

PageRank

May 22, 2018

0.1 Step 3.3

```
In [167]: import csv
import pandas as pd
import numpy as np

In [168]: pd_df=pd.read_csv("web-NotreDame.txt",delimiter="\t", skiprows=4, names=['from_node',
In [169]: pd_df=pd_df[(pd_df['from_node']<10000) & (pd_df['to_node']<10000)]

In [170]: fre_df1=pd_df.groupby(['from_node'])['from_node'].agg({"count": len}).sort_values("cou

In [171]: count_df=pd.merge(pd_df,fre_df1, left_on='from_node', right_on='from_node',how='inner')
#count_df=count_df.dropna

In [172]: def inverse(x):
return(1/x)

In [173]: count_df['count']=count_df['count'].apply(inverse)

In [174]: M=np.zeros((10000,10000))

In [175]: def create_matrix(row):
M[int(row.iloc[1])][int(row.iloc[0])]=row.iloc[2]

In [176]: df=count_df.apply(lambda row:create_matrix(row),axis=1)

In [177]: M1=M[10:30,10:30]
M1

Out[177]: array([[ 0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.],
                 [ 0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.],
                 [ 0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.],
                 [ 0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.],
                 [ 0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.],
                 [ 0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.,
                  0.,      0.,      0.,      0.,      0.,      0.]])
```

```

[ 0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ],
[ 0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ],
[ 0.01298701, 0.01818182, 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ],
[ 0.01298701, 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ],
[ 0.01298701, 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ],
[ 0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.0106383 ,
  0.00414938, 0.      , 0.      , 0.07142857, 0.07142857,
  0.03846154, 0.09090909, 0.07692308, 0.      , 0.      ],
[ 0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ],
[ 0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.0106383 ,
  0.00414938, 0.      , 0.      , 0.07142857, 0.07142857,
  0.03846154, 0.09090909, 0.07692308, 0.      , 0.      ],
[ 0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.0106383 ,
  0.00414938, 0.      , 0.      , 0.07142857, 0.07142857,
  0.03846154, 0.09090909, 0.07692308, 0.      , 0.      ],
[ 0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.0106383 ,
  0.00414938, 0.08333333, 0.      , 0.07142857, 0.07142857,
  0.03846154, 0.09090909, 0.07692308, 0.      , 0.      ],
[ 0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.      ],
[ 0.      , 0.      , 0.      , 0.      , 0.      ,
  0.      , 0.      , 0.      , 0.      , 0.0106383 ,
  0.00414938, 0.      , 0.      , 0.07142857, 0.07142857,
  0.03846154, 0.09090909, 0.07692308, 0.      , 0.      ],

```

```
[ 0.          ,  0.          ,  0.          ,  0.          ,  0.          ,
  0.          ,  0.          ,  0.          ,  0.          ,  0.0106383 ,
  0.00414938,  0.          ,  0.          ,  0.07142857,  0.07142857,
  0.03846154,  0.09090909,  0.07692308,  0.          ,  0.          ],
[ 0.          ,  0.          ,  0.          ,  0.          ,  0.          ,
  0.          ,  0.          ,  0.          ,  0.          ,  0.0106383 ,
  0.00414938,  0.          ,  0.          ,  0.07142857,  0.07142857,
  0.03846154,  0.09090909,  0.07692308,  0.          ,  0.          ],
[ 0.          ,  0.          ,  0.          ,  0.          ,  0.          ,
  0.          ,  0.          ,  0.          ,  0.          ,  0.          ,
  0.          ,  0.          ,  0.          ,  0.          ,  0.          ],
[ 0.          ,  0.          ,  0.          ,  0.          ,  0.          ,
  0.          ,  0.          ,  0.          ,  0.          ,  0.          ],
[ 0.          ,  0.          ,  0.          ,  0.          ,  0.          ,
  0.          ,  0.          ,  0.          ,  0.          ,  0.0106383 ,
  0.00414938,  0.          ,  0.          ,  0.07142857,  0.07142857,
  0.03846154,  0.09090909,  0.07692308,  0.          ,  0.          ],
[ 0.          ,  0.          ,  0.          ,  0.          ,  0.          ,
  0.          ,  0.          ,  0.          ,  0.          ,  0.          ,
  0.          ,  0.          ,  0.          ,  0.          ,  0.          ],
[ 0.          ,  0.          ,  0.          ,  0.          ,  0.          ,
  0.          ,  0.          ,  0.          ,  0.          ,  0.          ],
[ 0.          ,  0.          ,  0.          ,  0.          ,  0.          ,
  0.          ,  0.          ,  0.          ,  0.          ,  0.0106383 ,
  0.00414938,  0.          ,  0.          ,  0.07142857,  0.07142857,
  0.03846154,  0.09090909,  0.07692308,  0.          ,  0.          ]])
```

```
In [178]: def pagerank(M, alpha, num_iter):
          PR = np.ones((10000))
          PR=np.transpose(PR)
          one=np.ones(10000)
          one=np.transpose(one)
          for i in range(num_iter):
              PR=alpha*np.dot(M,PR)+(1-alpha)*one
          return PR

In [179]: page_rank=pagerank(M,0.85, 15)

In [180]: out_df = pd.DataFrame(page_rank,columns=['pagerank'])

In [181]: out_df=out_df.reset_index()

In [182]: out_df=out_df.sort_values('pagerank', ascending=False)
          out_df.reset_index()
          out_df.columns = ['node', 'pagerank']
          out_df.reset_index(drop = True, inplace = True)

In [183]: out_df.head(10)

Out[183]:   node  pagerank
0      0  224.702638
```

1	1973	189.250314
2	1790	53.438593
3	1828	50.954873
4	1	27.975911
5	238	26.779136
6	140	23.520898
7	14	22.232264
8	16	21.591054
9	162	18.283386