Supplementary Material: Structure-Aware Face Clustering on a Large-Scale Graph with 10^7 Nodes

1. Effect of Graph Parsing

Method	Pre	Recall	F_P
only graph refinement	91.84	70.87	80.00
only graph parsing	92.83	74.24	82.50
graph parsing+graph refinemnet	**95.50**	**85.91**	**90.45**

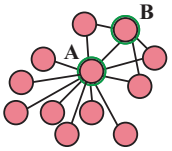
Table 1: Method comparison with different inference strategies. Train with the *part0_ train* and test on the *part1_ test* in MS1M [1].

In the proposed STAR-FC, we transform the face clustering task into two steps: **graph parsing** and **graph refinement**. The graph parsing step is based on a GCN edge confidence estimator, while the graph refinement step is based on the node intimacy (NI). In our paper, we have proved that these two steps are indispensable. Here we conduct more experiments to further demonstrate that clustering faces with a single step does not work well.

In Table 1 we compare the *pairwise F-score* under three different inference strategies. The *only graph refinement* method means that we directly perform pruning with NI on the *K*NN graph. In *only graph parsing* method, face clusters are obtained by dynamic edge pruning [8] based on the GCN predicted edge scores. As shown in Table 1, face clustering with single step achieves poor performance. We have analyzed the disadvantage of the *only graph parsing* method in the paper. Here we mainly discuss the poor performance of the *only graph refinement* method. We argue that since there exist lots of wrong connections in the initial *K*NN graph, different clusters are also densely linked and the NI of negetive neighbor may be very high. So the NI is difficult to work well on this graph. Therefore, performing graph parsing to get a relatively clear graph structure before employing the NI-based pruning is indispensable, and the combination of these two steps can achieve superior 90.45 *pairwise F-score*.

2. Comparing the NI with Jaccard similarity

Jaccard similarity coefficient is a statistic used for calculating the similarity and diversity of sample sets. Following the symbols defined in Sec.3.3, Jaccard [2, 4] $= \frac{k}{n_1+n_2-k}$,



Nodes	Edges	Shared Neighbors	NI	Jaccard
A	11	3	3/4	3/12
B	4			

Figure 1: An example to visualize the difference of Jaccard similarity coefficient and the proposed NI.

while NI $= max(\frac{k}{n_1}, \frac{k}{n_2})$. NI considers the attributes of two nodes respectively then uses max aggregation for judgment while Jaccard ignores the difference between the two nodes. As shown in Figure 1, *A* and *B* should be in the same cluster. Since *A* has many neighbors and *B* has a few, Jaccard gets a low score leading to misjudgment while NI can handle this case well. Therefore NI is more suitable for intimacy measures. In the added ablation study, the F_P of Jaccard and NI are 88.39 and 91.97, respectively.

3. More details on the SPSS

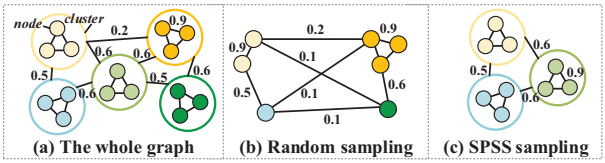


Figure 2: Comparison of random sampling method and the proposed SPSS sampling.

We use visualization to explain how SPSS works more vividly. Figure 2 illustrates the graph structure. (a) is the whole graph. (b) and (c) show the subgraph sampled in random way and SPSS respectively. Nodes with the same color belong to the same class. (c) preserves both the intra-cluster links and the **hard negative edges** between near clusters. **These two types of edges in (c) approximate the structure of edges in (a).** However the negative edges in (b) are mostly with low similarity that contribute less for training (Sec.3.2). Compared with the whole graph training (90.45 F_P), training with SPSS does not lead to performance loss and brings some extra accuracy gain (91.97 F_P), which further indicates that the global structure is preserved.

4. Discussion

In this section, we discuss core differences between some representative face clustering methods and the proposed STAR-FC.

L-GCN [5] predicts the linkage within some selected subgraphs. It relies heavily on a mass of subgraphs. Since there exist many overlapped neighbors in these subgraphs, it suffers from heavily redundant calculations which is a big drag on the inference speed. Besides, such local graph operations lack the comprehension of global graph structure which limits its performance upper bound. GCN-D [7] formulates face clustering as a detection and segmentation problem based on the affinity graph. However, it has the same problem as the L-GCN. It generates a large number of cluster proposals thus leading to inefficient inference. By contrast, the proposed STAR-FC performs face clustering inference based on the full-graph operation which can satisfy both efficiency and accuracy. ARO [3] computes the top-k nearest neighbors for each face in the dataset and performs face clustering based on the approximate rank-order metric. However, it lacks the parsing of the initial structure information. The coarse top-k nearest neighbors may contain a number of negative samples and the order of images may be far from the exact one, thus the effect of the rank-order metric will be greatly damaged. In the proposed STAR-FC, we use the edge scores predicted by the GCN to parse the graph in advance, therefore most wrong connections can be removed and the NI can work better on this graph with clearer structure. And the importance of graph parsing has been proved in Table 1. GCN-V+E [6] obtains face clusters through predicting the vertex confidence and edge connectivity. It takes the entire graph as input for GCN training. Due to the limitation of GPU memory, it is hard to handle larger-scale training set. The proposed STAR-FC proposes the structure-preserved subgraph sampling strategy to address this challenge and is able to explore larger-scale training data.

References

[1] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. MS-Celeb-1M: A dataset and benchmark for large-scale face recognition. In *ECCV*, 2016. 1

[2] Paul Jaccard. The distribution of the flora in the alpine zone. *New phytologist*, 1912. 1

[3] Charles Otto, Dayong Wang, and Anil K Jain. Clustering millions of faces by identity. *TPAMI*, 2017. 2

[4] Taffee T Tanimoto. Elementary mathematical theory of classification and prediction. 1958. 1

[5] Zhongdao Wang, Liang Zheng, Yali Li, and Shengjin Wang. Linkage based face clustering via graph convolution network. In *CVPR*, 2019. 2

[6] Lei Yang, Dapeng Chen, Xiaohang Zhan, Rui Zhao, Chen Change Loy, and Dahua Lin. Learning to cluster faces via confidence and connectivity estimation. In *CVPR*, 2020. 2

[7] Lei Yang, Xiaohang Zhan, Dapeng Chen, Junjie Yan, Chen Change Loy, and Dahua Lin. Learning to cluster faces on an affinity graph. In *CVPR*, 2019. 2

[8] Xiaohang Zhan, Ziwei Liu, Junjie Yan, Dahua Lin, and Chen Change Loy. Consensus-driven propagation in massive unlabeled data for face recognition. In *ECCV*, 2018. 1