Entity Hierarchy Embedding: Supplementary Material

1 Proof of Theorem 1

In this section we prove Theorem 1 (Section 2.2):

Theorem 1. $\forall h \in \mathcal{A}_e \cap \mathcal{A}_{e'}$, $h \in \mathcal{Q}_{e,e'}$ *iff it satisfies the two conditions: (1)* $|\mathcal{C}_h \cap (\mathcal{A}_e \cup \mathcal{A}_{e'})| \geq 2$*; (2)* $\exists a, b \in \mathcal{C}_h \cap (\mathcal{A}_e \cup \mathcal{A}_{e'})$ *s.t.* $t_a \neq t_b$*.*

Recall that $\mathcal{Q}_{e,e'}$ is the set of common ancestors of entity e and e' that are turning nodes of any $e \to e'$ paths; $\mathcal{A}_e$ is the ancestor nodes of entity e (including e itself); for a node $h \in \mathcal{A}_e \cup \mathcal{A}_{e'}$, its critical node t_h is the nearest (w.r.t the length of the shortest path) descendant of h (including h itself) that is in $\mathcal{Q}_{e,e'} \cup \{e, e'\}$; $\mathcal{C}_h$ be the set of immediate child nodes of h.

Lemma 2. $\forall h \in \mathcal{A}_e \cap \mathcal{A}_{e'}$, $t_h \in \mathcal{Q}_{e,e'}$.

Proof. $h \in \mathcal{A}_e \cap \mathcal{A}_{e'} \Rightarrow (h \in \mathcal{A}_e) \wedge (h \in \mathcal{A}_{e'})$.

As $h \in \mathcal{A}_e$, there's path $e \to \cdots \to h$ where the consecutive nodes are (child, parent) pairs. Similarly, there exists path $h \to \cdots \to e'$ where the consecutive nodes are (parent, child) pairs. Denote the set of intersections of the two paths as $\mathcal{I}$. Because the two paths intersects at h, $\mathcal{I} \neq \phi$.

Note that the nodes in the intersection set are also in the path $h \to \cdots \to e'$, so we can sort the nodes in $\mathcal{I}$ according to the topological order in path $h \to \cdots \to e'$. Denote the topologically lowest node in $\mathcal{I}$ as t. As t is in the intersection set of two paths, there exists path $e \to \cdots \to t$ where the consecutive nodes are (child, parent) pairs and path $t \to \cdots \to e'$ where the consecutive nodes are (parent, child) pairs. If the two paths $e \to \cdots \to t$ and $t \to \cdots \to e'$ have any intersections except for t, then the intersection will be topologically lower than t, which contradicts the definition of t. So paths $e \to \cdots \to t$ and $t \to \cdots \to e'$ have intersection only at t, so t is a turning node. So $Q_{e,e'} \neq \phi$. According to the construction of t, t is a descendant of h, therefore $t_h \in Q_{e,e'}$. $\square$

We next prove Theorem 1.

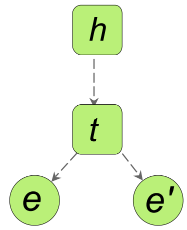

Figure 1: Illustration for Lemma 2. The topologically lowest intersection node is a turning node, which is also a descendant of h.

Proof. **Sufficiency**: Note that $e, e' \notin Q_{e,e'}$, we prove by enumerating possible situations: (i) $t_a = e, t_b = e'$, (ii) $t_a = e, t_b \in Q_{e,e'}$, (iii) $t_a, t_b \in Q_{e,e'}$. Case $t_a = e, t_b = e'$ is equivalent to case (i) if we swap e and e', and the cases $t_a = e', t_b \in Q_{e,e'}$, $t_a \in Q_{e,e'}, t_b = e(e')$ are equivalent to case (ii) if we swap the notations for variables a, b, e, e' properly. So the proof for cases (i), (ii) and (iii) is sufficient. An illustration of the cases is provided in Figure 2.

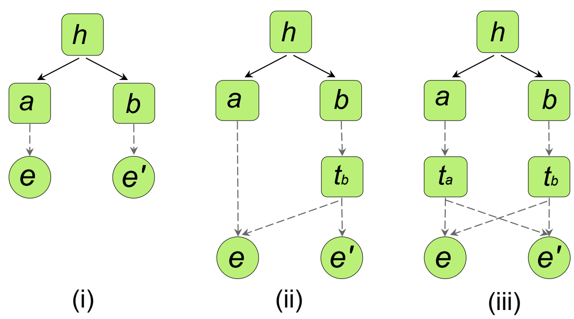

Figure 2: Three cases: (i) $t_a = e, t_b = e'$; (ii) $t_a = e, t_b \in Q_{e,e'}$; (iii) $t_a, t_b \in Q_{e,e'}$.

