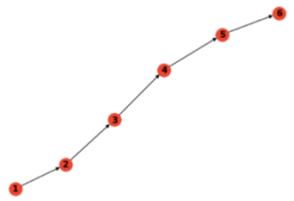
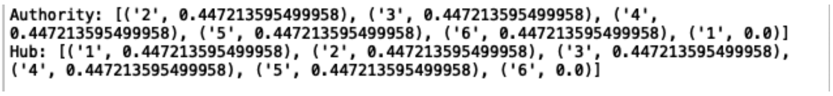
Datamining HW3 Report

Graph_1:



- HITS

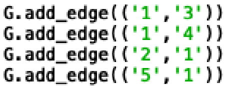
No node points to Node 1 so in this graph we can predict that Node 1 has lowest authority. Also from the graph we can tell Node 6 has 0 hub.



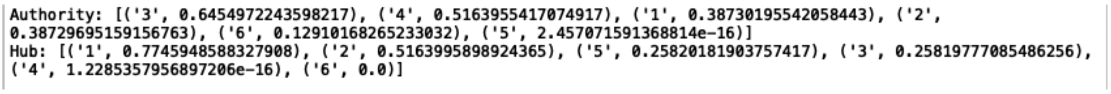
```
Authority: [('2', 0.447213595499958), ('3', 0.447213595499958), ('4',
0.447213595499958), ('5', 0.447213595499958), ('6', 0.447213595499958), ('1', 0.0)]
Hub: [('1', 0.447213595499958), ('2', 0.447213595499958), ('3', 0.447213595499958),
('4', 0.447213595499958), ('5', 0.447213595499958), ('6', 0.0)]
```

- Increase hub, authority:

To have a good authority Node 1 needs to have many nodes to point to it. To increase hub Node 1 needs to point to other nodes more than 1 time. Therefore, I add some links “from” Node 1 and “to” Node 1.



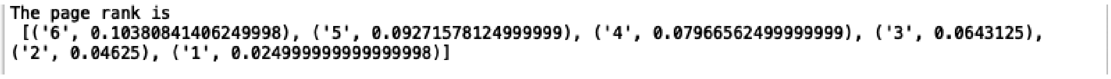
```
G.add_edge(('1','3'))
G.add_edge(('1','4'))
G.add_edge(('2','1'))
G.add_edge(('5','1'))
```



```
Authority: [('3', 0.6454972243598217), ('4', 0.5163955417074917), ('1', 0.38730195542058443), ('2',
0.38729695159156763), ('6', 0.12910168265233032), ('5', 2.457071591368814e-16)]
Hub: [('1', 0.7745948588327908), ('2', 0.5163995898924365), ('5', 0.25820181903757417), ('3', 0.25819777085486256),
('4', 1.2285357956897206e-16), ('6', 0.0)]
```

- PageRank

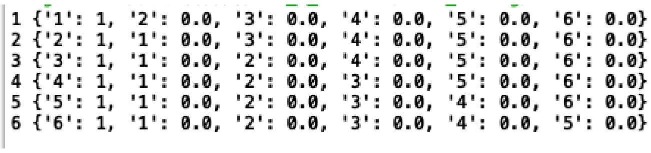
Different from authority and hub, pagerank consider the history status even if only Node 5 points to it however, Node 5 is pointed by Node4, Node 3, Node 2 and Node 1.



```
The page rank is
 [('6', 0.10380841406249998), ('5', 0.09271578124999999), ('4', 0.07966562499999999), ('3', 0.0643125),
('2', 0.04625), ('1', 0.024999999999999998)]
```

- SimRank

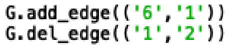
The first number means the Node number. And then shows the pair-wise similarity of nodes. S(1,1) = 1, S(1,2) = 0.0, S(1,3) = 0.0



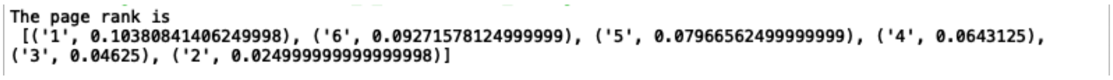
```
1 {'1': 1, '2': 0.0, '3': 0.0, '4': 0.0, '5': 0.0, '6': 0.0}
2 {'2': 1, '1': 0.0, '3': 0.0, '4': 0.0, '5': 0.0, '6': 0.0}
3 {'3': 1, '1': 0.0, '2': 0.0, '4': 0.0, '5': 0.0, '6': 0.0}
4 {'4': 1, '1': 0.0, '2': 0.0, '3': 0.0, '5': 0.0, '6': 0.0}
5 {'5': 1, '1': 0.0, '2': 0.0, '3': 0.0, '4': 0.0, '6': 0.0}
6 {'6': 1, '1': 0.0, '2': 0.0, '3': 0.0, '4': 0.0, '5': 0.0}
```

- Increase PageRank

To increase PageRank of Node 1 I let Node 1 become the last connected node. I deleted the link from Node 1 to Node 2 and added a link between Node 6 and Node 1.

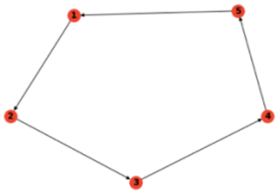


```
G.add_edge(('6','1'))
G.del_edge(('1','2'))
```

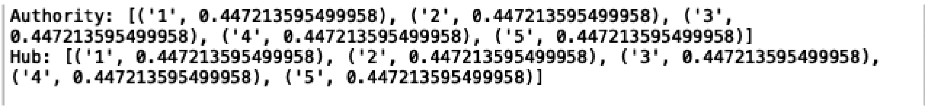


```
The page rank is
 [('1', 0.10380841406249998), ('6', 0.09271578124999999), ('5', 0.07966562499999999), ('4', 0.0643125),
('3', 0.04625), ('2', 0.024999999999999998)]
```

Graph_2:



