A Supplementary Material

A.1 Order Matters

In this section, we are going to discuss the important factor that affects the choice of the order for MSNE. In fact, the order of MSNE which generates best results mainly depends on the average number of node pairs with specific distance in the network.

Table 1. Average number of node pairs with distance ranging from 1 to 5.

Dataset	**Distance**					**Best Order**
	1	2	3	4	5	
BlogCatalog	64.77	6292.01	3899.67	53.51	0.05	4
Cora	3.90	31.88	91.30	244.94	438.38	3
arXiv-AstroPh	19.80	431.95	3597.52	4699.43	1591.66	4
arXiv-GrQc	5.81	26.58	109.31	143.37	790.18	3
PPI	17.69	610.45	2107.50	975.27	129.87	4

First of all, the average number of immediate neighbor nodes in the network affects the choice of order. If the number is small, the sampling space when generating the node sequence would be small, which results in *no gain* to use higher-order MSNE. Table 1 shows the average number of node pairs with distance ranging from 1 to 5. As shown in the Table 1, the average number of immediate neighbors in Cora and arXiv-GrQc is 3.90 and 5.81. Accordingly, MSNE with 3-order is enough to get expressive node sequences.

Secondly, if the distance of most node pairs in the network is smaller than a certain value (order), there is *few* information to be utilized for higher-order MSNE. Taking BlogCatalog as an example, the distance of most of node pairs is smaller than 5 (the average number of node pairs with distance of 5 is just 0.05), and thus MSNE-5^{th} cannot bring significant improvements over MSNE-4^{th} (Micro-F1 score of MSNE-5^{th} is 0.435, which is close to that of MSNE-4^{th}, i.e., 0.437.).

Besides, due to the randomness of MSNE and the greedy-based parameter search method, the high-order model may be worse than the low-order model due to *the introduction of noise*. This may be why MSNE-3^{rd} outperforms MSNE-4^{th} on the datasets Cora and arXiv-GrQc. Moreover, since MSNE filters out the candidates sequentially during applying strategies $\pi_2(x, u_2), \cdots, \pi_n(x, u_n)$ according to Eq. 7 and therefore MSNE with low parameter p is more likely to select next step earlier according to Eq. 8. In that case, MSNE with higher order is prone to getting similar performance of lower-order MSNE. As shown in table 1, although the average number of node pairs with a distance of 5 is large (1591.66) in arXiv-AstroPh, the probability of selecting the next step when using only 4-order MSNE is 78% due to the best parameter setting ($\{p_2 = 0.25, q_2 = 4; p_3 = 0.5, q_3 = 1; p_4 = 1, q_4 = 1\}$). Hence, the performance of high order MSNE

is similar to that of low order MSNE in arXiv-AstroPh. We leave finding a better way of parameter searching as a future work.