(i) $t_a = e, t_b = e'$:
As $t_a = e$, there's a path $e \to \cdots \to a \to h$ where the consecutive nodes are (child, parent) pairs. Similarly, there's a path $h \to b \to \cdots \to e'$ where the consecutive nodes are (parent, child) pairs. The above two paths only intersect at h, otherwise as a is the topologically highest node in path $e \to \cdots \to a \to h$ except for h, and e' is the topologically lowest node in path $h \to b \to \cdots \to e'$, e' would be a descendant of a. According to Lemma 2, $t_a \in Q_{e,e'}$, which contradicts $t_a = e$. So the two paths only intersect at h, and we can combine the two paths to construct a valid path $e \to \cdots \to a \to h \to b \to \cdots \to e'$, yielding h as a turning node.

(ii) $t_a = e, t_b \in Q_{e,e'}$:
$t_a = e \Rightarrow \exists e \to \cdots \to a \to h$ where the consecutive nodes are (child, parent) pairs. As $t_b \in Q_{e,e'}$, there exists path $h \to b \to \cdots \to t_b \to \cdots \to e'$ where the consecutive nodes are (parent, child) pairs. If the two paths $e \to \cdots \to a \to h$ and $h \to b \to \cdots \to t_b \to \cdots \to e'$ has any intersections except for h, then e' will be a descendant of a, thus $a \in \mathcal{A}_e \cup \mathcal{A}_{e'}$. According to Lemma 2, $t_a \in \mathcal{Q}_{e,e'}$, which contradicts the assumption that $t_a = e \notin \mathcal{Q}_{e,e'}$. So path $e \to \cdots \to a \to h \to b \to \cdots \to t_b \to \cdots \to e'$ is a valid path, yielding h as a turning node.

(iii) $t_a, t_b \in Q_{e,e'}$:
First of all, we prove that there exists path $e(e') \to \cdots \to t_a$ where the consecutive nodes are (child, parent) pairs and path $t_b \to \cdots \to e'(e)$ where the consecutive nodes are (parent, child) pairs and the two paths do not intersect with each other. If $t_b \to \cdots \to e'$ does not intersect with $e \to \cdots \to t_a$ (the existence of the paths is due to the definition of turning node), we've already got the construction. Otherwise, if $t_b \to \cdots \to e'$ intersects with $t_a \to \cdots \to e'$ at x before it intersects with $e \to \cdots \to t_a$, the path $e \to \cdots \to t_a$ and path $t_b \to \cdots \to x \to \cdots \to e'$ where the part $x \to \cdots \to e'$ is subpath of $t_a \to \cdots \to e'$ satisfies the above requirements. Similarly, if $t_b \to \cdots \to e'$ intersects with $e \to \cdots \to t_a$ at x before it intersects with $t_a \to \cdots \to e'$, the path $e' \to \cdots \to t_a$ and path $t_b \to \cdots \to x \to \cdots \to e$ where the part $x \to \cdots \to e$ is subpath of $t_a \to \cdots \to e$ satisfies the above requirements.

Using the above conclusion, if path $t_a \to \cdots \to a \to h$ (we choose the shortest path in the part $t_a \to \cdots \to a$ if there are multiple paths) intersects with $h \to b \to \cdots \to t_b$ (similarly, we choose the shortest path in the part $b \to \cdots \to t_b$) at any node except for h, we denote the topologically lowest one (w.r.t. path $h \to b \to \cdots \to t_b$) as x, then $t_a \to \cdots \to x$ has no intersection with $x \to \cdots \to t_b$ except for x, as any such intersection will be lower than x. So the path $e(e') \to \cdots \to t_a \to \cdots \to x \to \cdots \to t_b \to \cdots \to e'$ is a valid path, making x a turning node. As $t_a \neq t_b$, we have $(x \neq t_a) \vee (x \neq t_b)$. If $x \neq t_a$, x is closer to a as we've chosen the shortest path in part $t_a \to \cdots \to a$, contradicting the definition of t_a. Similarly, it is also impossible that $x \neq t_b$. So the two paths $t_a \to \cdots \to a \to h$ and $h \to b \to \cdots \to t_b$ do not intersect with each other.

Putting the above conclusions together, we can construct a valid path $e(e') \to \cdots \to t_a \to \cdots \to a \to h \to b \to \cdots \to t_b \to \cdots \to e'$, making h a turning node. Note that we also need to prove that the path $e \to \cdots \to t_a$ does not intersect with path $h \to b \to \cdots \to t_b$, which is analogous to the proof that path $t_a \to \cdots \to a \to h$ intersects with $h \to b \to \cdots \to t_b$ only at h.