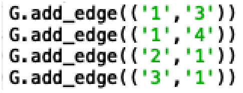
- HITS



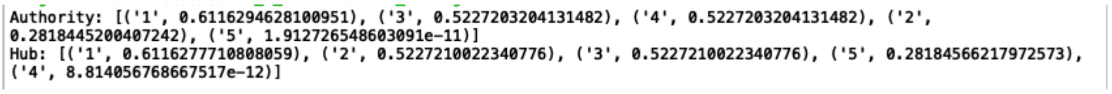
```
Authority: [('1', 0.447213595499958), ('2', 0.447213595499958), ('3',
0.447213595499958), ('4', 0.447213595499958), ('5', 0.447213595499958)]
Hub: [('1', 0.447213595499958), ('2', 0.447213595499958), ('3', 0.447213595499958),
('4', 0.447213595499958), ('5', 0.447213595499958)]
```

- Increase hub, authority:

After adding the links “from” and “to” Node 1 we can see there are some different between old hub/authority and new hub/authority.

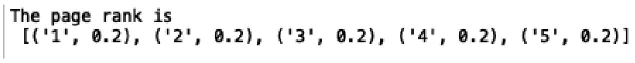


```
G.add_edge(('1','3'))
G.add_edge(('1','4'))
G.add_edge(('2','1'))
G.add_edge(('3','1'))
```



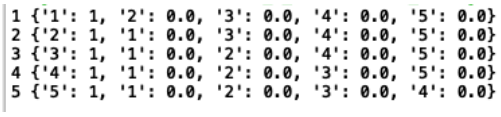
```
Authority: [('1', 0.6116294628100951), ('3', 0.5227203204131482), ('4', 0.5227203204131482), ('2',
0.2818445200407242), ('5', 1.912726548603091e-11)]
Hub: [('1', 0.6116277710808059), ('2', 0.5227210022340776), ('3', 0.5227210022340776), ('5', 0.28184566217972573),
('4', 8.814056768667517e-12)]
```

- PageRank



```
The page rank is
 [('1', 0.2), ('2', 0.2), ('3', 0.2), ('4', 0.2), ('5', 0.2)]
```

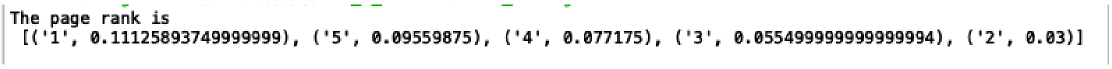
- SimRank



```
1 {'1': 1, '2': 0.0, '3': 0.0, '4': 0.0, '5': 0.0}
2 {'2': 1, '1': 0.0, '3': 0.0, '4': 0.0, '5': 0.0}
3 {'3': 1, '1': 0.0, '2': 0.0, '4': 0.0, '5': 0.0}
4 {'4': 1, '1': 0.0, '2': 0.0, '3': 0.0, '5': 0.0}
5 {'5': 1, '1': 0.0, '2': 0.0, '3': 0.0, '4': 0.0}
```

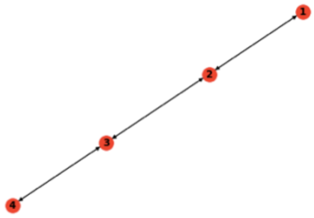
- Increase PageRank

The page rank in each node are the same because it is a one direction circular graph. I deleted the link from Node 1 to Node 2 and let Node 1 the last connected node.

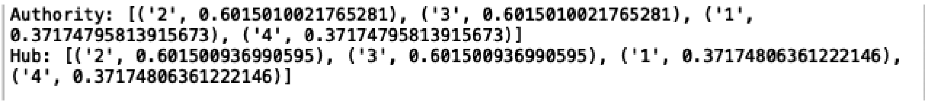


```
The page rank is
 [('1', 0.11125893749999999), ('5', 0.09559875), ('4', 0.077175), ('3', 0.055499999999999994), ('2', 0.03)]
```

Graph_3:



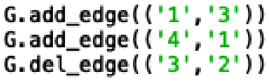
- HITS



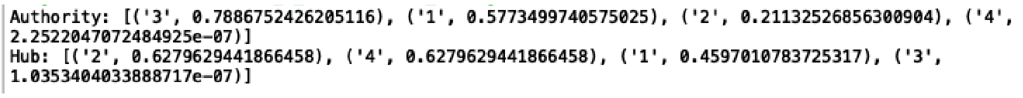
```
Authority: [('2', 0.6015010021765281), ('3', 0.6015010021765281), ('1',
0.37174795813915673), ('4', 0.37174795813915673)]
Hub: [('2', 0.601500936990595), ('3', 0.601500936990595), ('1', 0.37174806361222146),
('4', 0.37174806361222146)]
```

- Increase hub, authority:

The authority and hub in Node 2 and Node 3 are higher than Node 1, therefore, I deleted the link between Node 2 and Node 3 and added 1 extra link “from” and “to” Node 1.

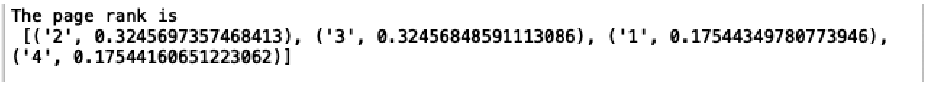


```
G.add_edge(('1','3'))
G.add_edge(('4','1'))
G.del_edge(('3','2'))
```



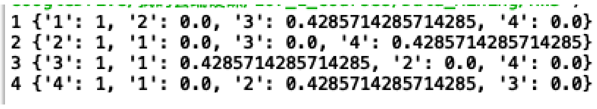
```
Authority: [('3', 0.7886752426205116), ('1', 0.5773499740575025), ('2', 0.21132526856300904), ('4',
2.2522047072484925e-07)]
Hub: [('2', 0.6279629441866458), ('4', 0.6279629441866458), ('1', 0.4597010783725317), ('3',
1.0353404033888717e-07)]
```

- PageRank



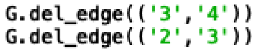
```
The page rank is
 [('2', 0.3245697357468413), ('3', 0.32456848591113086), ('1', 0.17544349780773946),
('4', 0.17544160651223062)]
```

