AdaSim: A Recursive Similarity Measure in Graphs

Accepted in the 30th ACM International Conference on Information and Knowledge Management, ACM CIKM 2021

APPENDIX

APPENDIX A

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

(1) **Symmetry**: according to Equation (10), if $a = b$, $S_k(a, b) = S_k(b, a) = 1$; if $a \neq b$, $S_k(a, b) = S_k(b, a) = \widehat{S}(a, b)$ for all $k \geq 0$.

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

According to Equation (10),

If $a = b$, then $S_0(a, b) = S_1(a, b) = \cdots = 1$; therefore, $0 \leq S_k(a, b) \leq 1$ for all values of k.

If $a \neq b$ and $I_a = \emptyset$ or $I_b = \emptyset$, then $S_0(a, b) = S_1(a, b) = \cdots = 0$; therefore, $0 \leq S_k(a, b) \leq 1$ for all values of k.

If $a \neq b$, $I_a \neq \emptyset$, and $I_b \neq \emptyset$, then $S_0(a, b) = \widehat{S}_0(a, b) = 0$; therefore, $0 \leq S_0(a, b) \leq 1$, which means the property holds for $k = 0$. Now, we assume that the property holds for k, which means $0 \leq S_k(a, b) = \widehat{S}_k(a, b) \leq 1$, According to the assumption $\widehat{S}_k(a, b) \geq 0$, thus

$$
\begin{aligned}
\widehat{S}_{k+1}(a, b) &= C \cdot \left(\frac{\alpha}{m} \sum_{i \in I_a \cap I_b} w_i + (1 - \alpha) \left(\frac{1}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot \widehat{S}_k(i, j) \cdot w_j \right) \right) \\
&\geq C \cdot \left(\frac{\alpha}{m} \sum_{i \in I_a \cap I_b} w_i + (1 - \alpha) \left(\frac{1}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot 0 \cdot w_j \right) \right) \\
&\geq C \cdot \frac{\alpha}{m} \sum_{i \in I_a \cap I_b} w_i
\end{aligned}
$$

m is the maximum Ada score in the dataset, thereby leading to the fact that $0 \leq \frac{1}{m} \sum_{i \in I_a \cap I_b} w_i \leq 1$; also, $0 < C < 1$ and $0 < \alpha \leq 1$, which means $\widehat{S}_{k+1}(a,b) = S_{k+1}(a,b) \geq 0$.
According to the assumption $\widehat{S}_k(a,b) \leq 1$, thus

$$\begin{aligned}\widehat{S}_{k+1}(a,b) &= C \cdot \left(\frac{\alpha}{m} \sum_{i \in I_a \cap I_b} w_i + (1-\alpha) \left(\frac{1}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot \widehat{S}_k(i,j) \cdot w_j \right) \right) \\ &\leq C \cdot \left(\frac{\alpha}{m} \sum_{i \in I_a \cap I_b} w_i + (1-\alpha) \left(\frac{1}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot 1 \cdot w_j \right) \right)\end{aligned}$$

since $\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t = \sum_{r \in I_a} \sum_{t \in I_b} w_r \cdot w_t = \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot 1 \cdot w_j$, then $\widehat{S}_{k+1}(a,b) \leq C \cdot \left(\frac{\alpha}{m} \sum_{i \in I_a \cap I_b} w_i + (1-\alpha) \cdot 1 \right)$; we also know that $\frac{1}{m} \sum_{i \in I_a \cap I_b} w_i \leq 1$, which means $\widehat{S}_{k+1}(a,b) \leq C \cdot \alpha + C \cdot (1-\alpha) = C$; since $0 < C < 1$, then also $\widehat{S}_{k+1}(a,b) = S_{k+1}(a,b) \leq 1$.

(3) **Monotonicity**: for every node-pair (a,b), the sequence $\{S_k(a,b)\}$ is non-decreasing as k increases.
If $a = b$, $S_0(a,b) = S_1(a,b) = \cdots = 1$; thus, the property holds.
If $a \neq b$ and $I_a = \emptyset$ or $I_b = \emptyset$, $S_0(a,b) = S_1(a,b) = \cdots = 0$; thus, the property holds.
If $a \neq b$, $I_a \neq \emptyset$, and $I_b \neq \emptyset$, according to Equation (10), $S_0(a,b) = 0$ and by the bounding property, $0 \leq S_1(a,b) \leq 1$; therefore, $S_0(a,b) \leq S_1(a,b)$, which means for $k = 0$, the property holds. We assume that the property holds for all k where $S_{k-1}(a,b) = \widehat{S}_{k-1}(a,b) \leq S_k(a,b) = \widehat{S}_k(a,b)$, which means $\widehat{S}_k(a,b) - \widehat{S}_{k-1}(a,b) \geq 0$. Now, we show the property holds for $k+1$ as follows:

