

756 APPENDIX

757
758 A SIMILARITY MEASURES

759
760 When comparing two binary arrays representing node features with n binary attributes, we compute
761 four quantities, often referred to as frequencies, from the given binary data:
762

- 763 • a – The number of attributes that are equal to 1 for both objects i and j .
- 764 • b – The number of attributes that are equal to 0 for object i but equal to 1 for object j .
- 765 • c – The number of attributes that are equal to 1 for object i but equal to 0 for object j .

766
767 Based on these frequencies, several popular similarity measures are defined as follows:
768

- 769 • **Jaccard Similarity** is defined as $\mathcal{J}(i, j) = \frac{a}{a + b + c}$
770
- 771 • **Positive Similarity** is defined as $\mathcal{P}(i, j) = a$
772
- 773 • **Cosine Similarity** is defined as $\mathcal{C}(i, j) = \frac{a}{\sqrt{(a + b)(a + c)}}$
774

775 Table 5 displays the performance of the W-GCN model across various similarity measures. Following
776 an assessment of these measures, we have chosen *positive similarity* for this study, particularly suited
777 for datasets containing binary node feature vectors.
778

779 B CLASSWISE LANDMARKS FOR EACH DATASET

780
781 Our landmark selection methods are highly flexible and can be applied to various datasets with
782 different attribute vector formats. In this section, we provide details on how these methods were
783 implemented in the benchmark datasets.
784

785 **CORA:** The Cora dataset is a widely used benchmark in the field of machine learning, particularly
786 for tasks involving graph neural networks and semi-supervised learning. It consists of a citation
787 network of scientific publications classified into one of seven classes. The dataset contains 2,708
788 nodes, each representing a document, and 5,429 edges, each representing a citation link between two
789 documents. Each document is described by a 1,433-dimensional binary word vector, indicating the
790 absence or presence of specific words from a predefined dictionary. Since the dataset has categorical
791 domain features, we consider two class-wise landmarks: one is inclusive and the other is selective.
792 The inclusive landmark, ξ_i^j , is a binary word vector of length 1433, where each entry is 1 if the
793 corresponding word is present in any binary vectors X_u of class C_j , and 0 otherwise. For the selective
794 landmark vector, ξ_s^j is defined as a binary word vector indicating the presence of the corresponding
795 word in at least 10% of nodes in class C_j . Thus, we have 7 class-wise landmarks for the inclusive
796 choice and another 7 class-wise landmarks for the selective choice, totaling 14 class-wise landmarks.
797 By considering the similarity measure with these landmarks for a given node, we can create a
14-dimensional Positional embedding for the TWGNN model.

798 **CITeseer:** It consists of a citation network of scientific publications classified into one of six classes.
799 The dataset contains 3,327 nodes, each representing a document, and 4,732 edges, each representing
800 a citation link between two documents. Each document is described by a 3,703-dimensional binary
801 word vector, indicating the absence or presence of specific words from a predefined dictionary. Since
802 the dataset shares similar characteristics with the Cora dataset, we employ the same technique for
803 classWISE landmarks selection. This involves creating six inclusive and six selective landmarks,
804 resulting in a total of twelve landmarks. These landmarks capture important information about the
805 presence of specific words within each class of nodes. Subsequently, we utilize these landmarks to
806 generate a twelve-dimensional positional embedding for our model.

807 **PUBMED:** The PUBMED dataset consists of a directed graph containing 19,717 scientific publi-
808 cations obtained from the PubMed database, specifically targeting diabetes research, and divided
809 into three classes. Each node within the graph represents a publication and is defined by a TF/IDF
weighted word vector derived from a vocabulary of 500 unique words. As the domain features are

810 real-valued, we select three landmarks by computing the mean value along each dimension of the
811 word vectors, resulting in three 500-dimensional landmarks. Utilizing the Euclidean distance as a
812 similarity measure, we construct a three-dimensional initial positional embedding for our model,
813 enabling the effective representation of the dataset’s structural and textual characteristics.

814 **TEXAS, CORNELL, and WISCONSIN:** The Texas, Cornell, and Wisconsin datasets all represent
815 web-page networks, where nodes correspond to web pages and edges signify hyperlinks between
816 them. The Wisconsin dataset comprises 251 nodes, while both the Cornell and Wisconsin datasets
817 contain 183 nodes each. Each node is characterized by a feature vector of 1,703 binary attributes,
818 denoting the presence or absence of specific words within the web page content. Additionally, the
819 nodes are classified into one of five categories based on the type of web page they represent. Given
820 that the feature vector is categorical, akin to the CORA dataset, we can employ a similar technique for
821 feature extraction. This involves creating five selective and five inclusive landmarks. Consequently,
822 we can generate a ten-dimensional positional embedding for our model.

823 **CHAMELEON:** The datasets represent page-page networks centered around specific topics such as
824 chameleons. In these networks, nodes correspond to articles, while edges signify mutual links between
825 them. Node features are extracted from informative nouns found in the corresponding Wikipedia
826 pages. The presence of a feature in the feature list indicates the occurrence of an informative noun in
827 the text of the Wikipedia article. These two datasets also share categorical features and possess feature
828 vectors of dimensions 2325 and 2089 respectively. Both datasets are structured into five classes.
829 Employing our categorical class-wise landmark selection technique, we establish five inclusive and
830 five selective landmarks with dimensions 2325 and 2089.

831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863