- SimRank

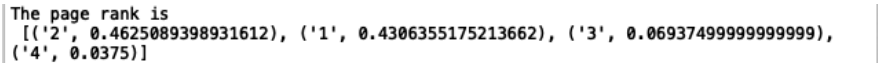


```
1 {'1': 1, '2': 0.0, '3': 0.4285714285714285, '4': 0.0}
2 {'2': 1, '1': 0.0, '3': 0.0, '4': 0.4285714285714285}
3 {'3': 1, '1': 0.4285714285714285, '2': 0.0, '4': 0.0}
4 {'4': 1, '1': 0.0, '2': 0.4285714285714285, '3': 0.0}
```

- Increase PageRank

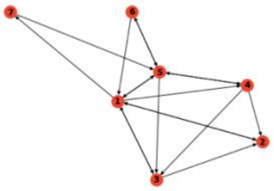


```
G.del_edge(('3','4'))
G.del_edge(('2','3'))
```

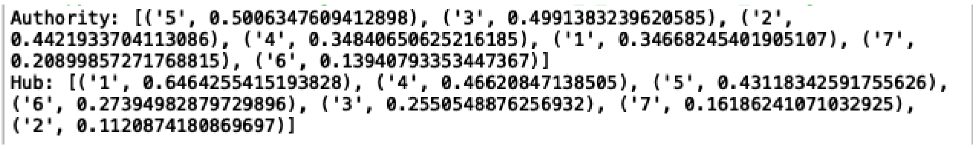


```
The page rank is
 [('2', 0.4625089398931612), ('1', 0.4306355175213662), ('3', 0.06937499999999999),
('4', 0.0375)]
```

Graph_4:

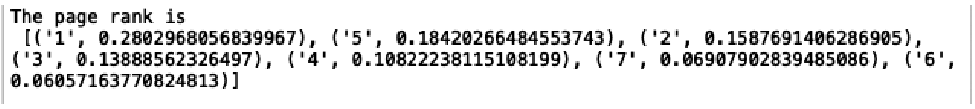


- HITS



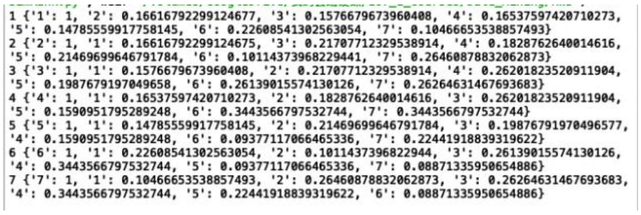
```
Authority: [('5', 0.5006347609412898), ('3', 0.4991383239620585), ('2',
0.4421933704113086), ('4', 0.34840650625216185), ('1', 0.34668245401905107), ('7',
0.20899857271768815), ('6', 0.13940793353447367)]
Hub: [('1', 0.6464255415193828), ('4', 0.46620847138505), ('5', 0.43118342591755626),
('6', 0.27394982879729896), ('3', 0.2550548876256932), ('7', 0.16186241071032925),
('2', 0.1120874180869697)]
```

- PageRank



```
The page rank is
 [('1', 0.2802968056839967), ('5', 0.18420266484553743), ('2', 0.1587691406286905),
('3', 0.13888562326497), ('4', 0.10822238115108199), ('7', 0.06907902839485086), ('6',
0.06057163770824813)]
```

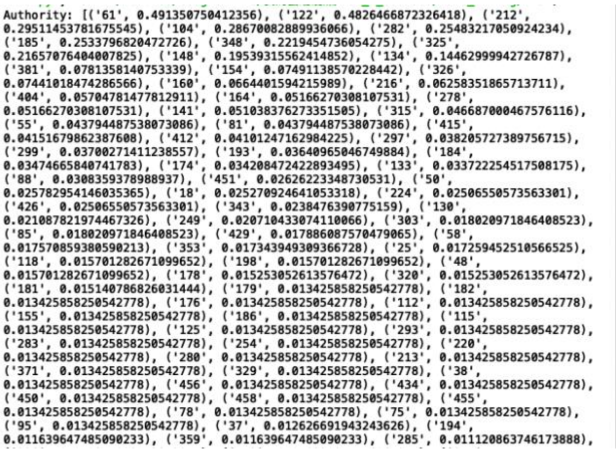
- SimRank



```
1 {'1': 1, '2': 0.16616792299124677, '3': 0.1576679673960408, '4': 0.16537597420710273,
'5': 0.14785559917758145, '6': 0.22608541302563054, '7': 0.10466653538857493}
2 {'2': 1, '1': 0.16616792299124675, '3': 0.21707712329538914, '4': 0.1828762640014616,
'5': 0.21469699646791784, '6': 0.10114373968229441, '7': 0.26460878832062873}
3 {'3': 1, '1': 0.1576679673960408, '2': 0.21707712329538914, '4': 0.26201823520911904,
'5': 0.1987679197049658, '6': 0.26139015574130126, '7': 0.26264631467693683}
4 {'4': 1, '1': 0.16537597420710273, '2': 0.1828762640014616, '3': 0.26201823520911904,
'5': 0.1590951795289248, '6': 0.3443566797532744, '7': 0.3443566797532744}
5 {'5': 1, '1': 0.14785559917758145, '2': 0.21469699646791784, '3': 0.19876791970496577,
'4': 0.1590951795289248, '6': 0.09377117066465336, '7': 0.22441918839319622}
6 {'6': 1, '1': 0.22608541302563054, '2': 0.1011437396822944, '3': 0.26139015574130126,
'4': 0.3443566797532744, '5': 0.09377117066465336, '7': 0.08871335950654886}
7 {'7': 1, '1': 0.10466653538857493, '2': 0.26460878832062873, '3': 0.26264631467693683,
'4': 0.3443566797532744, '5': 0.22441918839319622, '6': 0.08871335950654886}
```

