subgraph2vec

Article	Data	Data	Comments or Questions
Narayanan A , Chandramohan M , Chen L , et al. subgraph2vec: Learning Distributed Representations of Rooted Sub-graphs from Large Graphs[J]. 2016.	**Aim:** Learning latent representations of rooted subgraphs from large graphs. **Contributions:** ● Propose subgraph2vec. ● Develop a modified version of the skipgram language model. ● Discuss how subgraph2vec's representation learning technique would help to build the deep learning variant of WL kernel. **Applications:** ● graph classification and clustering tasks ● code clone detection ● malware detection	**Key results:** ● Benchmark Datasets Table 3: Average Accuracy (± std dev.) for subgraph2vec and state-of-the-art graph kernels on benchmark graph classification datasets Dataset | MUTAG | PTC | PROTEINS | NCI1 | NCI109 WL [6] | 80.63 ± 3.07 | 56.91 ± 2.79 | 72.92 ± 0.56 | 80.01 ± 0.50 | 80.12 ± 0.34 Deep WL_{YV} [7] | 82.95 ± 1.96 | 59.04 ± 1.09 | 73.30 ± 0.82 | **80.31** ± 0.46 | **80.32** ± 0.33 subgraph2vec | **87.17** ± 1.72 | **60.11** ± 1.21 | **73.38** ± 1.09 | 78.05 ± 1.15 | 78.39 ± 1.89 ● Clone Detection Table 5: Clone Detection - Results Kernel | WL [6] | Deep WL_{YV} [7] | subgraph2vec Pre-training duration | - | 421.7 s | 409.28 s ARI | 0.67 | 0.71 | **0.88** ● Malware Detection Table 7: Malware Detection - Results Classifier | WL [6] | Deep WL_{YV} [7] | subgraph2vec Pre-training duration | - | 2631.17 s | 2219.28 s Accuracy | 66.15 | 71.03 | **74.48** **Background:** ● language model **Methods:** ● Generate rooted subgraphs: WL Relabeling Process ● Embeddings of those subgraphs: Radial Skip Gram Model ● Approximate the probability distribution: Negative Sampling ➢ **for** *each* $sg_{cont} \in context_v^{(d)}$ **do** $J(\Phi) = -\log \Pr(sg_{cont} \mid \Phi(sg_v^{(d)}))$	● When subgraph of degree 0 are considered, subgraph2vec provides node embeddings. ● How can we get the initial labels for each node? ● What is the Deep WL kernel? ☑