Necessity: If h was a turning node, there would be a path $e \to \cdots a \to h \to b \to \cdots \to e'$, where the consecutive nodes before h are (child, parent) pairs and (parent, child) pairs after h, and we denote the two direct children of h in the path as a and b, in which a is ascendant of e (or e itself) and b ascendant of e' (or e' itself). So $|\mathcal{C}_h \cap (\mathcal{A}_e \cup \mathcal{A}_{e'})| \geq |\{a, b\}| = 2$.

Then we prove that $\exists a, b \in \mathcal{C}_h \cap (\mathcal{A}_e \cup \mathcal{A}_{e'})$ s.t. $t_a \neq t_b$ by contradiction. Suppose that $\forall a, b \in \mathcal{C}_h \cap (\mathcal{A}_e \cup \mathcal{A}_{e'})$ we have $t_a = t_b$. Using the same notation as above, denote a, b as the direct children of h in the path $e \to \cdots a \to h \to b \to \cdots \to e'$ which makes h a turning node. W.l.o.g. we consider two cases: $t_a = t_b = e$, and $t_a = t_b \in Q_{e,e'}$. For the first case, $t_b = e \Rightarrow e$ is a descendant of b, and from the definition of b we know that e' is a descendant of b, so $b \in \mathcal{A}_{e,e'}$. From Lemma 2, $t_b \in Q_{e,e'}$, contradicts $t_b = e$.

For the second case $t_a = t_b \in Q_{e,e'}$, denote $t_{a,b} = t_a = t_b$. As h is a turning node, there exists a path $e \to \cdots a \to h \to b \to \cdots \to e'$. Then the subpaths $e \to \cdots \to a$ and $b \to \cdots \to e'$ has no common nodes according to the definition of a path. So at least one of the subpaths does not include $t_{a,b}$, w.l.o.g assume subpath $b \to \cdots \to e'$ does not include $t_{a,b}$. As $t_{a,b}$ is a descendant of b, there exists paths $b \to \cdots \to t_{a,b}$, and we pick up the shortest one. We'll prove that there's no intersection between path $b \to \cdots \to t_{a,b}$ and path $b \to \cdots \to e'$: Assume that there exists such intersections, and denote the topologically lowest intersection (w.r.t. path $b \to \cdots \to t_{a,b}$) as x, then as we've assumed that subpath $b \to \cdots \to e'$ does not include $t_{a,b}$, we have $x \neq t_{a,b}$. Then we can prove that x is a turning node: If subpath $x \to \cdots \to e'$ does not intersect with path $e \to \cdots \to t_{a,b}$, then we can construct a path $e \to \cdots \to t_{a,b} \to x \to \cdots \to e'$, yielding x as a turning node. Otherwise, if $x \to \cdots \to e'$ intersects with $t_{a,b} \to \cdots \to e'$ before it intersects with $e \to \cdots \to t_{a,b}$ or it does not intersect with $e \to \cdots \to t_{a,b}$ at all, then denote the intersection node as y, we have a valid path $e \to \cdots \to t_{a,b} \to x \to \cdots \to y \to \cdots \to e'$ in which the part $y \to \cdots \to e'$ is a subpath of $t_{a,b} \to \cdots \to e'$, yielding x as a turning node. By similar construction, we can prove that if if $x \to \cdots \to e'$ intersects with $e \to \cdots \to t_{a,b}$ before it intersects with $t_{a,b} \to \cdots \to e'$ or it does not intersect with $t_{a,b} \to \cdots \to e'$ at all, x is also a turning node. However, x is nearer to b than $t_{a,b}$, which contradicts the definition of $t_{a,b}$. So we have proved that there's no intersection between path $b \to \cdots \to t_{a,b}$ and path $b \to \cdots \to e'$. Then we can prove that $t_{a,b} = b$: If path $b \to \cdots \to e'$ does not intersect with $e \to \cdots \to t_{a,b}$, then a valid path $e \to \cdots \to t_{a,b} \to \cdots \to b \to \cdots \to e'$ will make b a turning node, so $t_{a,b} = b$. Otherwise, if $b \to \cdots \to e'$ intersects with $t_{a,b} \to \cdots \to e'$ at z before it intersects with $e \to \cdots \to t_{a,b}$, then a valid path $e \rightarrow \cdots \rightarrow t_{a,b} \rightarrow \cdots \rightarrow b \rightarrow \cdots \rightarrow z \rightarrow \cdots \rightarrow e'$ where the part $z \rightarrow \cdots \rightarrow e'$ is subpath of $t_{a,b} \rightarrow e'$ will make b a turning node. If $b \rightarrow \cdots \rightarrow e'$ intersects with $e \rightarrow \cdots \rightarrow t_{a,b}$ at z before it intersects with $t_{a,b} \rightarrow \cdots \rightarrow e'$, then through similar construction we can also prove $t_{a,b} = b$. This contradicts the assumption that subpath $b \rightarrow \cdots \rightarrow e'$ does not include $t_{a,b}$, so the second case is also impossible. $\square$