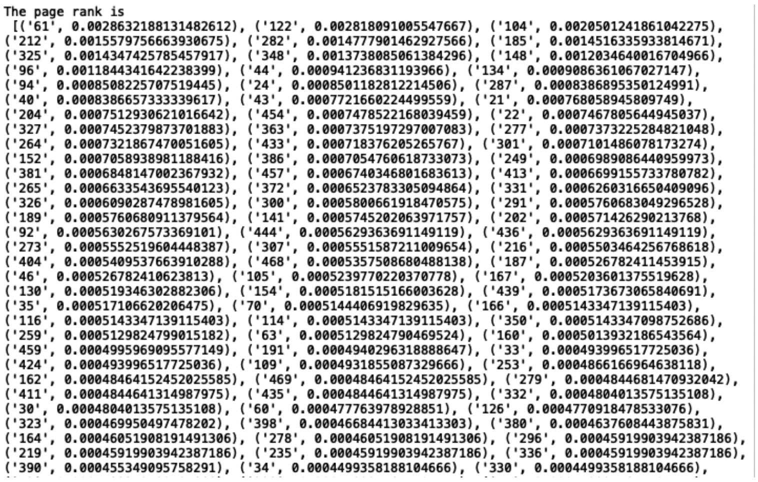
Graph_5:

- HITS



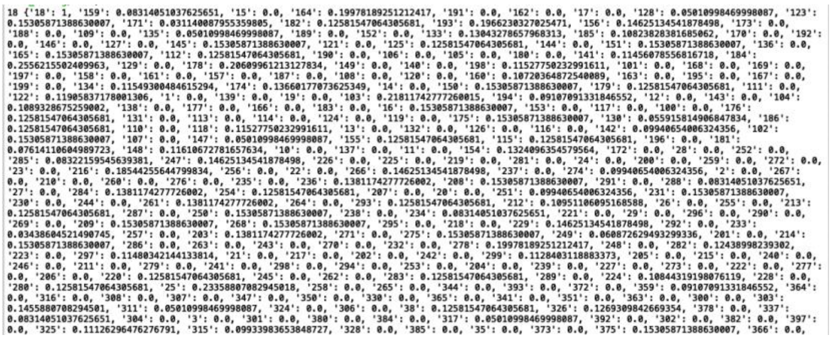
```
Authority: [('61', 0.491350750412356), ('122', 0.4826466872326418), ('212',
0.29511453781675545), ('104', 0.28670082889936066), ('282', 0.25483217050924234),
('185', 0.2533796820472726), ('348', 0.2219454736054275), ('325',
0.21657076404007825), ('148', 0.19539315562414852), ('134', 0.14462999942726787),
('381', 0.0781358140753339), ('154', 0.07491138570228442), ('326',
0.07441018474286566), ('160', 0.0664401594215989), ('216', 0.06258351865713711),
('404', 0.05704781477812911), ('164', 0.05166270308107531), ('278',
0.05166270308107531), ('141', 0.051038376273351505), ('315', 0.046687000467576116),
('55', 0.043794487538073086), ('81', 0.043794487538073086), ('415',
0.04151679862387608), ('412', 0.04101247162984225), ('297', 0.038205727389756715),
('299', 0.03700271411238557), ('193', 0.03640965046749884), ('184',
0.03474665840741783), ('174', 0.034208472422893495), ('133', 0.033722254517508175),
('88', 0.0308359378988937), ('451', 0.02626223348730531), ('50',
0.025782954146035365), ('18', 0.025270924641053318), ('224', 0.02506550573563301),
('426', 0.02506550573563301), ('343', 0.0238476390775159), ('130',
0.021087821974467326), ('249', 0.020710433074110066), ('303', 0.018020971846408523),
('85', 0.018020971846408523), ('429', 0.017886087570479065), ('58',
0.017570859380590213), ('353', 0.017343949309366728), ('25', 0.017259452510566525),
('118', 0.015701282671099652), ('198', 0.015701282671099652), ('48',
0.015701282671099652), ('178', 0.015253052613576472), ('320', 0.015253052613576472),
('181', 0.015140786826031444), ('179', 0.013425858250542778), ('182',
0.013425858250542778), ('176', 0.013425858250542778), ('112', 0.013425858250542778),
('155', 0.013425858250542778), ('186', 0.013425858250542778), ('115',
0.013425858250542778), ('125', 0.013425858250542778), ('293', 0.013425858250542778),
('283', 0.013425858250542778), ('254', 0.013425858250542778), ('220',
0.013425858250542778), ('280', 0.013425858250542778), ('213', 0.013425858250542778),
('371', 0.013425858250542778), ('329', 0.013425858250542778), ('38',
0.013425858250542778), ('456', 0.013425858250542778), ('434', 0.013425858250542778),
('450', 0.013425858250542778), ('458', 0.013425858250542778), ('455',
0.013425858250542778), ('78', 0.013425858250542778), ('75', 0.013425858250542778),
('95', 0.013425858250542778), ('37', 0.012626691943243626), ('194',
0.011639647485090233), ('359', 0.011639647485090233), ('285', 0.011120863746173888),
```

