

## Proofs : Local Algorithms for Approximate Inference in Minor-Excluded Graphs

**Proof of Lemma 1.** This essentially follows from arguments in [10, 11]. We sketch proof of the property (a) for  $\mathcal{B}$  being  $(r/\Delta, O(\Delta))$ -decomposition; the proof of (b) follows from Theorem 4.2 [10]. To see property (a), consider an edge  $e \in E$ . If  $e \notin \mathcal{B}$  in the beginning of iteration  $0 \leq i \leq r-1$ , then it will be present exactly once in a breadth-first tree, say  $T_j^i$ . This edge will be chosen in  $\mathcal{B}_j^i$  only if it is at level  $L_j^i + k\Delta, k \geq 0$ . The probability of this is at most  $1/\Delta$  since  $L_j^i$  is chosen u.a.r. from  $\{0, \dots, \Delta-1\}$ . By union bound, it follows that the probability that an edge is chosen in any of the  $r$  iterations is at most  $r/\Delta$ . This completes the proof of Lemma 1.  $\square$

**Proof of Lemma 2.** First, we prove properties of  $\log \hat{Z}_{\text{LB}}, \log \hat{Z}_{\text{UB}}$  as follows:

$$\begin{aligned}
\log \hat{Z}_{\text{LB}} &\stackrel{(o)}{=} \sum_{j=1}^K \log Z_j + \sum_{(i,j) \in \mathcal{B}} \psi_{ij}^L \\
&\stackrel{(a)}{=} \log \left[ \sum_{\mathbf{x} \in \Sigma^n} \exp \left( \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E \setminus \mathcal{B}} \psi_{ij}(x_i, x_j) + \sum_{(i,j) \in \mathcal{B}} \psi_{ij}^L \right) \right] \\
&\stackrel{(b)}{\leq} \log \left[ \sum_{\mathbf{x} \in \Sigma^n} \exp \left( \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E \setminus \mathcal{B}} \psi_{ij}(x_i, x_j) + \sum_{(i,j) \in \mathcal{B}} \psi_{ij}(x_i, x_j) \right) \right] = \log Z \\
&\stackrel{(c)}{\leq} \log \left[ \sum_{\mathbf{x} \in \Sigma^n} \exp \left( \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E \setminus \mathcal{B}} \psi_{ij}(x_i, x_j) + \sum_{(i,j) \in \mathcal{B}} \psi_{ij}^U \right) \right] \\
&\stackrel{(d)}{=} \sum_{j=1}^K \log Z_j + \sum_{(i,j) \in \mathcal{B}} \psi_{ij}^U \\
&= \log \hat{Z}_{\text{UB}}. \tag{2}
\end{aligned}$$

We justify (a)-(d) as follows: (a) holds because by removal of edges  $\mathcal{B}$ , the  $G$  decomposes into disjoint connected components  $S_1, \dots, S_K$ ; (b) holds because of the definition of  $\psi_{ij}^L$ ; (c) holds by definition  $\psi_{ij}^U$  and (d) holds for a similar reason as (a). The claim about difference  $\log \hat{Z}_{\text{UB}} - \log \hat{Z}_{\text{LB}}$  in the statement of Lemma 2 follows directly from definitions (i.e. subtract RHS (o) from (d)). This completes proof of claimed relation between bounds  $\log \hat{Z}_{\text{LB}}, \log \hat{Z}_{\text{UB}}$ .

For running time analysis, note that LOG PARTITION performs two main tasks: (i) Decomposing  $G$  using DECOMP algorithm, which by definition take  $T_{\text{DECOMP}}$  time. (ii) Computing  $Z_j$  for each component  $S_j$  through exhaustive computation, which takes  $O(|E| \Sigma^{|S_j|})$  time and producing  $\log \hat{Z}_{\text{LB}}, \log \hat{Z}_{\text{UB}}$  takes addition  $|E|$  operations at the most. Since there are  $K$  components in total with max-size of component being  $|S^*|$  we obtain that running time for this task is  $O(|E|K \Sigma^{|S^*|})$ . Putting (i) and (ii) together, we obtain the desired bound. This completes the proof of Lemma 2.  $\square$

