Random Walk Kernel and DBSCAN

Andreas Dahlberg

December 2018

Exercise 1

Exercise 1d)

Let's assume that G and G' have n nodes each. The weight matrix for the product graph is then an $n^2 \times n^2$ matrix since the nodes in the product graph is just the cartesian product of the nodes of G and G'. In my code there are many transformations being performed between the graph represented as a nx.Graph and as a numpy matrix. Let's assume that these transformations are $O(1)$. The most expensive calculation in calculating the random walk kernel is then inverting an $n^2 \times n^2$ matrix. This takes $O(n^6)$ so the time complexity for the random walk kernel is $O(n^6)$. All operations besides inverting the matrix is much faster than $O(n^6)$ so if we can find a way to calculate the random walk kernel without explicitly calculating the inverse of the weight matrix, it would save us a lot of time.

Exercise 2

Exercise 2a)

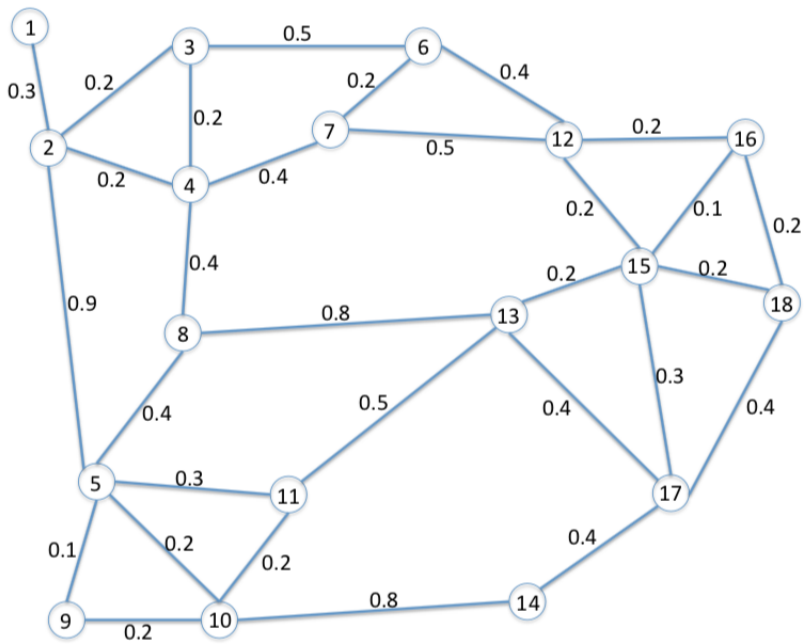

Figure 1: Graph

Below are the steps performed by the DBSCAN clustering algorithm for the graph in figure 1 in the case when $\epsilon = 0.3$ and MinPts=4. The nodes are processed from node 1 to node 18. We let $|\epsilon\_neigh(v)|$ be the size of the ϵ-neighborhood for node v.

1. $|\epsilon\_neigh(1)|$=2 so node 1 is noise.
2. $|\epsilon\_neigh(2)|$=4 so create a new cluster, cluster1 and add node 2 to cluster1. Node 2 is a corepoint.
3. Add node 1 to cluster1, node 1 is a border point.
4. Add node 3 to cluster1, $|\epsilon\_neigh(3)|$=3 so node 3 is a border point.
5. Add node 4 to cluster1, $|\epsilon\_neigh(4)|$=3 so node 4 is a border point.
6. $|\epsilon\_neigh(5)|$=4 so create a new cluster, cluster2 and add node 5 to cluster2. Node 5 is a core point.
7. Add node 9 to cluster2. $|\epsilon\_neigh(9)|$=3 so node 9 is a border point.
8. Add node 10 to cluster2. $|\epsilon\_neigh(10)|$=4 so node 10 is a core point.
9. Add node 11 to cluster2. $|\epsilon\_neigh(11)|$=3 so node 11 is a border point.
10. $|\epsilon\_neigh(6)|$=2 so node 6 is noise.
11. $|\epsilon\_neigh(7)|$=2 so node 7 is noise.

12. $|\epsilon\_neigh(8)|$=1 so node 8 is noise.
13. $|\epsilon\_neigh(12)|$=3 so node 12 is noise.
14. $|\epsilon\_neigh(13)|$=2 so node 13 is noise.
15. $|\epsilon\_neigh(14)|$=1 so node 14 is noise.
16. $|\epsilon\_neigh(15)|$=6 so node 15 is a core point. Create a new cluster, cluster3 and add node 15 to cluster3.
17. Add node 12 to cluster3, node 12 is a border point.
18. Add node 13 to cluster3, node 13 is a border point.
19. Add node 16 to cluster3. $|\epsilon\_neigh(16)|$=4 so node 16 is a core point.
20. Add node 18 to cluster3. $|\epsilon\_neigh(18)|$=3 so node 18 is a border point.
21. Add node 17 to cluster3. $|\epsilon\_neigh(17)|$=2 so node 17 is a border point.

For ϵ=0.3 and MinPts=4 we get:
Core points: 2,5,10,15,16
Border points: 1,3,4,9,11,12,13,17,18
Noise points: 6,7,8,14
Cluster1 = $\{1, 2, 3, 4\}$, Cluster2 = $\{5, 9, 10, 11\}$, Cluster3=$\{15, 12, 13, 16, 18, 17\}$

For $\epsilon = 0.4$ and MinPts=4 we get:
Core points: 2,4,5,10,12,15,16,17,18
Border points: 1,3,6,7,8,9,11,13,14
Noise points:
Cluster1 = $\{2, 1, 3, 4, 7, 8\}$, Cluster2 = $\{5, 9, 10, 11\}$, Cluster3=$\{12, 6, 15, 13, 16, 18, 17, 14\}$

For $\epsilon = 0.4$ and MinPts=5 we get:
Core points: 4,5,15,17
Border points: 2,3,7,8,9,10,11,12,13,14,16,18
Noise points: 1,6
Cluster1 = $\{4, 2, 3, 7, 8\}$, Cluster2 = $\{5, 9, 10, 11\}$, Cluster3=$\{15, 12, 13, 16, 17, 14, 18\}$

Exercise 2b)

When increasing ϵ from 0.3 to 0.4 while keeping MinPts fixed at 4 we see that the number of noise points decreases from 4 to 0 and also that the number of core points increases. These results are expected since increasing ϵ means that the ϵ-neighbourhood for a point increases and hence it is more likely to be a core point.

When increasing MinPts from 4 to 5 while keeping ϵ fixed at 0.4 we observe the opposite effect: the number of core points decreases and the number of noise points increases. These results are also expected since increasing MinPts implies a tougher condition for a point to be a core point since it has to have more points in its ϵ-neighbourhood in order to be a core point.