- PageRank



```
The page rank is
 [('61', 0.0028632188131482612), ('122', 0.002818091005547667), ('104', 0.0020501241861042275),
('212', 0.0015579756663930675), ('282', 0.0014777901462927566), ('185', 0.0014516335933814671),
('325', 0.0014347425785457917), ('348', 0.0013738085061384296), ('148', 0.0012034640016704966),
('96', 0.0011844341642238399), ('44', 0.000941236831193966), ('134', 0.0009086361067027147),
('94', 0.0008508225707519445), ('24', 0.0008501182812214506), ('287', 0.0008386895350124991),
('40', 0.0008386657333339617), ('43', 0.0007721660224499559), ('21', 0.000768058945809749),
('204', 0.0007512930621016642), ('454', 0.0007478522168039459), ('22', 0.0007467805644945037),
('327', 0.0007452379873701883), ('363', 0.0007375197297007083), ('277', 0.0007373225284821048),
('264', 0.0007321867470051605), ('433', 0.000718376205265767), ('301', 0.0007101486078173274),
('152', 0.0007058938981188416), ('386', 0.0007054760618733073), ('249', 0.0006989086440959973),
('381', 0.0006848147002367932), ('457', 0.0006740346801683613), ('413', 0.0006699155733780782),
('265', 0.0006633543695540123), ('372', 0.0006523783305094864), ('331', 0.0006260316650409096),
('326', 0.0006090287478981605), ('300', 0.0005800661918470575), ('291', 0.0005760683049296528),
('189', 0.0005760680911379564), ('141', 0.0005745202063971757), ('202', 0.000571426290213768),
('92', 0.0005630267573369101), ('444', 0.0005629363691149119), ('436', 0.0005629363691149119),
('273', 0.0005552519604448387), ('307', 0.0005551587211009654), ('216', 0.0005503464256768618),
('404', 0.0005409537663910288), ('468', 0.0005357508680488138), ('187', 0.000526782411453915),
('46', 0.000526782410623813), ('105', 0.0005239770220370778), ('167', 0.0005203601375519628),
('130', 0.000519346302882306), ('154', 0.0005181515166003628), ('439', 0.0005173673065840691),
('35', 0.000517106620206475), ('70', 0.0005144406919829635), ('166', 0.0005143347139115403),
('116', 0.0005143347139115403), ('114', 0.0005143347139115403), ('350', 0.0005143347098752686),
('259', 0.0005129824799015182), ('63', 0.0005129824790469524), ('160', 0.0005013932186543564),
('459', 0.0004995969095577149), ('191', 0.0004940296318888647), ('33', 0.000493996517725036),
('424', 0.000493996517725036), ('109', 0.0004931855087329666), ('253', 0.0004866166964638118),
('162', 0.00048464152452025585), ('469', 0.00048464152452025585), ('279', 0.0004844681470932042),
('411', 0.0004844641314987975), ('435', 0.0004844641314987975), ('332', 0.0004804013575135108),
('30', 0.0004804013575135108), ('60', 0.000477763978928851), ('126', 0.0004770918478533076),
('323', 0.000469950497478202), ('398', 0.00046684413033413303), ('380', 0.0004637608443875831),
('164', 0.00046051908191491306), ('278', 0.00046051908191491306), ('296', 0.00045919903942387186),
('219', 0.00045919903942387186), ('235', 0.00045919903942387186), ('336', 0.00045919903942387186),
('390', 0.000455349095758291), ('34', 0.0004499358188104666), ('330', 0.0004499358188104666),
```