**Proof of Lemma 3.** Assign weight  $w_{ij} = \psi_{ij}^U - \psi_{ij}^L$  to an edge  $(i, j) \in E$ . Since graph has maximum vertex degree  $D$ , by Vizing's theorem there exists an edge-coloring of the graph using at most  $D+1$  colors. Edges with the same color form a matching of the  $G$ . A standard application of Pigeon-hole's principle implies that there is a color with weight at least  $\frac{1}{D+1} (\sum_{(i,j) \in E} w_{ij})$ . Let  $M \subset E$  denote these set of edges. That is,

$$\sum_{(i,j) \in M} (\psi_{ij}^U - \psi_{ij}^L) \geq \frac{1}{D+1} \left( \sum_{(i,j) \in E} (\psi_{ij}^U - \psi_{ij}^L) \right).$$

Now, consider a  $Q \subset \Sigma^n$  of size  $2^{|M|}$  created as follows. For  $(i, j) \in M$  let  $(x_i^U, x_j^U) \in \arg \max_{(x, x') \in \Sigma^2} \psi_{ij}(x, x')$ . For each  $i \in V$ , choose  $x_i^L \in \Sigma$  arbitrarily. Then,

$$Q = \{ \mathbf{x} \in \Sigma^n : \forall (i, j) \in M, (x_i, x_j) = (x_i^U, x_j^U) \text{ or } (x_i^L, x_j^L); \text{ for all other } i \in V, x_i = x_i^L \}.$$

Note that we have used the fact that  $M$  is a matching for  $Q$  to be well-defined.

By definition  $\phi_i, \psi_{ij}$  are non-negative function (hence, their exponents are at least 1). Using this property, we have the following:

$$\begin{aligned}
Z &\geq \left[ \sum_{\mathbf{x} \in Q} \exp \left( \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E} \psi_{ij}(x_i, x_j) \right) \right] \\
&\stackrel{(o)}{\geq} \left[ \sum_{\mathbf{x} \in Q} \exp \left( \sum_{(i,j) \in M} \psi_{ij}(x_i, x_j) \right) \right] \\
&\stackrel{(a)}{\geq} 2^{|M|} \prod_{(i,j) \in M} \frac{\exp(\psi_{ij}^L) + \exp(\psi_{ij}^U)}{2} \\
&\stackrel{(b)}{=} \prod_{(i,j) \in M} (1 + \exp(\psi_{ij}^U - \psi_{ij}^L)) \exp(\psi_{ij}^L) \\
&\stackrel{(c)}{\geq} \prod_{(i,j) \in M} \exp(\psi_{ij}^U - \psi_{ij}^L) = \exp \left( \sum_{(i,j) \in M} \psi_{ij}^U - \psi_{ij}^L \right). \tag{3}
\end{aligned}$$

Justification of (o)-(c): (o) follows since  $\psi_{ij}, \phi_i$  are non-negative functions. (a) consider the following probabilistic experiment: assign  $(x_i, x_j)$  for each  $(i, j) \in M$  equal to  $(x_i^U, x_j^U)$  or  $(x_i^L, x_j^L)$  with probability  $1/2$  each. Under this experiment, the expected value of the  $\exp(\sum_{(i,j) \in M} \psi_{ij}(x_i, x_j))$ , which is  $\prod_{(i,j) \in M} \frac{\exp(\psi_{ij}(x_i^L, x_j^L)) + \exp(\psi_{ij}(x_i^U, x_j^U))}{2}$ , is equal to  $2^{-|M|} [\sum_{\mathbf{x} \in Q} \exp(\sum_{(i,j) \in M} \psi_{ij}(x_i, x_j))]$ . Now, use the fact that  $\psi_{ij}(x_i^L, x_j^L) \geq \psi_{ij}^L$ . (b) follows from simple algebra and (c) follows by using non-negativity of function  $\psi_{ij}$ . Therefore,

$$\log Z \geq \sum_{(i,j) \in M} (\psi_{ij}^U - \psi_{ij}^L) \geq \frac{1}{D+1} \left( \sum_{(i,j) \in E} (\psi_{ij}^U - \psi_{ij}^L) \right), \tag{4}$$

using fact about weight of  $M$ . This completes the proof of Lemma 3.  $\square$

