

ELTRA: An Embedding Method based on Learning-to-Rank to Preserve Asymmetric Information in Directed Graphs

Submitted to the 32th ACM International Conference on Information and
Knowledge Management, ACM CIKM 2023

APPENDIX

APPENDIX A

In this section, we prove that the CRW scores are asymmetric, bounded, monotonic, unique, and always existent.

(1) **Asymmetry**: according to Equation (2), if $u \neq v$, $S_k(u, v)$ is computed by considering O_u and I_u , while $S_k(v, u)$ is computed by considering O_v and I_v ; since $O_u \neq O_v$ and $I_u \neq I_v$, then $S(u, v) \neq S(v, u)$

(2) **Bounding**: for all k , $0 \leq S_k(u, v) \leq 1$.

According to Equation (2), if $u \neq v$, then $S_0(u, v) = 0$, otherwise $S_0(u, v) = 1$; therefore $0 \leq S_0(u, v) \leq 1$. It means the property holds for $k=0$. Now, we assume that the property holds for k , which means $0 \leq S_k(u, v) \leq 1$ for *any* node-pairs (u, v) ; according to the assumption $S_k(u, v) \geq 0$, thus

$$\begin{aligned} S_{k+1}(u, v) &= \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} S_k(i, v)}{|O_u|} + \frac{\sum_{j \in I_u} S_k(j, v)}{|I_u|} \right) \\ &\geq \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} (0)}{|O_u|} + \frac{\sum_{j \in I_u} (0)}{|I_u|} \right) \\ &\geq 0 \end{aligned}$$

also, according to the assumption $S_k(u, v) \leq 1$, thus

$$\begin{aligned}
 S_{k+1}(u, v) &= \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} S_k(i, v)}{|O_u|} + \frac{\sum_{j \in I_u} S_k(j, v)}{|I_u|} \right) \\
 &\leq \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} (1)}{|O_u|} + \frac{\sum_{j \in I_u} (1)}{|I_u|} \right) \\
 &\leq \frac{C}{2} \cdot (1 + 1) \\
 &\leq C
 \end{aligned}$$

since $0 < C < 1$, then $S_{k+1}(u, v) \leq 1$.

(3) **Monotonicity:** for every node-pair (u, v) , the sequence $\{S_k(u, v)\}$ is non-decreasing as k increases.

If $u = v$, $S_0(u, v) = S_1(u, v) = \dots = 1$; thus, the property holds.

If $u \neq v$, according to Equation (2), $S_0(u, v) = 0$ and by the bounding property, $0 \leq S_1(u, v) \leq 1$; therefore, $S_0(u, v) \leq S_1(u, v)$, which means the property holds for $k = 0$. We assume that the property holds for all k where $S_{k-1}(u, v) \leq S_k(u, v)$ for *any* node-pairs (u, v) , which means $S_k(u, v) - S_{k-1}(u, v) \geq 0$. Now, we show the property holds for $k + 1$ as follows:

$$\begin{aligned}
 S_{k+1}(u, v) - S_k(u, v) &= \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} S_k(i, v)}{|O_u|} + \frac{\sum_{j \in I_u} S_k(j, v)}{|I_u|} \right) - \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} S_{k-1}(i, v)}{|O_u|} + \frac{\sum_{j \in I_u} S_{k-1}(j, v)}{|I_u|} \right) \\
 &= \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} S_k(i, v) - S_{k-1}(i, v)}{|O_u|} + \frac{\sum_{j \in I_u} S_k(j, v) - S_{k-1}(j, v)}{|I_u|} \right)
 \end{aligned}$$

according to the assumptions, $S_k(u, v) - S_{k-1}(u, v) \geq 0$ and we already know that $C > 0$ therefore, $S_{k+1}(u, v) - S_k(u, v) \geq 0$, which means $S_{k+1}(u, v) \geq S_k(u, v)$.

(4) **Existence:** the fixed points $S(*, *)$ of the CRW equation always exists.

By the bounding and monotonicity properties, for any node-pairs (u, v) , $S_k(u, v)$ is bounded and non-decreasing as k increases. A sequence $S_k(u, v)$ converges to $\lim S(u, v) \in [0, 1]$, according to the Completeness Axiom of calculus. $\lim_{k \rightarrow \infty} S_{k+1}(u, v) = \lim_{k \rightarrow \infty} S_k(u, v) = S(u, v)$

and the limit of a sum is identical to the sum of the limits, therefore

$$\begin{aligned}
S(u, v) &= \lim_{k \rightarrow \infty} S_{k+1} = \frac{C}{2} \cdot \left(\frac{\lim_{k \rightarrow \infty} \sum_{i \in O_u} S_k(i, v)}{|O_u|} + \frac{\lim_{k \rightarrow \infty} \sum_{j \in I_u} S_k(j, v)}{|I_u|} \right) \\
&= \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} \lim_{k \rightarrow \infty} S_k(i, v)}{|O_u|} + \frac{\sum_{j \in I_u} \lim_{k \rightarrow \infty} S_k(j, v)}{|I_u|} \right) \\
&= \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} S(i, v)}{|O_u|} + \frac{\sum_{j \in I_u} S(j, v)}{|I_u|} \right) \\
&= S(u, v)
\end{aligned}$$

(5) **Uniqueness:** the solution for the fixed-point $S(*, *)$ is always unique.

Suppose that $S(*, *)$ and $S'(*, *)$ are two solutions for the CRW equation. Also, for *all* node-pairs (u, v) , let $\delta(u, v) = S(u, v) - S'(u, v)$ be the difference between these two solutions. Let $M = \max_{(u, v)} |\delta(u, v)|$ be the maximum absolute value of all differences observed for some nod-pairs (u, v) (i.e., $|\delta(u, v)| = M$). We need to prove that $M = 0$. If $u = v$, $M = 0$ since $S(u, v) = S'(u, v) = 1$. If $u \neq v$, $S(u, v)$ and $S'(u, v)$ are computed by CRW, and we have

$$\begin{aligned}
\delta(u, v) &= S(u, v) - S'(u, v) \\
&= \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} S(i, v) - S'(i, v)}{|O_u|} + \frac{\sum_{j \in I_u} S(j, v) - S'(j, v)}{|I_u|} \right) \\
&= \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} \delta(i, v)}{|O_u|} + \frac{\sum_{j \in I_u} \delta(j, v)}{|I_u|} \right)
\end{aligned}$$

thus,

$$\begin{aligned}
M = |\delta(u, v)| &= \left| \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} \delta(i, v)}{|O_u|} + \frac{\sum_{j \in I_u} \delta(j, v)}{|I_u|} \right) \right| \\
&\leq \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} |\delta(i, v)|}{|O_u|} + \frac{\sum_{j \in I_u} |\delta(j, v)|}{|I_u|} \right) \\
&\leq \frac{C}{2} \cdot \left(\frac{\sum_{i \in O_u} M}{|O_u|} + \frac{\sum_{j \in I_u} M}{|I_u|} \right) \\
&= \frac{C}{2} \cdot (M + M) = C \cdot M
\end{aligned}$$

Since $0 < C < 1$, surely $M = 0$.