\mathbf{a}^{(f)})^\top \mathbf{m}_x^{(f)} = \beta \underbrace{(\mathbf{a}^{(f)})^\top \tilde{\mathbf{a}}^{(g)}}_{\leq 1} + \underbrace{(\mathbf{a}^{(f)})^\top \mathbf{x}}_{< b^{(f)} - \varepsilon} \Rightarrow \varepsilon < \beta \quad (30)$$

Now, apply Lemma 25

$$\begin{aligned} (\tilde{\mathbf{a}}^{(g)})^\top \mathbf{m}_x^{(f)} &< \tilde{b}^{(g)} + \varepsilon \\ (\tilde{\mathbf{a}}^{(g)})^\top \mathbf{x} + (\tilde{\mathbf{a}}^{(g)})^\top \tilde{\mathbf{a}}^{(g)} \beta &< \tilde{b}^{(g)} + \varepsilon \\ (\tilde{\mathbf{a}}^{(g)})^\top \mathbf{x} &< \tilde{b}^{(g)} \end{aligned} \quad (31)$$

Recall we chose face  $g \in [h]$  arbitrarily, so this holds for all halfspaces in the convex polytope. Hence, we have  $A\mathbf{x} < \mathbf{b}$ . ■

### B. Existence proof for decomposing expert cycles into LOGS

We can show that a decomposition of any curl consistent cycle into acyclic orientations must exist. Consider  $\mathbf{v}$  the vector of edge-weights of an arbitrary cycle with  $v^{(i)} = f_d^{(y^{(i)}, y^{(i+1)})}$ . Acyclic orientations in this framework correspond to the set  $A = \{0, 1\}^\ell \setminus \{\mathbf{0}_\ell, \mathbf{1}_\ell\}$ .

**Lemma 26.** *Given  $\mathbf{v} \in (0, 1)^\ell$ , such that  $1 < \|\mathbf{v}\|_1 < \ell - 1$ , then for some  $d(\cdot) : A \mapsto (0, 1)$  in  $\Delta_{2^\ell - 2}$ , we have*

$$\mathbf{v} = \sum_{\mathbf{t} \in G} d(\mathbf{t}) \mathbf{t} \quad (32)$$

*Proof:* Consider  $H = \{\mathbf{h} \in [0, 1]^\ell : 1 \leq \|\mathbf{h}\|_1 \leq \ell - 1\}$ . Note that  $\mathbf{v} \in H$ . Since  $A$  and  $H$  are closed and bounded sets, they are compact. Further,  $H$  is convex. Therefore,  $\text{Co}(A)$  is also convex and compact using compactness and Carathéodory's theorem [3]. We can now prove the required statement by showing  $H = \overline{\text{Co}}(A)$ .

- 1)  $\overline{\text{Co}}(A) \subset H$ : Let  $\mathbf{y} \in \overline{\text{Co}}(A)$ , then  $\mathbf{y} = \sum_{\mathbf{t} \in A} d(\mathbf{t}) \mathbf{t}$ , for some  $d(\cdot) : A \mapsto (0, 1)$  in  $\Delta_{2^\ell - 2}$ . We notice that  $\mathbf{y} \in [0, 1]^\ell$  and  $\|\mathbf{y}\|_1 \in [1, \ell - 1]$ . Therefore  $\mathbf{y} \in H$  and hence  $\overline{\text{Co}}(A) \subset H$ .
- 2)  $H \subset \overline{\text{Co}}(A)$ : Let  $E$  be the set of extreme points of  $H$ . Since  $H$  is convex and compact, we can use the Krein-Milman Theorem [3] to get  $H = \overline{\text{Co}}(E)$ . Further, we notice that  $E \subset A$  [3], therefore  $H \subset \overline{\text{Co}}(A)$ . ■

