Theorem 2: *Incorporating POI contextual information into node representations increases the entropy of the node embedding.* Specifically,

$$H(\tilde{G}) \geq H(G), \tag{1}$$

where $H(\tilde{G})$ represents the entropy under multiple features and $H(G)$ represents the entropy under a single feature.

Proof

The entropy H of a graph's node embedding distribution is defined as:

$$H(G) = -\sum_{x \in \mathcal{H}} P(x) \log P(x), \tag{2}$$

where $P(x)$ is the probability of a node embedding being assigned to state x, and $\mathcal{H}$ represents the set of embedding states.

Assuming the feature space size is k. In a single-feature graph G, similar as (2) in Figure 2 (d), the embedding space are distributed uniformly:

$$P(x) = \frac{1}{k}, \quad x \in \mathcal{H}. \tag{3}$$

The entropy of G is:

$$H(G) = -\sum_{x=1}^{k} \frac{1}{k} \log \frac{1}{k} = \log k. \tag{4}$$

With multiple features, let the feature space grow exponentially with dimension d, i.e., $k_d = k^d$. The embedding states are uniformly distributed:

$$P(x) = \frac{1}{k^d}, \quad x \in \mathcal{H}. \tag{5}$$

The entropy of $\tilde{G}$ is:

$$H(\tilde{G}) = -\sum_{x=1}^{k^d} \frac{1}{k^d} \log \frac{1}{k^d} = \log k^d = d \log k. \tag{6}$$

Then we compare the entropies of $H(G)$ and $H(\tilde{G})$:

$$H(\tilde{G}) = d \log k \quad \text{and} \quad H(G) = \log k. \tag{7}$$

Since $d > 1$, it follows that:

$$H(\tilde{G}) \geq H(G). \tag{8}$$

Therefore, incorporating multiple features increases the entropy of the node embeddings.

Instead of exponential growth, let the feature space expand linearly with dimension d, such that $k_d = k \cdot d$. The embedding states are uniformly distributed:

$$P(x) = \frac{1}{k \cdot d}, \quad x \in \mathcal{H}. \tag{9}$$

The entropy of $\tilde{G}$ is:

$$H(\tilde{G}) = -\sum_{x=1}^{k \cdot d} \frac{1}{k \cdot d} \log \frac{1}{k \cdot d} = \log(k \cdot d). \tag{10}$$

Then we compare the entropies of $H(G)$ and $H(\tilde{G})$:

$$H(\tilde{G}) = \log(k \cdot d) \quad \text{and} \quad H(G) = \log k. \tag{11}$$

Expanding $H(\tilde{G})$:

$$H(\tilde{G}) = \log k + \log d. \tag{12}$$

Since $\log d > 0$ for $d > 1$, it follows that:

$$H(\tilde{G}) \geq H(G). \tag{13}$$

Thus, even with linear growth of features, incorporating multiple features increases the entropy of the node embeddings.