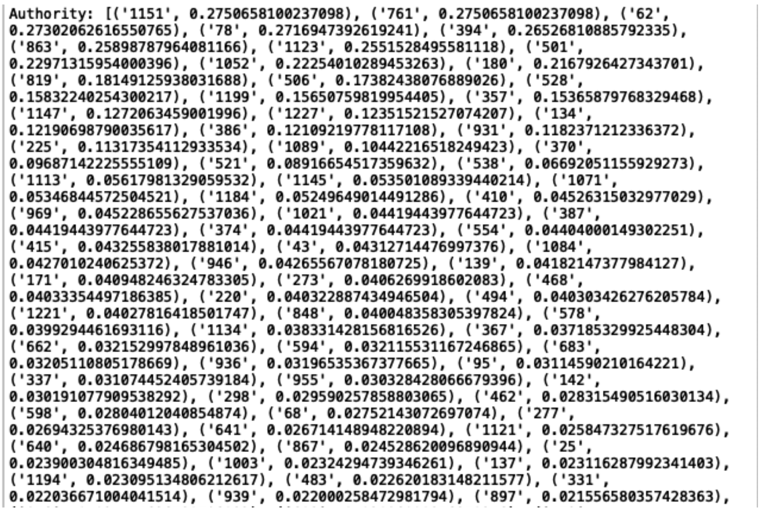
- SimRank



```
18 {'18': 1, '159': 0.08314051037625651, '15': 0.0, '164': 0.19978189251212417, '191': 0.0, '162': 0.0, '17': 0.0, '128': 0.05010998469998087, '123':
0.15305871388630007, '171': 0.031140087955359805, '182': 0.12581547064305681, '193': 0.1966230327025471, '156': 0.14625134541878498, '173': 0.0,
'188': 0.0, '109': 0.0, '135': 0.05010998469998087, '189': 0.0, '152': 0.0, '133': 0.13043278657968313, '185': 0.10823828381685062, '170': 0.0, '192':
0.0, '146': 0.0, '127': 0.0, '145': 0.15305871388630007, '121': 0.0, '125': 0.12581547064305681, '144': 0.0, '151': 0.15305871388630007, '136': 0.0,
'165': 0.15305871388630007, '112': 0.12581547064305681, '190': 0.0, '106': 0.0, '105': 0.0, '180': 0.0, '141': 0.11456078556816718, '184':
0.2556215502409963, '129': 0.0, '178': 0.20609961213127834, '149': 0.0, '140': 0.0, '198': 0.11527750232991611, '101': 0.0, '168': 0.0, '169': 0.0,
'197': 0.0, '158': 0.0, '161': 0.0, '157': 0.0, '187': 0.0, '108': 0.0, '120': 0.0, '160': 0.10720364872540089, '163': 0.0, '195': 0.0, '167': 0.0,
'199': 0.0, '134': 0.11549300484615294, '174': 0.13660177073625349, '14': 0.0, '150': 0.15305871388630007, '179': 0.12581547064305681, '111': 0.0,
'122': 0.11905837178001306, '1': 0.0, '139': 0.0, '19': 0.0, '103': 0.21811742777260015, '194': 0.09107091331846552, '12': 0.0, '143': 0.0, '104':
0.1089328675259002, '138': 0.0, '177': 0.0, '166': 0.0, '183': 0.0, '16': 0.15305871388630007, '153': 0.0, '117': 0.0, '100': 0.0, '176':
0.12581547064305681, '131': 0.0, '113': 0.0, '114': 0.0, '124': 0.0, '119': 0.0, '175': 0.15305871388630007, '130': 0.055915814906847834, '186':
0.12581547064305681, '110': 0.0, '118': 0.11527750232991611, '13': 0.0, '132': 0.0, '126': 0.0, '116': 0.0, '142': 0.09940654006324356, '102':
0.15305871388630007, '107': 0.0, '147': 0.05010998469998087, '155': 0.12581547064305681, '115': 0.12581547064305681, '196': 0.0, '181':
0.07614110604989723, '148': 0.11610672781657634, '10': 0.0, '137': 0.0, '11': 0.0, '154': 0.1324096354579564, '172': 0.0, '28': 0.0, '252': 0.0,
'285': 0.08322159545639381, '247': 0.14625134541878498, '226': 0.0, '225': 0.0, '219': 0.0, '281': 0.0, '24': 0.0, '200': 0.0, '259': 0.0, '272': 0.0,
'23': 0.0, '216': 0.18544255644799834, '256': 0.0, '22': 0.0, '266': 0.14625134541878498, '237': 0.0, '274': 0.09940654006324356, '2': 0.0, '267':
0.0, '210': 0.0, '260': 0.0, '276': 0.0, '235': 0.0, '236': 0.1381174277726002, '208': 0.15305871388630007, '291': 0.0, '288': 0.08314051037625651,
'27': 0.0, '284': 0.1381174277726002, '254': 0.12581547064305681, '207': 0.0, '20': 0.0, '251': 0.09940654006324356, '231': 0.15305871388630007,
'230': 0.0, '244': 0.0, '261': 0.1381174277726002, '264': 0.0, '293': 0.12581547064305681, '212': 0.10951106095168588, '26': 0.0, '255': 0.0, '213':
0.12581547064305681, '287': 0.0, '250': 0.15305871388630007, '238': 0.0, '234': 0.08314051037625651, '221': 0.0, '29': 0.0, '296': 0.0, '290': 0.0,
'269': 0.0, '209': 0.15305871388630007, '268': 0.15305871388630007, '295': 0.0, '218': 0.0, '229': 0.14625134541878498, '292': 0.0, '233':
0.03438604521490745, '257': 0.0, '203': 0.1381174277726002, '271': 0.0, '275': 0.15305871388630007, '249': 0.060872629493299336, '201': 0.0, '214':
0.15305871388630007, '286': 0.0, '263': 0.0, '243': 0.0, '270': 0.0, '232': 0.0, '278': 0.19978189251212417, '248': 0.0, '282': 0.12438998239302,
'223': 0.0, '297': 0.11480342144133814, '21': 0.0, '217': 0.0, '202': 0.0, '242': 0.0, '299': 0.1128403118883373, '205': 0.0, '215': 0.0, '240': 0.0,
'246': 0.0, '211': 0.0, '279': 0.0, '241': 0.0, '298': 0.0, '294': 0.0, '253': 0.0, '204': 0.0, '239': 0.0, '227': 0.0, '273': 0.0, '222': 0.0, '277':
0.0, '206': 0.0, '220': 0.12581547064305681, '245': 0.0, '262': 0.0, '283': 0.12581547064305681, '289': 0.0, '224': 0.10844319198076119, '228': 0.0,
'280': 0.12581547064305681, '25': 0.23358807082945018, '258': 0.0, '265': 0.0, '344': 0.0, '393': 0.0, '372': 0.0, '359': 0.09107091331846552, '364':
0.0, '316': 0.0, '308': 0.0, '307': 0.0, '347': 0.0, '350': 0.0, '330': 0.0, '365': 0.0, '341': 0.0, '351': 0.0, '363': 0.0, '300': 0.0, '303':
0.1455880708294501, '311': 0.05010998469998087, '324': 0.0, '306': 0.0, '38': 0.12581547064305681, '326': 0.1269309842669354, '378': 0.0, '337':
0.08314051037625651, '304': 0.0, '3': 0.0, '301': 0.0, '380': 0.0, '384': 0.0, '317': 0.05010998469998087, '392': 0.0, '302': 0.0, '382': 0.0, '397':
0.0, '325': 0.11126296476276791, '315': 0.09933983653848727, '328': 0.0, '385': 0.0, '35': 0.0, '373': 0.0, '375': 0.15305871388630007, '366': 0.0,
```

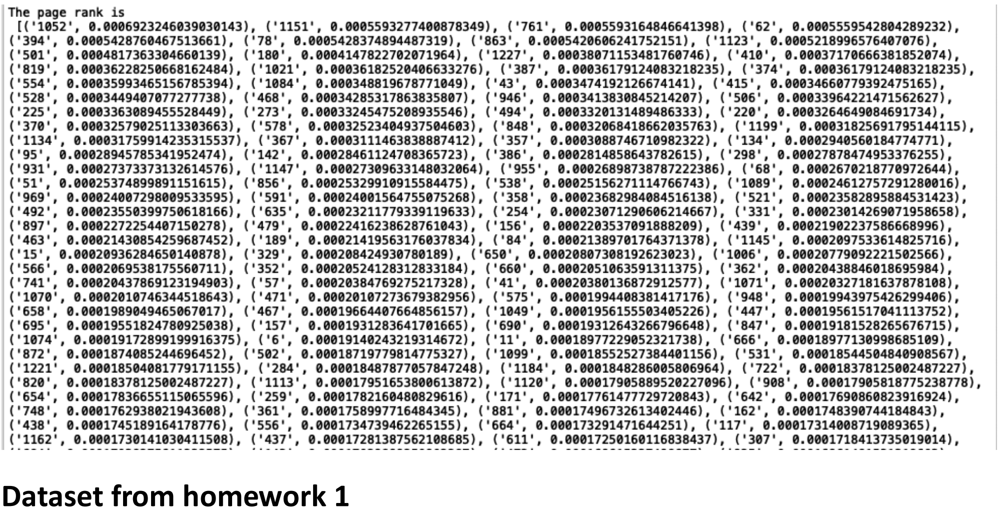
Graph_6:

