Assignment 2

Jacky(Hsiu-Hsuan) Yeh

```
import rdata
import pandas as pd
import numpy as np
from libpysal.weights import full2W
from spreg import ML_Lag
import networkx as nx


def preprocess(adjacency_matrix):
    np.random.seed(1126)
    N = adjacency_matrix.shape[0]

    non_zero = np.argwhere(adjacency_matrix.diagonal() == 1)
    if len(non_zero) != 0:
        for i in non_zero[0]:
            adjacency_matrix[i][i] = 0

    # iso = np.where(~adjacency_matrix.any(axis=1))[0]
    # for i in iso:
    #     idx = np.random.choice(range(N))
    #     while idx == i:
    #         idx = np.random.choice(range(N))
    #     adjacency_matrix[i][idx] = 1


data_matrix, adjacency_matrix = rdata.read_rda(
    'dataset_network_interactions.RData'
)['dataset_lecture3']
preprocess(adjacency_matrix)
W = full2W(adjacency_matrix)
```

```
/home/r12323011/.local/lib/python3.10/site-packages/libpysal/weights/util.py:768: UserWarning
```

```
There are 132 disconnected components.
There are 93 islands with ids: 3, 8, 25, 30, 32, 33, 35, 48, 53, 60, 61, 68, 72, 74, 76, 80
 return W(neighbors, weights, id_order=ids, **kwargs)
```

```
_X = pd.DataFrame(
    data_matrix,
    columns=[
        'gpa', 'smoke_freq', 'drink', 'club', 'exer',
        'male', 'black', 'hisp', 'asian', 'race_other',
        'both_par', 'less_hs', 'more_hs', 'momedu_miss',
        'Prof', 'Home', 'job_other'
    ]
)
y = _X.gpa.to_numpy()
name_x = [
    'male', 'black', 'hisp', 'asian',
    'race_other', 'both_par', 'less_hs',
    'more_hs', 'momedu_miss', 'Prof',
    'Home', 'job_other'
]
X = _X[name_x]
```

Problem 1

Using the data set dataset_network_interactions provided in the course demonstration, estimate the spatial Durbin model where the regressors WX are included. Report your estimation results.

```
W.transform = 'r'
WX = np.dot(W.full()[0], X)
X_extended = np.hstack([X, WX])

model = ML_Lag(
    y, X_extended, W,
    method='full',
    name_y = 'gpa',
    name_x = name_x + [f'spatial_{i}'for i in name_x]
)
print(model.summary)
```

```
REGRESSION RESULTS
```

```
------------------

SUMMARY OF OUTPUT: MAXIMUM LIKELIHOOD SPATIAL LAG (METHOD = FULL)
-----------------------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :         gpa                Number of Observations:         336
Mean dependent var  :      2.9779                Number of Variables   :          26
S.D. dependent var  :      0.6925                Degrees of Freedom    :         310
Pseudo R-squared    :      0.2196
Spatial Pseudo R-squared:  0.2096
Log likelihood      :   -311.7780
Sigma-square ML     :      0.3732                Akaike info criterion :     675.556
S.E of regression   :      0.6109                Schwarz criterion     :     774.801

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT         3.07745         0.18805        16.36527         0.00000
                male        -0.11223         0.07242        -1.54973         0.12121
               black        -0.31892         0.14309        -2.22878         0.02583
                hisp        -0.46580         0.17252        -2.70002         0.00693
               asian        -0.06983         0.15069        -0.46337         0.64310
          race_other        -0.49197         0.39402        -1.24859         0.21182
            both_par        -0.01411         0.08690        -0.16242         0.87098
             less_hs         0.18006         0.16361         1.10052         0.27110
             more_hs         0.11881         0.10584         1.12257         0.26162
         momedu_miss         0.02249         0.13476         0.16685         0.86749
                Prof        -0.03509         0.12624        -0.27794         0.78106
                Home        -0.18428         0.15067        -1.22309         0.22130
           job_other        -0.15493         0.12564        -1.23311         0.21753
        spatial_male         0.12237         0.10217         1.19776         0.23101
       spatial_black        -0.31118         0.15433        -2.01639         0.04376
        spatial_hisp        -0.27512         0.26477        -1.03911         0.29875
       spatial_asian         0.15733         0.18979         0.82896         0.40713
  spatial_race_other         0.58706         0.45905         1.27887         0.20094
    spatial_both_par        -0.20306         0.12637        -1.60690         0.10808
     spatial_less_hs        -0.18725         0.23405        -0.80005         0.42368
     spatial_more_hs         0.10232         0.16445         0.62218         0.53383
 spatial_momedu_miss         0.10709         0.21110         0.50731         0.61194
        spatial_Prof         0.11018         0.19381         0.56852         0.56968
        spatial_Home         0.09747         0.22165         0.43975         0.66012
   spatial_job_other         0.23712         0.18707         1.26754         0.20496
```

```
            W_gpa         0.09614          0.04626          2.07835          0.03768
--------------------------------------------------------------------------------
================================ END OF REPORT =================================
```