### C. Algorithmically decomposing expert cycles into LOGS

For this section, consider a knowledge cycle  $G = (\mathcal{Y}, E, f(\cdot))$  with edges  $y^{(1)} \rightarrow y^{(2)} \rightarrow \dots \rightarrow y^{(n)} \rightarrow y^{(1)}$ . Let the edge weights be represented by a vector  $\mathbf{f}_0 \in (0, 1)^n$  with  $f_0^{(i)} = f_d^{(y^{(i)}, y^{(i+1)})}$  for  $i < n$ , and  $f_0^{(n)} = f_d^{(y^{(n)}, y^{(1)})}$ . We will refer to *not fully decomposed* components as vectors  $\mathbf{f}_j \in [0, 1]^\ell$  and all *orientation* components as  $\mathbf{t}_j \in \{0, 1\}^\ell$  with weight  $d(\mathbf{t}_j)$ . We will also use  $\mathbf{1}_\ell[I] \in \{0, 1\}^\ell$  to denote a vector for which  $\mathbf{1}_\ell[I]^{(i)} = \mathbb{1}[i \in I]$ .

We begin by observing that all  $\mathbf{t}_j \in \{0, 1\}^\ell$  are allowed other than  $\mathbf{1}_\ell$  and the origin.

**Observation 27.** *We can decompose a scaled version of  $\mathbf{1}_\ell$  as follows:*

$$\left(1 - \frac{1}{\ell}\right) \mathbf{1}_\ell = \sum_{i=1}^{\ell-1} \frac{\mathbf{1}_\ell[\{j : j \neq i\}]}{\ell} = \sum_{i=1}^{\ell-1} \frac{\mathbf{1}_\ell[-\{j\}]}{\ell} \quad (33)$$

**Observation 28.** *If  $\|\mathbf{f}_j\|_1 = 1$ , we can decompose:*

$$\mathbf{f}_j = \sum_{i=1}^{\ell-1} f_j^{(i)} \mathbf{1}_\ell[\{i\}] \quad (34)$$

**Definition 15.** We define the support of vector  $\mathbf{f}_j \in [0, 1]^\ell$  to be the set of nonzero indices.

$$\text{Supp}(\mathbf{f}_j) = \{i : \mathbf{f}_j^{(i)} > 0\} \quad (35)$$

**Lemma 29.** *Let  $\gamma$  be a factor for handling the case of full support:*

$$\gamma = 1 - \frac{\mathbb{1}[|S_j| = \ell]}{\ell} \quad (36)$$

Given  $\mathbf{f}_j \in [0, 1]^\ell$  with  $\text{Supp}(\mathbf{f}_j) = S_j$  and  $\|\mathbf{f}_j\|_1 \in (1, \gamma|S_j|)$ , we can decompose

$$\mathbf{f}_j = z_j \gamma \mathbf{1}_\ell[S_j] + (1 - z_j) \mathbf{f}_{j+1} \quad (37)$$

where either  $\|\mathbf{f}_{j+1}\|_1 = 1$  or

- (i)  $|\text{Supp}(\mathbf{f}_{j+1})| \leq |S_j| - 1$ .
- (ii)  $\|\mathbf{f}_{j+1}\|_1 \in (1, |S_j| - 1)$ .

*Proof:* We begin by first observing

$$f_{j+1}^{(i)} = \frac{f_j^{(i)} - z_j \gamma \mathbb{1}[i \in S_j]}{1 - z_j} \leq \frac{f_j^{(i)} - z_j \mathbb{1}[i \in S_j]}{1 - z_j} \quad (38)$$

Recall that if  $f_j^{(i)} > 0$ , then  $i \in S_j$ . Hence  $\mathbf{f}_{j+1} \in [0, 1]^\ell$ . We notice that the upper bound of (ii)  $\|\mathbf{f}_{j+1}\|_1 \leq |S_j| - 1$  now follows from (i). It remains to find  $z_j$  such that  $\|\mathbf{f}_{j+1}\|_1 \geq 1$  and if  $\|\mathbf{f}_{j+1}\|_1 > 1$  then (i) holds.

$$\begin{aligned} \zeta_a &= \frac{1}{\gamma} \min_{i \in S_j} f_j^{(i)} \\ \zeta_b &= \frac{\|\mathbf{f}_j\|_1 - 1}{\gamma|S_j| - 1} \\ z_j &= \min(\zeta_a, \zeta_b) \end{aligned} \quad (39)$$