**Proof of Lemma 4.** From Lemma 2, Lemma 3 and definition of  $(\delta, \Delta)$ -decomposition, we have the following.

$$\begin{aligned}
\mathbb{E} \left[ \log \hat{Z}_{UB} - \log \hat{Z}_{LB} \right] &\leq \mathbb{E} \left[ \sum_{(i,j) \in \mathcal{B}} (\psi_{ij}^U - \psi_{ij}^L) \right] = \sum_{(i,j) \in E} \Pr((i,j) \in \mathcal{B}) (\psi_{ij}^U - \psi_{ij}^L) \\
&\leq \delta \left[ \sum_{(i,j) \in E} (\psi_{ij}^U - \psi_{ij}^L) \right] \leq \delta(D+1) \log Z. \tag{5}
\end{aligned}$$

Now to estimate the running time, note that under  $(\delta, \Delta)$  decomposition  $\mathcal{B}$ , with probability 1 the  $G' = (V, E \setminus \mathcal{B})$  is divided into connected components with diameter at most  $\Delta$  with respect to  $G'$ . Since maximum vertex degree is  $D$ , it follows easily that each of these component has at most  $D^\Delta$  nodes. Now, the running time bound of Lemma 2 implies the desired result.  $\square$

**Proof of Theorem 1.** The justification about the estimates  $\log \hat{Z}_{LB}, \log \hat{Z}_{UB}$  follows from  $(r/\Delta, O(\Delta))$ -decomposition property of **DeC** algorithm (Lemma 1) and Lemma 4. The bound on running time follows from Lemma 4 as well.  $\square$

**Proof of Lemma 5.** By definition of MAP  $\mathbf{x}^*$ , we have  $\mathcal{H}(\widehat{\mathbf{x}}^*) \leq \mathcal{H}(\mathbf{x}^*)$ . Now, consider the following.

$$\begin{aligned}
\mathcal{H}(\mathbf{x}^*) &= \max_{\mathbf{x} \in \Sigma^n} \left[ \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E} \psi_{ij}(x_i, x_j) \right] \\
&= \max_{\mathbf{x} \in \Sigma^n} \left[ \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E \setminus \mathcal{B}} \psi_{ij}(x_i, x_j) + \sum_{(i,j) \in \mathcal{B}} \psi_{ij}(x_i, x_j) \right] \\
&\stackrel{(a)}{\leq} \max_{\mathbf{x} \in \Sigma^n} \left[ \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E \setminus \mathcal{B}} \psi_{ij}(x_i, x_j) + \sum_{(i,j) \in \mathcal{B}} \psi_{ij}^U \right] \\
&\stackrel{(b)}{=} \sum_{j=1}^K \left[ \max_{\mathbf{x}^j \in \Sigma^{|\mathcal{S}_j|}} \mathcal{H}(\mathbf{x}^j) \right] + \left[ \sum_{(i,j) \in \mathcal{B}} \psi_{ij}^U \right] \\
&\stackrel{(c)}{=} \sum_{j=1}^K \mathcal{H}(\mathbf{x}^{*,j}) + \left[ \sum_{(i,j) \in \mathcal{B}} \psi_{ij}^U \right] \\
&\stackrel{(d)}{\leq} \mathcal{H}(\widehat{\mathbf{x}}^*) + \left[ \sum_{(i,j) \in \mathcal{B}} \psi_{ij}^U - \psi_{ij}^L \right]. \tag{6}
\end{aligned}$$

We justify (a)-(d) as follows: (a) holds because for each edge  $(i, j) \in \mathcal{B}$ , we have replaced its effect by maximal value  $\psi_{ij}^U$ ; (b) holds because by placing constant value  $\psi_{ij}^U$  over  $(i, j) \in \mathcal{B}$ , the maximization over  $G$  decomposes into maximization over the connected components of  $G' = (V, E \setminus \mathcal{B})$ ; (c) holds by definition of  $\mathbf{x}^{*,j}$  and (d) holds because when we obtain global assignment  $\mathbf{x}^*$  from  $\mathbf{x}^{*,j}$ ,  $1 \leq j \leq K$  and compute its global value, the additional terms get added for each  $(i, j) \in \mathcal{B}$  which add at least  $\psi_{ij}^L$  amount.