Problem 2

Change the specification of weights matrix to binary values, i.e., do not row-standardize the weights matrix, and estimate the above spatial Durbin model again. Report your estimation results.

```
W.transform = 'b'
WX = np.dot(W.full()[0], X)
X_extended = np.hstack([X, WX])

model = ML_Lag(
    y, X_extended, W,
    method='full',
    name_y = 'gpa',
    name_x = name_x + [f'spatial_{i}'for i in name_x]
)
print(model.summary)
```

```
REGRESSION RESULTS
------------------

SUMMARY OF OUTPUT: MAXIMUM LIKELIHOOD SPATIAL LAG (METHOD = FULL)
-----------------------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :         gpa                Number of Observations:         336
Mean dependent var  :      2.9779                Number of Variables   :          26
S.D. dependent var  :      0.6925                Degrees of Freedom    :         310
Pseudo R-squared    :      0.1854
Spatial Pseudo R-squared:  0.1773
Log likelihood      :   -318.6715
Sigma-square ML     :      0.3895                Akaike info criterion :     689.343
S.E of regression   :      0.6241                Schwarz criterion     :     788.588

--------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
--------------------------------------------------------------------------------
```

```
            CONSTANT         3.22754         0.18861        17.11204         0.00000
                male        -0.10860         0.07365        -1.47441         0.14037
               black        -0.42628         0.14336        -2.97356         0.00294
                hisp        -0.47935         0.17408        -2.75370         0.00589
               asian         0.05779         0.15179         0.38075         0.70339
          race_other        -0.22913         0.38991        -0.58765         0.55677
            both_par        -0.04574         0.08958        -0.51060         0.60963
             less_hs         0.18068         0.16825         1.07388         0.28288
             more_hs         0.11524         0.10930         1.05440         0.29170
         momedu_miss        -0.00849         0.13948        -0.06086         0.95147
                Prof        -0.03637         0.12922        -0.28143         0.77838
                Home        -0.08751         0.15420        -0.56753         0.57035
           job_other        -0.14233         0.12971        -1.09730         0.27251
        spatial_male         0.07309         0.06099         1.19837         0.23077
       spatial_black        -0.07201         0.08104        -0.88863         0.37420
        spatial_hisp        -0.19482         0.14812        -1.31525         0.18843
       spatial_asian        -0.08069         0.10247        -0.78749         0.43100
  spatial_race_other         0.35969         0.34129         1.05390         0.29193
    spatial_both_par        -0.11759         0.08170        -1.43937         0.15005
     spatial_less_hs        -0.02532         0.13221        -0.19151         0.84812
     spatial_more_hs        -0.00721         0.08818        -0.08176         0.93483
 spatial_momedu_miss         0.10036         0.10633         0.94382         0.34526
        spatial_Prof         0.07818         0.11698         0.66835         0.50391
        spatial_Home         0.01584         0.13977         0.11330         0.90979
   spatial_job_other         0.09404         0.10899         0.86283         0.38823
               W_gpa         0.03513         0.02486         1.41325         0.15758
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

```
/home/r12323011/.local/lib/python3.10/site-packages/spreg/ml_lag.py:634: RuntimeWarning: inva
  jacob = np.log(np.linalg.det(a))
```

Problem 3

Redo 2. by adding an additional regressor of individual network degree. Report your estimation results.

```
G = nx.from_numpy_array(adjacency_matrix, create_using=nx.DiGraph)
in_degree_centrality = nx.in_degree_centrality(G)
out_degree_centrality = nx.out_degree_centrality(G)
```

```
in_degree = np.array(list(in_degree_centrality.values()))
out_degree = np.array(list(out_degree_centrality.values()))
degree = (in_degree + out_degree).reshape(-1, 1)
```

```
X_extended = np.hstack([degree, X, WX])

model = ML_Lag(
    y, X_extended, W,
    method='full',
    name_y = 'gpa',
    name_x = ['total_degree'] + name_x + [f'spatial_{i}'for i in name_x]
)
print(model.summary)
```

```
REGRESSION RESULTS
------------------

SUMMARY OF OUTPUT: MAXIMUM LIKELIHOOD SPATIAL LAG (METHOD = FULL)
-----------------------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :         gpa                Number of Observations:         336
Mean dependent var  :      2.9779                Number of Variables   :          27
S.D. dependent var  :      0.6925                Degrees of Freedom    :         309
Pseudo R-squared    :      0.1944
Spatial Pseudo R-squared:  0.1802
Log likelihood      :   -317.5684
Sigma-square ML     :      0.3853                Akaike info criterion :     689.137
S.E of regression   :      0.6207                Schwarz criterion     :     792.199

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT         3.26107         0.18873        17.27879         0.00000
        total_degree       -48.49033        31.29689        -1.54937         0.12129
                male        -0.10955         0.07327        -1.49526         0.13485
               black        -0.45007         0.14364        -3.13342         0.00173
                hisp        -0.51526         0.17448        -2.95318         0.00315
               asian         0.02279         0.15232         0.14963         0.88106
          race_other        -0.25173         0.38802        -0.64875         0.51650
            both_par        -0.04018         0.08911        -0.45089         0.65207
```

