Theoretical Proof

Proposition 1. *Let M_k represent the number of vertexes at the k-th level of the call graph. We denote $P_{gen\_width}(M_k = m_k)$ as the probability that the k-th level of the call graph has m_k vertexes according to a random graph model, and $P_{real\_width}(M_k = m_k)$ as the probability that the k-th level of the real call graph has m_k vertexes. Then,*

$$P_{gen\_width}(M_k = m_k) = P_{real\_width}(M_k = m_k) \tag{1}$$

Proof. Let E_S denote the event that service S is selected. Then, $P_{\text{gen\_width}}(M_k = m_k \mid E_S) = P_{\text{real\_width}}(M_k = m_k \mid E_S)$. This is formalized using induction.

When $k = 2$, the layer 1 consists solely of the vertex 'user', the probability of m_2 vertexes at layer 2 in the generated graph equals the probability of the size of the children set of 'user' equals m_2. Thus, the probability of generation equals that of S.Assume the hypothesis holds for k. For $k + 1$:

$$P_{\text{gen\_width}}(M_{k+1} = m_{k+1} \mid E_S) = \left|\bigcup_{u \in V_k} C_u\right| \tag{2}$$

Based on the process of generating call graphs from a random graph model, we can establish that the size distributions of children sets for each vertex identical. Therefore, the probability of a generated call graph having m_k vertexes in k-th layer is equal to the probability of service S's.

$$\begin{aligned} P_{\text{gen\_width}}(M_k = m_k) &= \sum_{\text{all } S} P(E_S) \cdot P_{\text{gen\_width}}(M_k = m_k \mid E_S) \\ &= \sum_{\text{all } S} P(E_S) \cdot P_{\text{real\_width}}(M_k = m_k \mid E_S) \end{aligned} \tag{3}$$

$$P_{\text{real\_width}}(M_k = m_k) = \sum_{\text{all } S} P(E_S) \cdot P_{\text{real\_width}}(M_k = m_k \mid E_S) \tag{4}$$

$$P_{\text{gen\_width}}(M_k = m_k) = P_{\text{real\_width}}(M_k = m_k) \tag{5}$$

□

Proposition 2. *Let H denote the total depth of a call graph. Let $P_{gen\_depth}(H = h)$ denote the probability that a generated call graph has depth h, and $P_{real\_depth}(H = h)$ denote the probability of depth h in the real dataset. Then,*

$$P_{gen\_depth}(H = h) = P_{real\_depth}(H = h) \tag{6}$$

Proof. We proceed with the following steps using conditional probabilities:

Let E_S denote the event that service S is selected. We aim to show:

$$P_{\text{gen\_depth}}(H = h \mid E_S) = P_{\text{real\_depth}}(H = h \mid E_S) \tag{7}$$

Under the assumption that service S is selected (E_S), the probability $P_{\text{gen\_depth}}(H = h \mid E_S)$ represents the likelihood that a generated call graph has depth h. This can be expressed using the conditions: 1. The depth of the generated call graph must be at least h, given by $1 - \sum_{i=1}^{h-1} P_{\text{real\_depth}}(H = i)$. 2. At depth h, every vertex $u \in V_h$ must have an empty children set, denoted $P(|C_u| = 0) \mid E_S$.

Thus,

$$P_{\text{gen\_depth}}(H = h \mid E_S) = \left(1 - \sum_{i=1}^{h-1} P_{\text{real\_depth}}(H = i)\right) \cdot \prod_{u \in V_h} P(|C_u| = 0 \mid E_S) \tag{8}$$

Based on the process of generating call graphs from a random graph model, we can establish that the size distributions of children sets for each vertex and the distributions of vertex counts at each depth are identical. Therefore, the probability of a generated call graph having depth h is equal to the probability of service S's call graph having depth h.

Next, to prove $P_{\text{gen\_depth}}(H = h) = P_{\text{real\_depth}}(H = h)$:

$$\begin{aligned} P_{\text{gen\_depth}}(H = h) &= \sum_{\text{all } S} P(E_S) \cdot P_{\text{gen\_depth}}(H = h \mid E_S) \\ &= \sum_{\text{all } S} P(E_S) \cdot P_{\text{real\_depth}}(H = h \mid E_S) \end{aligned} \tag{9}$$

$$P_{\text{real\_depth}}(H = h) = \sum_{\text{all } S} P(E_S) \cdot P_{\text{real\_depth}}(H = h \mid E_S) \tag{10}$$

$$P_{\text{gen\_depth}}(H = h) = P_{\text{real\_depth}}(H = h) \tag{11}$$

Therefore, $P_{\text{gen\_depth}}(h) = P_{\text{real\_depth}}(h)$, demonstrating that the probability of a generated call graph having depth h equals the probability of service S's call graph having depth h for any selected service S. $\square$

Proposition 3. *Let N denote the total number of vertexes in a call graph, $P_{gen\_num}(N = n)$ denote the probability that the generated call graph has n microservices. And $P_{real\_num}(N = n)$ denote the probability that the real call graph has n microservices.Then:*

$$P_{gen\_num}(N = n) = P_{real\_num}(N = n) \tag{12}$$

Proof. $P_{\text{gen\_num}}(N = n)$ can be expressed mathematically as:

$$P_{\text{gen\_num}}(N = n) = \sum_{h=2}^{h_{max}} P_{\text{gen\_depth}}(H = h) \cdot P_{\text{gen\_width}}\Big(\sum_{i=2}^{h} m_i = n\Big) \tag{13}$$

Here, $P_{\text{gen\_depth}}(H = h)$ is the probability that the generated call graph has depth h,m_i is the num of vertex counts at depth i in the generated call graph.We have previously established that the distribution of vertex counts at each layer and the overall number of layers in the generated call graph are equivalent to those in the real call graph. Therefore, we conclude that $P_{\text{gen\_num}}(N = n) =$

$P_{\text{real\_num}}(N = n)$, indicating that the probability distribution of the number of microservices in the generated call graph is identical to that in the real call graph. $\square$