The running time analysis of MODE is exactly the same as that of LOG PARTITION in Lemma 2. Hence, we skip the details here. This completes the proof of Lemma 5.  $\square$

**Proof of Lemma 6.** Assign weight  $w_{ij} = \psi_{ij}^U$  to an edge  $(i, j) \in E$ . Using argument of Lemma 3, we obtain that there exists a matching  $M \subset E$  such that

$$\sum_{(i,j) \in M} \psi_{ij}^U \geq \frac{1}{D+1} \left( \sum_{(i,j) \in E} \psi_{ij}^U \right).$$

Now, consider an assignment  $\mathbf{x}^M$  as follows: for each  $(i, j) \in M$  set  $(x_i^M, x_j^M) = \arg \max_{(x, x') \in \Sigma^2} \psi_{ij}(x, x')$ ; for remaining  $i \in V$ , set  $x_i^M$  to some value in  $\Sigma$  arbitrarily. Note that for above assignment to be possible, we have used matching property of  $M$ . Therefore, we have

$$\begin{aligned}
\mathcal{H}(\mathbf{x}^M) &= \sum_{i \in V} \phi_i(x_i^M) + \sum_{(i,j) \in E} \psi_{ij}(x_i^M, x_j^M) \\
&= \sum_{i \in V} \phi_i(x_i^M) + \sum_{(i,j) \in E \setminus M} \psi_{ij}(x_i^M, x_j^M) + \sum_{(i,j) \in M} \psi_{ij}(x_i^M, x_j^M) \\
&\stackrel{(a)}{\geq} \sum_{(i,j) \in M} \psi_{ij}(x_i^M, x_j^M) \\
&= \sum_{(i,j) \in M} \psi_{ij}^U \\
&\geq \frac{1}{D+1} \left[ \sum_{(i,j) \in E} \psi_{ij}^U \right]. \tag{7}
\end{aligned}$$

Here (a) follows because  $\psi_{ij}, \phi_i$  are non-negative valued functions. Since  $\mathcal{H}(\mathbf{x}^*) \geq \mathcal{H}(\mathbf{x}^M)$  and  $\psi_{ij}^L \geq 0$  for all  $(i, j) \in E$ , we obtain the Lemma 6.  $\square$

**Proof of Lemma 7.** From Lemma 5, Lemma 6 and definition of  $(\delta, \Delta)$ -decomposition, we have the following.

$$\begin{aligned}
\mathbb{E} \left[ \mathcal{H}(\mathbf{x}^*) - \mathcal{H}(\widehat{\mathbf{x}}^*) \right] &\leq \mathbb{E} \left[ \sum_{(i,j) \in \mathcal{B}} (\psi_{ij}^U - \psi_{ij}^L) \right] \\
&= \sum_{(i,j) \in E} \Pr((i,j) \in \mathcal{B}) (\psi_{ij}^U - \psi_{ij}^L) \\
&\leq \delta \left[ \sum_{(i,j) \in E} (\psi_{ij}^U - \psi_{ij}^L) \right] \\
&\leq \delta(D+1)\mathcal{H}(\mathbf{x}^*). \tag{8}
\end{aligned}$$

The running time bound can be obtained using arguments similar to those in Lemma 4.  $\square$

**Proof of Theorem 2.** The justification about the bound on estimate  $\mathcal{H}(\widehat{\mathbf{x}}^*)$  follows from  $(r/\Delta, O(\Delta))$ -decomposition property of **DeC** algorithm (Lemma 1) and Lemma 7. The bound on running time follows from Lemma 7 as well.  $\square$