```
             less_hs         0.18246          0.16734          1.09036          0.27555
             more_hs         0.11309          0.10873          1.04008          0.29830
         momedu_miss        -0.00566          0.13877         -0.04077          0.96748
                Prof        -0.03547          0.12858         -0.27586          0.78265
                Home        -0.08797          0.15339         -0.57347          0.56632
           job_other        -0.14320          0.12908         -1.10940          0.26726
        spatial_male         0.08207          0.06109          1.34352          0.17910
       spatial_black        -0.00382          0.09399         -0.04059          0.96762
        spatial_hisp        -0.10869          0.15572         -0.69798          0.48519
       spatial_asian        -0.00729          0.11288         -0.06455          0.94853
  spatial_race_other         0.45906          0.34443          1.33284          0.18259
    spatial_both_par        -0.06929          0.08677         -0.79853          0.42456
     spatial_less_hs        -0.00546          0.13188         -0.04141          0.96697
     spatial_more_hs         0.04775          0.09382          0.50898          0.61077
 spatial_momedu_miss         0.12596          0.10695          1.17769          0.23892
       spatial_Prof          0.12534          0.12177          1.02935          0.30331
       spatial_Home          0.07980          0.14661          0.54429          0.58624
   spatial_job_other         0.14468          0.11509          1.25717          0.20869
               W_gpa         0.06359          0.02912          2.18390          0.02897
------------------------------------------------------------------------------------
================================ END OF REPORT =====================================
```

```
/home/r12323011/.local/lib/python3.10/site-packages/spreg/ml_lag.py:634: RuntimeWarning: inva
  jacob = np.log(np.linalg.det(a))
```

contextualized degree centrality

```
X_degree = pd.concat([pd.DataFrame(degree), X], axis=1).rename(columns={0: 'degree'})
WX = np.dot(W.full()[0], X_degree)
X_extended = np.hstack([X_degree, WX])

model = ML_Lag(
    y, X_extended, W,
    method='full',
    name_y = 'gpa',
    name_x = ['total_degree'] + name_x + ['spatial_total_degree'] + [f'spatial_{i}'for i in n
)
print(model.summary)
```

```
REGRESSION RESULTS
```

```
------------------

SUMMARY OF OUTPUT: MAXIMUM LIKELIHOOD SPATIAL LAG (METHOD = FULL)
-----------------------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :         gpa                Number of Observations:         336
Mean dependent var  :      2.9779                Number of Variables   :          28
S.D. dependent var  :      0.6925                Degrees of Freedom    :         308
Pseudo R-squared    :      0.1945
Spatial Pseudo R-squared:  0.1808
Log likelihood      :   -317.5531
Sigma-square ML     :      0.3852                Akaike info criterion :     691.106
S.E of regression   :      0.6207                Schwarz criterion     :     797.985

--------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
--------------------------------------------------------------------------------
            CONSTANT         3.26036         0.18875        17.27319         0.00000
        total_degree       -47.43946        31.53471        -1.50436         0.13249
                male        -0.11028         0.07340        -1.50235         0.13301
               black        -0.45060         0.14366        -3.13655         0.00171
                hisp        -0.51492         0.17449        -2.95109         0.00317
               asian         0.02276         0.15232         0.14944         0.88120
          race_other        -0.25028         0.38806        -0.64495         0.51896
            both_par        -0.03982         0.08914        -0.44675         0.65506
             less_hs         0.18086         0.16756         1.07939         0.28042
             more_hs         0.11346         0.10874         1.04342         0.29675
         momedu_miss        -0.00708         0.13903        -0.05092         0.95939
                Prof        -0.03607         0.12864        -0.28040         0.77917
                Home        -0.08817         0.15339        -0.57480         0.56543
           job_other        -0.14405         0.12919        -1.11503         0.26484
spatial_total_degree        -0.52640         3.01605        -0.17453         0.86145
        spatial_male         0.08356         0.06172         1.35392         0.17576
       spatial_black        -0.00227         0.09457        -0.02397         0.98087
        spatial_hisp        -0.10840         0.15577        -0.69588         0.48650
       spatial_asian        -0.00441         0.11399        -0.03872         0.96911
  spatial_race_other         0.45727         0.34451         1.32730         0.18441
    spatial_both_par        -0.06728         0.08757        -0.76830         0.44231
     spatial_less_hs        -0.00671         0.13204        -0.05085         0.95945
     spatial_more_hs         0.04727         0.09386         0.50364         0.61451
 spatial_momedu_miss         0.12682         0.10705         1.18465         0.23616
       spatial_Prof          0.12585         0.12180         1.03321         0.30151
```

```
        spatial_Home         0.07661         0.14766         0.51878         0.60391
   spatial_job_other         0.14568         0.11524         1.26411         0.20619
               W_gpa         0.06348         0.02929         2.16701         0.03023
------------------------------------------------------------------------------------
================================= END OF REPORT ====================================


/home/r12323011/.local/lib/python3.10/site-packages/spreg/ml_lag.py:634: RuntimeWarning: inva
  jacob = np.log(np.linalg.det(a))
```