Evaluating node embeddings for low degree nodes

The figure below shows the improvement in classification accuracy when considering only low degree nodes. Sorting the nodes in increasing order according to their degree, we compute the macro-F1 scores for node classification using DeepWalk and SmoothDeepWalk embeddings only for the $k\%$ of all nodes having the lowest degrees for $k \in [20, 40, 60, 80, 100]$. Thus, the lower k the lower the degree of the nodes, and for $k = 100$ we consider all nodes. We report the improvement over DeepWalk in percentage. For example, if DeepWalk yields a Macro-F1 score of 0.8 and SmoothDeepWalk yields 0.82, the improvement is 2.5%. As we see, for all graphs the gains become more significant for a lower degree, except for Flickr which is the graph with the lowest skew in positive pair frequency distribution.

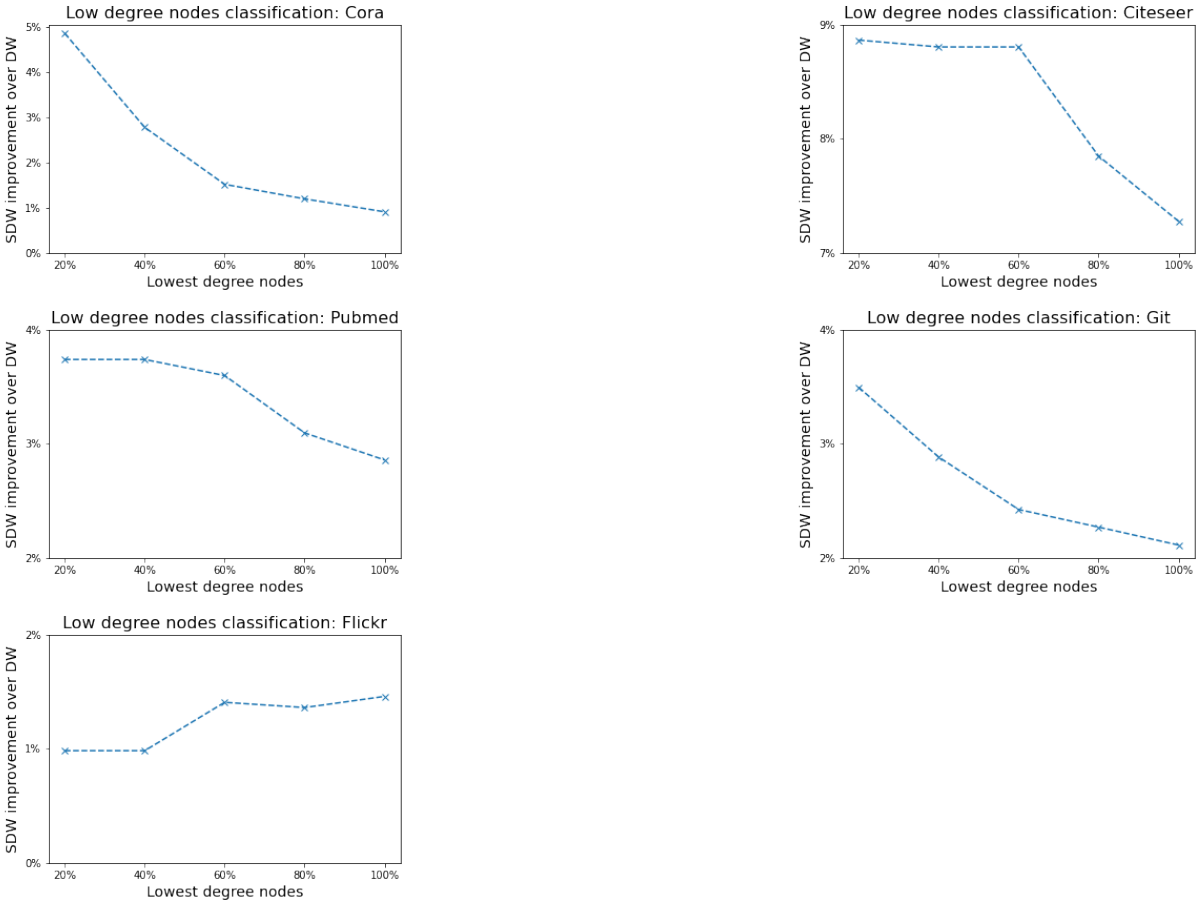

Figure 1: The improvement of macro-F1 node classification scores of SmoothDeepWalk over DeepWalk for varying percentage lowest degree nodes