$$
\begin{aligned}
\widehat{S}_{k+1}(a,b) - \widehat{S}_k(a,b) &= C \cdot \left(\frac{\alpha}{m} \sum_{i \in I_a \cap I_b} w_i + (1-\alpha) \left(\frac{1}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot \widehat{S}_k(i,j) \cdot w_j \right) \right) \\
&- C \cdot \left(\frac{\alpha}{m} \sum_{i \in I_a \cap I_b} w_i + (1-\alpha) \left(\frac{1}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot \widehat{S}_{k-1}(i,j) \cdot w_j \right) \right) \\
&= \frac{C \cdot (1-\alpha)}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \left(\sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot \widehat{S}_k(i,j) \cdot w_j - \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot \widehat{S}_{k-1}(i,j) \cdot w_j \right) \\
&= \frac{C \cdot (1-\alpha)}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \left(\sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot w_j \cdot \left(\widehat{S}_k(i,j) - \widehat{S}_{k-1}(i,j) \right) \right)
\end{aligned}
$$

according to the assumptions, $\widehat{S}_k(a,b) - \widehat{S}_{k-1}(a,b) \geq 0$ and we already know that $C \cdot (1-\alpha) \geq 0$ and $\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t \geq 0$; therefore, $\widehat{S}_{k+1}(a,b) - \widehat{S}_k(a,b) \geq 0$, which means $\widehat{S}_{k+1}(a,b) = S_{k+1}(a,b) \geq \widehat{S}_k(a,b) = S_k(a,b)$.

(4) **Existence**: the fixed points $S(*,*)$ of the AdaSim equation always exists.
By the bounding and monotonicity properties, for any node-pairs (a,b), $\widehat{S}_k(a,b)$ is bounded and non-decreasing as k increases. A sequence $\widehat{S}_k(a,b)$ converges to a $\lim \widehat{S}(a,b) = S(a,b)$ in $[0,1]$, according to the Completeness Axiom of calculus. $\lim\limits_{k \to \infty} \widehat{S}_{k+1}(a,b) = \lim\limits_{k \to \infty} \widehat{S}_k(a,b) = \widehat{S}(a,b)$ and the limit of a sum is identical to the sum of the limits, therefore

$$
\begin{aligned}
\widehat{S}(a,b) = \lim_{k \to \infty} \widehat{S}_{k+1} &= C \cdot \left(\frac{\alpha}{m} \sum_{i \in I_a \cap I_b} w_i + \frac{(1-\alpha)}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \lim_{k \to \infty} \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot \widehat{S}_k(i,j) \cdot w_j \right) \\
&= C \cdot \left(\frac{\alpha}{m} \sum_{i \in I_a \cap I_b} w_i + \frac{(1-\alpha)}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot w_j \cdot \lim_{k \to \infty} \widehat{S}_k(i,j) \right) \\
&= C \cdot \left(\frac{\alpha}{m} \sum_{i \in I_a \cap I_b} w_i + \frac{(1-\alpha)}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot w_j \cdot \widehat{S}(i,j) \right) \\
&= S(a,b)
\end{aligned}
$$

(5) **Uniqueness**: the solution for the fixed-point $S(*,*)$ is always unique.
Suppose that $S(*,*)$ and $S'(*,*)$ are two solutions for the AdaSim equation. Also, for all node-pairs (a,b), let $\delta(a,b) = S(a,b) - S'(a,b)$ be the difference between these two solutions. Let $M = \max_{(a,b)} |\delta(a,b)|$ be the maximum absolute value of all differences observed for some nod-pairs (a,b) (i.e., $|\delta(a,b)| = M$). We need to prove that $M = 0$. If $a = b$, $M = 0$ since $S(a,b) = S'(a,b) = 1$. If $a \neq b$ and $I_a = \emptyset$ or $I_b = \emptyset$, $M = 0$ since $S(a,b) = S'(a,b) = 0$. Otherwise, $S(a,b) = \widehat{S}(a,b)$ and $S'(a,b) = \widehat{S}'(a,b)$ are computed by AdaSim.
When $\alpha = 1$, $M = 0$ since $\widehat{S}(a,b) = \widehat{S}'(a,b) = \frac{C}{m} \sum_{i \in I_a \cap I_b} w_i$.
When $0 < \alpha < 1$, we have

$$\begin{aligned}
\delta(a,b) &= \widehat{S}(a,b) - \widehat{S}'(a,b) \\
&= \frac{C \cdot (1-\alpha)}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot w_j \cdot (\widehat{S}(i,j) - \widehat{S}'(i,j)) \\
&= \frac{C \cdot (1-\alpha)}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot w_j \cdot \delta(i,j)
\end{aligned}$$

thus,

$$\begin{aligned}
M = |\delta(a,b)| &= \left| \frac{C \cdot (1-\alpha)}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot w_j \cdot \delta(i,j) \right| \\
&\leq \frac{C \cdot (1-\alpha)}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot w_j \cdot |\delta(i,j)| \\
&\leq \frac{C \cdot (1-\alpha)}{\sum_{r \in I_a} w_r \cdot \sum_{t \in I_b} w_t} \cdot \sum_{i \in I_a} \sum_{j \in I_b} w_i \cdot w_j \cdot M \\
&= C \cdot (1-\alpha) \cdot M
\end{aligned}$$

Since $0 < \alpha < 1$ and $0 < C < 1$, then $0 < C \cdot (1-\alpha) < 1$, which means $M = 0$.