

A Algorithm Proof

In this section, we present the detailed proofs for the formula in Algorithm 1. Note that the symbols follow the definition in the main paper.

Proof 1 Given one vertex as i , the Eq. 8 proves the backward process of $\frac{\partial \text{loss}}{\partial \mathbf{x}}$ in Algorithm 1.

$$\begin{aligned} \frac{\partial \text{loss}}{\partial \mathbf{x}_i} &= \frac{\partial f(\mathbf{x}_i)}{\partial \mathbf{x}_i} \sum_{j \in \Omega} \frac{\partial \text{loss}}{\partial \mathbf{y}_j} \frac{\partial \mathbf{y}_j}{\partial \mathbf{x}_i} \\ &= \frac{\partial f(\mathbf{x}_i)}{\partial \mathbf{x}_i} \sum_{j \in \Omega} S(\mathbf{E}_{i,j}) \left(\frac{\partial \text{loss}}{\partial \mathbf{y}_j} \frac{1}{z_j} \right) \\ &= \frac{\partial f(\mathbf{x}_i)}{\partial \mathbf{x}_i} \boldsymbol{\psi}_i \end{aligned} \quad (8)$$

Proof 2 Given one edge with a pair of connected vertices i and j , the Eq. 9-11 proves the backward process of $\frac{\partial \text{loss}}{\partial \boldsymbol{\omega}}$ in Algorithm 1, where Ω_i indicates the set of children of vertex i in the tree whose root is j and Ω_j indicates the set of children of vertex j in the tree whose root is i .

$$\begin{aligned} \frac{\partial \text{loss}}{\partial \boldsymbol{\omega}_{i,j}} &= \frac{\partial S(\mathbf{E}_{i,j})}{\partial \boldsymbol{\omega}_{i,j}} \sum_{m \in \Omega} \sum_{m \in \Omega} \frac{\partial \text{loss}}{\partial \mathbf{y}_m} \frac{\partial \mathbf{y}_m}{\partial S(\mathbf{E}_{i,j})} \\ &= \frac{\partial S(\mathbf{E}_{i,j})}{\partial \boldsymbol{\omega}_{i,j}} \sum_{m \in \Omega} \left(\sum_{m \in \Omega} \frac{1}{z_m} \frac{\partial \text{loss}}{\partial \mathbf{y}_m} \frac{\partial z_m \mathbf{y}_m}{\partial S(\mathbf{E}_{i,j})} - \sum_{m \in \Omega} \frac{\partial \text{loss}}{\partial \mathbf{y}_m} \frac{\mathbf{y}_m}{z_m} \frac{\partial z_m}{\partial S(\mathbf{E}_{i,j})} \right) \\ &= \frac{\partial S(\mathbf{E}_{i,j})}{\partial \boldsymbol{\omega}_{i,j}} \sum (\gamma_i^s - \gamma_i^z) \end{aligned} \quad (9)$$

The computational complexity of the component γ_i^s can be reduce to linear by the dynamic programming procedure below.

$$\begin{aligned} \gamma_i^s &= \sum_{m \in \Omega} \frac{\partial \text{loss}}{\partial \mathbf{y}_m} \frac{1}{z_m} \frac{\partial z_m \mathbf{y}_m}{\partial S(\mathbf{E}_{i,j})} \\ &= \sum_{m \in \Omega} \left(\frac{\phi}{z} \right)_m \sum_{k \in \Omega} \frac{\partial S(\mathbf{E}_{m,k}) f(\mathbf{x}_k)}{\partial S(\mathbf{E}_{i,j})} \\ &= \sum_{m \in \Omega_i} \left(\frac{\phi}{z} \right)_m \sum_{k \in \Omega_j} S(\mathbf{E}_{m,i}) S(\mathbf{E}_{i,k}) f(\mathbf{x}_k) + \sum_{m \in \Omega_j} \left(\frac{\phi}{z} \right)_m \sum_{k \in \Omega_i} S(\mathbf{E}_{m,j}) S(\mathbf{E}_{j,k}) f(\mathbf{x}_k) \\ &= \sum_{m \in \Omega_i} S(\mathbf{E}_{m,i}) \left(\frac{\phi}{z} \right)_m \sum_{k \in \Omega_j} S(\mathbf{E}_{i,k}) f(\mathbf{x}_k) + \sum_{m \in \Omega_j} S(\mathbf{E}_{m,j}) \left(\frac{\phi}{z} \right)_m \sum_{k \in \Omega_i} S(\mathbf{E}_{j,k}) f(\mathbf{x}_k) \\ &= \hat{\boldsymbol{\psi}}_i \cdot \hat{\boldsymbol{\rho}}_j + \hat{\boldsymbol{\psi}}_j \cdot \hat{\boldsymbol{\rho}}_i \\ &= \hat{\boldsymbol{\psi}}_i \cdot (\boldsymbol{\rho}_i - S(\mathbf{E}_{i,j}) \hat{\boldsymbol{\rho}}_i) + \hat{\boldsymbol{\rho}}_i \cdot (\boldsymbol{\psi}_i - S(\mathbf{E}_{j,i}) \hat{\boldsymbol{\psi}}_i) \\ &= \hat{\boldsymbol{\psi}}_i \cdot \boldsymbol{\rho}_i + \boldsymbol{\psi}_i \cdot \hat{\boldsymbol{\rho}}_i - 2S(\mathbf{E}_{i,j}) \hat{\boldsymbol{\psi}}_i \cdot \hat{\boldsymbol{\rho}}_i \end{aligned} \quad (10)$$

The same procedure can be easily adapted to obtain the component γ_i^z .

$$\gamma_i^z = \sum_{m \in \Omega} \frac{\partial \text{loss}}{\partial \mathbf{y}_m} \frac{\mathbf{y}_m}{z_m} \frac{\partial z_m}{\partial S(\mathbf{E}_{i,j})} = \hat{\boldsymbol{\nu}}_i z_i + \boldsymbol{\nu}_i \hat{z}_i - 2S(\mathbf{E}_{i,j}) \hat{\boldsymbol{\nu}}_i \hat{z}_i \quad (11)$$