We can apply the Pigeonhole principle to  $\sum_{i \in S_j} f_j^{(i)} \leq \ell - 1$  to get  $z_j \leq \zeta_a \leq 1$ . If  $z_j = 1$  then the second term of the decomposition is irrelevant. We can treat the L1 norm as a linear function when the vectors are all the same sign giving:

$$\|\mathbf{f}_{j+1}\|_1 = \frac{\|\mathbf{f}_j\|_1 - z_j \gamma |S_j|}{1 - z_j} \quad (40)$$

Now consider two cases:

- 1) ( $z_j = \zeta_b$ ) Notice  $\zeta_b$  has been carefully chosen so that,

$$\|\mathbf{f}_{j+1}\|_1 = \frac{\|\mathbf{f}_j\|_1 - \zeta_b \gamma |S_j|}{1 - \zeta_b} = 1. \quad (41)$$

- 2) ( $z_j = \zeta_a$ ) Notice that if  $f_j^{(i^*)} = \min_{i \in S_j} f_j^{(i)}$ , then  $f_{j+1}^{(i^*)} = 0$ , which satisfies condition (i).

To show the lower bound of (ii), rewrite Equation 40:

$$\|\mathbf{f}_{j+1}\|_1 = \|\mathbf{f}_j\|_1 \frac{1 - z_j \left( \frac{\gamma |S_j|}{\|\mathbf{f}_j\|_1} \right)}{1 - z_j} \quad (42)$$

Recall that  $\|\mathbf{f}_j\|_1 < \gamma |S_j|$ , so  $\frac{\gamma |S_j|}{\|\mathbf{f}_j\|_1} > 1$ . So, if

Equation 42 gives  $\|\mathbf{f}_{j+1}\|_1 = 1$  with  $z_j = \zeta_b$ , then  $z_j \leq \zeta_b$  gives  $\|\mathbf{f}_{j+1}\|_1 \geq 1$ . ■

**Lemma 30.** *We can decompose any expert cycle  $G_d^P = (\mathcal{Y}, E, f_d(\cdot))$  into a linear combination of elements of  $\mathcal{H}_{\mathcal{R}}$ :*

$$G_d^P = \sum_{r \in \mathcal{R}} w(r) H_r$$

As usual, this means  $f_d(e) = \sum_{r \in \mathcal{R}} w(r) f_r(e) \forall e \in E$ .

