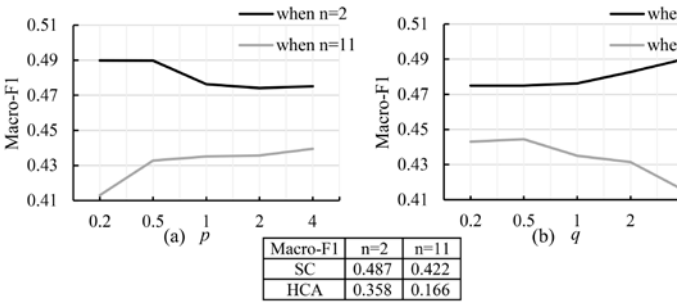

Macro-F1	n=2	n=11
SC	0.487	0.422
HCA	0.358	0.166

Figure 3. Changes of Macro-F1 versus hyperparameters *p* and *q*: when the search strategy is more inclined to BFS (p<1 or q >1), Macro-F1 scores of structural equivalence get higher and when the search strategy is more inclined to DFS (p>1 or q<1), Macro-F1 of homophily get higher. Additionally, with most settings of hyperparameters, PEM outperforms SC and HCA in both cases.

As mentioned in Section 3.1.2, we could deduce that adopting BFS helps discover the structural equivalence (when n=2) while using DFS helps find the homophily (when n=11) in this network. Figure 3 shows that when the search strategy is more inclined to BFS (*p*<1 or *q* >1), Macro-F1 scores of structural equivalence get higher and when the search strategy is more inclined to DFS (*p*>1 or *q*<1), Macro-F1 of homophily get higher. Additionally, with most settings of hyperparameters, PEM outperforms SC and HCA in both cases. Thus, it proves that PEM is flexible, which is able to discover different kinds of structure through tuning hyperparameters. And it is also accurate when the values of *p* and *q* is proper.

3.3 Parameter Sensitivity

PEM involves a set of parameters and we examine the recommendations about how to choose the parameters in Section 2.2. Here, the basic parameters are set as *d*=5, *l*=10, *r*=10, *k*=5 and the hyperparameters are set as *p*=4, *q*=1. The influences of parameters on the detection of homophily and structural equivalence in this network are shown as Figure 4.

As seen in Figure 4, the overall effect of parameters is positive: Macro-F1 scores are on the rise despite the fluctuations. It is also could be seen that the parameters have little effect on the results when n=2. As mentioned above, when *p*=4 and *q*=1, it's DFS guiding the random walk and we are likely to discover structure of homophily in this network. Thus, under the significant influence of hyperparameters, it is reasonable that Macro-F1 scores are not hardly affected by the parameters when n=2 while Macro-F1 scores rise obviously. Among the parameters, *k* has a significant influence on the clustering, and *d* as well as *r* follows. The influence of *l* is limited.

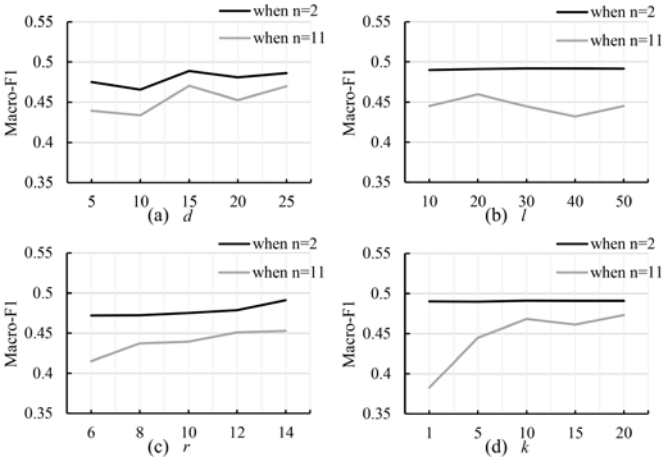

Figure 4. Parameter Sensitivity: the overall effect of parameters is positive: Macro-F1 scores are on the rise despite the fluctuations[1].

3.4 Perturbation Analysis

Since many real-world datasets co-exists with uncertain noise, we perform a perturbation study where we added noise data. The distribution of noise obeys Gaussian distribution $N(0,\sigma^2)$ ($\sigma \in (0.2,0.5,1,2,4)$) and Poisson distribution $P(\lambda)$ ($\lambda \in (1,2,4,8,16)$) respectively and the value of created noise matrices is processed into [0, 1]. The experimental results are shown as following. It could be seen that Macro-F1 scores change little whatever the distribution of noise is. Thus, PEM is proved to be able to handle the problem with high noise.

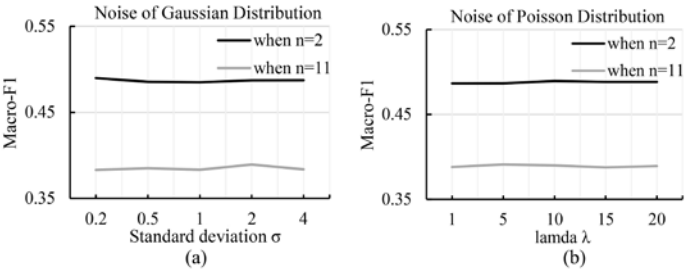

Figure 5. Perturbation analysis: The steady Macro-F1 results show that PEM could handle the problem with high noise

4. CASE STUDY: URBAN STRUCTURE IN SHANGHAI (CHINA) USING TAXI TRAJECTORY DATA

4.1 Experimental Setup

The studied area: We take the central area within the outer ring of Shanghai as the studied area and divide the area into 4422 uniform grids whose size is 500m×500m (Figure 6).

[1] Since the accurate values of experimental results are not repeatable, the Macro-F1 scores have a little difference. But the trends reflected by the experiments are provable.