-- Part A
-- Problem 2

SELECT SDO_NET.GET_NO_OF_NODES('airport') FROM DUAL;

SDO_NET.GET_NO_OF_NODES('AIRPORT')
----------------------------------
             293

-- Problem 3
SELECT SDO_NET.GET_NODE_DEGREE('airport', AIRPORT_ID)
FROM airport_list
WHERE AIRPORT_NAME = 'Philadelphia, PA'

SDO_NET.GET_NODE_DEGREE('AIRPORT', AIRPORT_ID)
---------------------------------------------
             91

-- Problem 4
SELECT AVG(SDO_NET.GET_NODE_OUT_DEGREE('airport', airport_ID))
FROM airport_list

AVG(SDO_NET.GET_NODE_OUT_DEGREE('AIRPORT', AIRPORT_ID))
------------------------------------------------------
       13.9692833

-- Part B
-- Problem 1
WITH
one_from_gustavus (START_NODE_ID, END_NODE_ID, CHAIN_LEVEL) AS
  (
    SELECT START_NODE_ID, END_NODE_ID, 0 CHAIN_LEVEL
    FROM airport_link$
    WHERE START_NODE_ID = 11997
    UNION ALL
    SELECT link.START_NODE_ID, link.END_NODE_ID, o.CHAIN_LEVEL+1
    FROM one_from_gustavus o, airport_link$ link
    WHERE o.END_NODE_ID = link.START_NODE_ID AND o.CHAIN_LEVEL = 0
  )
SELECT list.AIRPORT_ID, list.AIRPORT_NAME
FROM one_from_gustavus o, airport_list list
WHERE CHAIN_LEVEL <= 1
  AND o.END_NODE_ID = list.AIRPORT_ID

AIRPORT_ID AIRPORT_NAME
---------------------------
12523 Juneau, AK
14256 Petersburg, AK
15991 Yakutat, AK
14828 Sitka, AK
10299 Anchorage, AK
11997 Gustavus, AK
12819 Ketchikan, AK
14747 Seattle, WA
-- Problem 2

WITH

    OR_to_KY(airport_id, end_id, link_level) AS
    (
        SELECT start_node_id, end_node_id, 0 link_level
        FROM airport_link$
        WHERE start_node_id = 14057
        UNION ALL
        SELECT a.start_node_id, a.end_node_id, t.link_level+1
        FROM OR_to_KY t, airport_link$ a
        WHERE t.end_id = a.start_node_id AND t.link_level < 2 AND
        t.airport_id <> a.start_node_id
    )

SELECT MIN(link_level)
FROM OR_to_KY
WHERE end_id = 14006

MIN(LINK_LEVEL)
---------------

1

SELECT C2.COUNTRY_NAME FROM COUNTRY2018 C1, COUNTRY2018 C2 WHERE
C1.COUNTRY_NAME='Norway' AND C2.COUNTRY_NAME<>'Norway' AND SDO_NN(C2.GEOM,
C1.GEOM, 'SDO_NUM_RES=5')='TRUE';

COUNTRY_NAME
---------------

Russia
Denmark
Finland
Sweden

SELECT DISTINCT c.COUNTRY_NAME
FROM RIVER2018 r, COUNTRY2018 c
WHERE RIVER_NAME = 'Nile'
AND SDO_GEOM.SDO_INTERSECTION(c.geom, r.geom) is not null;

COUNTRY_NAME
---------------

Sudan
Uganda
Egypt

SQL> SELECT C2.COUNTRY_NAME FROM COUNTRY2018 C1, COUNTRY2018 C2 WHERE
C1.COUNTRY_NAME='Norway' AND SDO_TOUCH(C1.GEOM, C2.GEOM)='TRUE';

COUNTRY_NAME
---------------

Finland
Russia
Sweden
SELECT C1.COUNTRY_NAME, C2.COUNTRY_NAME
FROM COUNTRY2018 C1, COUNTRY2018 C2, USER_SDO_GEOM_METADATA M
WHERE M.TABLE_NAME='COUNTRY2018' AND C1.COUNTRY_NAME <> C2.COUNTRY_NAME AND
SDO_GEOM.WITHIN_DISTANCE(C1.GEOM, M.DIMINFO, 1, C2.GEOM, M.DIMINFO) = 'TRUE';
TIME ~ 8hrs

SELECT C1.COUNTRY_NAME, C2.COUNTRY_NAME
FROM COUNTRY2018 C1, COUNTRY2018 C2
WHERE SDO_WITHIN_DISTANCE(C1.GEOM, C2.GEOM, 'DISTANCE = 1') = 'TRUE' and
C1.COUNTRY_NAME <> C2.COUNTRY_NAME;
TIME: 1:04.79mins

It takes too long (more than 8 hours) to get the running time for Problem 1 since it is not spatially indexed. The query in Problem 2 is much faster (about 1 min).

Part D
- The results are not statistically significant. 8 clusters were detected with the minimum p-value 0.202 and maximum p-value 0.996.
- A total of 3 clusters were detected with only 1 cluster having a p-value to be 0.01 implying that it is statistically significant with the coordinates (495.177, 504.398) and radius 79.21.

From this visualization it looks like a valid hotspot.
• It detected a total of 8 hotspots with the following details:
  o Hotspot 1:
    ▪ Coordinates / radius: (529402,181041) / 132.56
    ▪ Log likelihood ratio: 60.376872
    ▪ P-value: 0.001
  o Hotspot 2:
    ▪ Coordinates / radius: (529591,181161) / 7.50
    ▪ Log likelihood ratio: 11.371095
    ▪ P-value: 0.352
  o Hotspot 3:
    ▪ Coordinates / radius: (529247,181060) / 6.95
    ▪ Log likelihood ratio: 8.810967
    ▪ P-value: 0.775
  o Hotspot 4:
    ▪ Coordinates / radius: (529568,181142) / 7.16
    ▪ Log likelihood ratio: 8.584994
    ▪ P-value: 0.816
  o Hotspot 5:
    ▪ Coordinates / radius: (529304,180925) / 8.03
    ▪ Log likelihood ratio: 7.719316
    ▪ P-value: 0.927
  o Hotspot 6:
    ▪ Coordinates / radius: (529522,180941) / 8.34
    ▪ Log likelihood ratio: 7.428179
    ▪ P-value: 0.950
  o Hotspot 7:
    ▪ Coordinates / radius: (529570,181062) / 15.08
    ▪ Log likelihood ratio: 6.918324
    ▪ P-value: 0.982
  o Hotspot 8:
    ▪ Coordinates / radius: (529223,180979) / 12.46
    ▪ Log likelihood ratio: 6.635646
    ▪ P-value: 0.989

Out of all the hotspots detected, there is only 1 which is statistically significant i.e., Hotspot 1 having a p-value of 0.001 which is less than 0.01. All the other hotspots have a p-value > 0.01.

• It detected only 1 hotspot:
  o Hotspot 1:
    ▪ Coordinates / radius: (529427,181055) / 231.10
    ▪ Log likelihood ratio: 856.396950
    ▪ P-value: 0.001

With the updated range of study, the result is more filtered giving only 1 hotspot with a p-value > 0.01. The center for the hotspot in this range and the range in previous question are almost the same but there is a big difference in the radius of the circle. Also, the log likelihood ration increased a lot. This difference is because of the distribution of points in the dataset. Number of points detected as a hotspot in this polygon are scattered which were thought of as separate hotspots by the first polygon although they were statistically insignificant.