- HITS



```
Authority: [('1151', 0.2750658100237098), ('761', 0.2750658100237098), ('62',
0.27302062616550765), ('78', 0.2716947392619241), ('394', 0.26526810885792335),
('863', 0.25898787964081166), ('1123', 0.2551528495581118), ('501',
0.22971315954000396), ('1052', 0.22254010289453263), ('180', 0.2167926427343701),
('819', 0.18149125938031688), ('506', 0.17382438076889026), ('528',
0.15832240254300217), ('1199', 0.15650759819954405), ('357', 0.15365879768329468),
('1147', 0.1272063459001996), ('1227', 0.12351521527074207), ('134',
0.12190698790035617), ('386', 0.12109219778117108), ('931', 0.1182371212336372),
('225', 0.11317354112933534), ('1089', 0.10442216518249423), ('370',
0.09687142225555109), ('521', 0.08916654517359632), ('538', 0.06692051155929273),
('1113', 0.05617981329059532), ('1145', 0.053501089339440214), ('1071',
0.05346844572504521), ('1184', 0.05249649014491286), ('410', 0.04526315032977029),
('969', 0.045228655627537036), ('1021', 0.04419443977644723), ('387',
0.04419443977644723), ('374', 0.04419443977644723), ('554', 0.04404000149302251),
('415', 0.043255838017881014), ('43', 0.04312714476997376), ('1084',
0.0427010240625372), ('946', 0.04265567078180725), ('139', 0.04182147377984127),
('171', 0.040948246324783305), ('273', 0.0406269918602083), ('468',
0.04033354497186385), ('220', 0.040322887434946504), ('494', 0.040303426276205784),
('1221', 0.04027816418501747), ('848', 0.040048358305397824), ('578',
0.0399294461693116), ('1134', 0.038331428156816526), ('367', 0.037185329925448304),
('662', 0.032152997848961036), ('594', 0.032115531167246865), ('683',
0.03205110805178669), ('936', 0.03196535367377665), ('95', 0.03114590210164221),
('337', 0.031074452405739184), ('955', 0.030328428066679396), ('142',
0.030191077909538292), ('298', 0.029590257858803065), ('462', 0.028315490516030134),
('598', 0.02804012040854874), ('68', 0.02752143072697074), ('277',
0.026943253769801 43), ('641', 0.026714148948220894), ('1121', 0.025847327517619676),
('640', 0.024686798165304502), ('867', 0.024528620096890944), ('25',
0.023900304816349485), ('1003', 0.02324294739346261), ('137', 0.023116287992341403),
('1194', 0.023095134806212617), ('483', 0.022620183148211577), ('331',
0.022036671004041514), ('939', 0.022000258472981794), ('897', 0.021556580357428363),
```

- PageRank

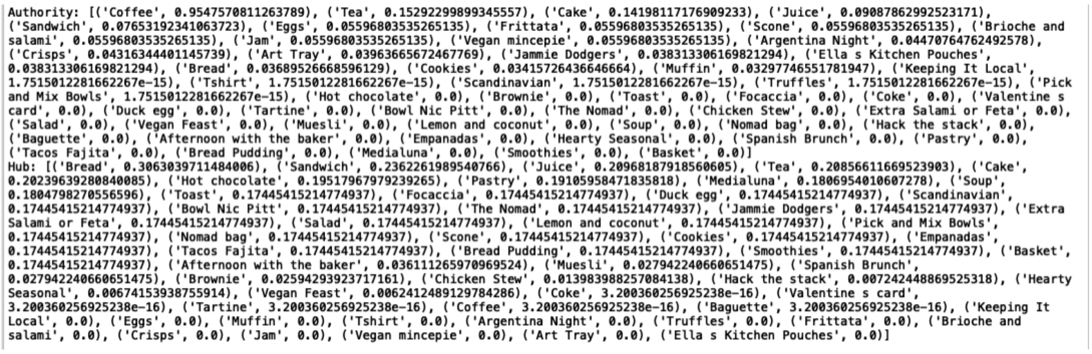


```
The page rank is
 [('1052', 0.0006923246039030143), ('1151', 0.0005593277400878349), ('761', 0.0005593164846641398), ('62', 0.0005559542804289232),
('394', 0.0005428760467513661), ('78', 0.0005428374894487319), ('863', 0.0005420606241752151), ('1123', 0.0005218996576407076),
('501', 0.0004817363304660139), ('180', 0.0004147822702071964), ('1227', 0.00038071153481760746), ('410', 0.00037170666381852074),
('819', 0.00036228250668162484), ('1021', 0.00036182520406633276), ('387', 0.00036179124083218235), ('374', 0.00036179124083218235),
('554', 0.00035993465156785394), ('1084', 0.000348819678771049), ('43', 0.0003474192126674141), ('415', 0.00034660779392475165),
('528', 0.0003449407077277738), ('468', 0.00034285317863835807), ('946', 0.0003413830845214207), ('506', 0.00033964221471562627),
('225', 0.0003363089455528449), ('273', 0.00033245475208935546), ('494', 0.0003320131489486333), ('220', 0.0003264649084691734),
('370', 0.00032579025113303663), ('578', 0.00032523404937504603), ('848', 0.00032068418662035763), ('1199', 0.00031825691795144115),
('1134', 0.00031759914235315537), ('367', 0.0003111463838887412), ('357', 0.0003088746710982322), ('134', 0.0002940560184774771),
('95', 0.00028945785341952474), ('142', 0.00028461124708365723), ('386', 0.0002814858643782615), ('298', 0.00027878474953376255),
('931', 0.00027373373132614576), ('1147', 0.00027309633148032064), ('955', 0.00026898738787222386), ('68', 0.0002670218770972644),
('51', 0.00025374899891151615), ('856', 0.00025329910915584475), ('538', 0.00025156271114766743), ('1089', 0.00024612757291280016),
('969', 0.00024007298009533595), ('591', 0.00024001564755075268), ('358', 0.00023682984084516138), ('521', 0.00023582895884531423),
('492', 0.00023550399750618166), ('635', 0.00023211779339119633), ('254', 0.00023071290606214667), ('331', 0.00023014269071958658),
('897', 0.0002272254407150278), ('479', 0.00022416238628761043), ('156', 0.0002203537091888209), ('439', 0.00021902237586668996),
('463', 0.00021430854259687452), ('189', 0.00021419563176037834), ('84', 0.00021389701764371378), ('1145', 0.0002097533614825716),
('15', 0.00020936284650140878), ('329', 0.000208424930780189), ('650', 0.00020807308192623023), ('1006', 0.00020779092221502566),
('566', 0.0002069538175560711), ('352', 0.00020524128312833184), ('660', 0.0002051063591311375), ('362', 0.00020438846018695984),
('741', 0.00020437869123194903), ('57', 0.00020384769275217328), ('41', 0.00020380136872912577), ('1071', 0.00020327181637878108),
('1070', 0.0002010746344518643), ('471', 0.00020107273679382956), ('575', 0.0001994408381417176), ('948', 0.00019943975426299406),
('658', 0.0001989049465067017), ('467', 0.00019664407664856157), ('1049', 0.0001956155503405226), ('447', 0.00019561517041113752),
('695', 0.00019551824780925038), ('157', 0.0001931283641701665), ('690', 0.00019312643266796648), ('847', 0.00019181528265676715),
('1074', 0.00019172899199916375), ('6', 0.00019140243219314672), ('11', 0.00018977229052321738), ('666', 0.00018977130998685109),
('872', 0.0001874085244696452), ('502', 0.00018719779814775327), ('1099', 0.00018552527384401156), ('531', 0.00018544504840908567),
('1221', 0.00018504081779171155), ('284', 0.00018487877057847248), ('1184', 0.0001848286005806964), ('722', 0.00018378125002487227),
('820', 0.00018378125002487227), ('1113', 0.00017951653800613872), ('1120', 0.00017905889520227096), ('908', 0.00017905818775238778),
('654', 0.00017836655115065596), ('259', 0.0001782160480829616), ('171', 0.00017761477729720843), ('642', 0.00017690860823916924),
('748', 0.0001762938021943608), ('361', 0.0001758997716484345), ('881', 0.00017496732613402446), ('162', 0.0001748390744184843),
('438', 0.0001745189164178776), ('556', 0.0001734739462265155), ('664', 0.000173291471644251), ('117', 0.00017314008719089365),
('1162', 0.0001730141030411508), ('437', 0.00017281387562108685), ('611', 0.00017250160116838437), ('307', 0.0001718413735019014),
```