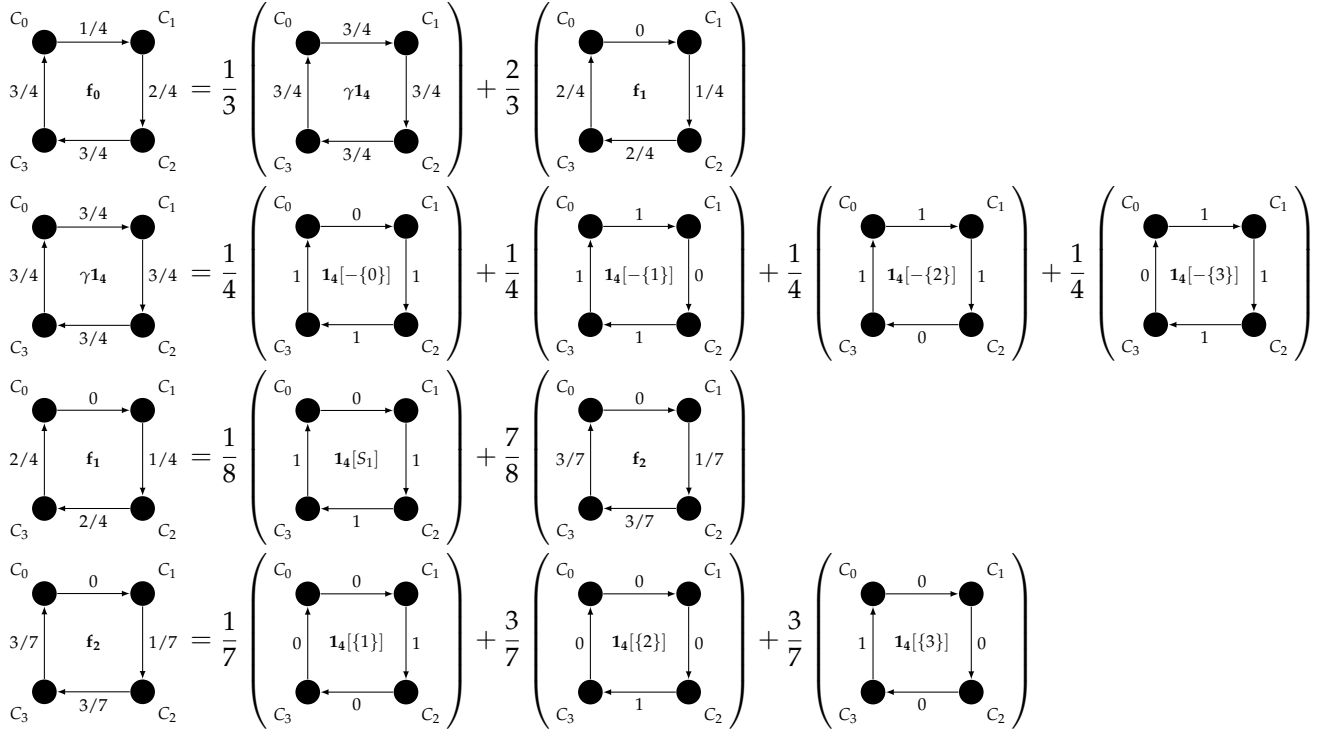


Fig. 11. Decomposing an expert graph  $\mathbf{f}_0$  following the procedure given in Appendix Section C. The first line pulls off a scaled  $\mathbf{1}_4$ . The second line decomposes this by Observation 27. The third line decomposes  $\mathbf{f}_1$  into  $\mathbf{1}_4[S_1]$  and  $\mathbf{f}_2$  with  $\|\mathbf{f}_2\|_1 = 1$ . The fourth line decomposes  $\mathbf{f}_2$  according to Observation 28.

*Proof:* Recall that edge-weight vectors for  $H_r$  can be all  $\mathbf{t}_j \in \{0,1\}^\ell$  other than  $\mathbf{1}_\ell$  and the origin. The task is now to decompose the edge-weight vector  $\mathbf{f}_0 = (f_d(y^{(1)}, y^{(2)}), f_d(y^{(2)}, y^{(3)}), \dots, f_d(y^{(\ell)}, y^{(1)}))^\top$  into a convex combination of  $\mathbf{t}_j$ . Curl consistency gives  $\|\mathbf{f}_0\|_1 \in (1, \ell - 1)$ .

Use Lemma 29 to decompose

$$\mathbf{f}_0 = z_0 \gamma \mathbf{1}_\ell[S_0] + (1 - z_0) \mathbf{f}_1 \quad (43)$$

If  $\mathbf{f}_0$  had full support, then  $\gamma = 1 - \frac{1}{\ell-1}$  and  $\mathbf{1}_\ell[S_0] = \mathbf{1}_\ell$ . Thus, we can use Observation 27 to decompose  $\gamma \mathbf{1}_\ell$  into acyclic orientations. If  $\text{Supp}(\mathbf{f}_0) < \ell$ , then  $\gamma = 1$  and  $\mathbf{1}_\ell[S_0]$  already represents an acyclic orientation.

Repeatedly apply Lemma 29. In this process if we reach some  $\mathbf{f}_j$  with  $\|\mathbf{f}_j\|_1 = 1$ , we can apply Observation 28 to decompose  $\mathbf{f}_j$  into acyclic orientations. If we do not, we will terminate at  $\mathbf{f}_\ell$  with  $|S_\ell| = 0$ . See Figure 11 for an example of this decomposition. ■