Dataset from homework 1

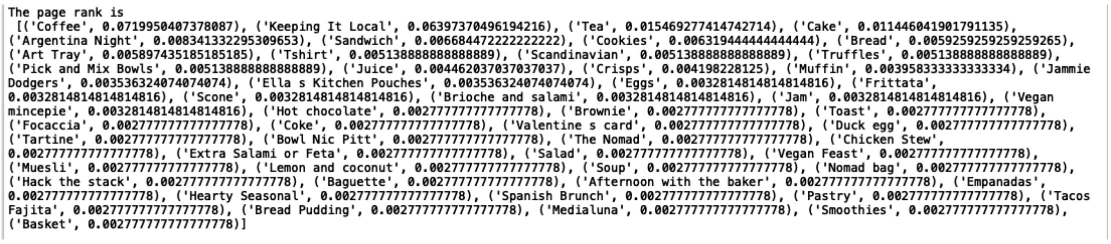
Kaggle_dataset:

- HITS



```
Authority: [('Coffee', 0.9547570811263789), ('Tea', 0.15292299899345557), ('Cake', 0.14198117176909233), ('Juice', 0.09087862992523171),
('Sandwich', 0.07653192341063723), ('Eggs', 0.05596803535265135), ('Frittata', 0.05596803535265135), ('Scone', 0.05596803535265135), ('Brioche and
salami', 0.05596803535265135), ('Jam', 0.05596803535265135), ('Vegan mincepie', 0.05596803535265135), ('Argentina Night', 0.04470764762492578),
('Crisps', 0.04316344401145739), ('Art Tray', 0.03963665672467769), ('Jammie Dodgers', 0.03831330616982l294), ('Ella s Kitchen Pouches',
0.03831330616982l294), ('Bread', 0.03689526668596129), ('Cookies', 0.03415726436646664), ('Muffin', 0.03297746551781947), ('Keeping It Local',
1.7515012281662267e-15), ('Tshirt', 1.7515012281662267e-15), ('Scandinavian', 1.7515012281662267e-15), ('Truffles', 1.7515012281662267e-15), ('Pick
and Mix Bowls', 1.7515012281662267e-15), ('Hot chocolate', 0.0), ('Brownie', 0.0), ('Toast', 0.0), ('Focaccia', 0.0), ('Coke', 0.0), ('Valentine s
card', 0.0), ('Duck egg', 0.0), ('Tartine', 0.0), ('Bowl Nic Pitt', 0.0), ('The Nomad', 0.0), ('Chicken Stew', 0.0), ('Extra Salami or Feta', 0.0),
('Salad', 0.0), ('Vegan Feast', 0.0), ('Muesli', 0.0), ('Lemon and coconut', 0.0), ('Soup', 0.0), ('Nomad bag', 0.0), ('Hack the stack', 0.0),
('Baguette', 0.0), ('Afternoon with the baker', 0.0), ('Empanadas', 0.0), ('Hearty Seasonal', 0.0), ('Spanish Brunch', 0.0), ('Pastry', 0.0),
('Tacos Fajita', 0.0), ('Bread Pudding', 0.0), ('Medialuna', 0.0), ('Smoothies', 0.0), ('Basket', 0.0)]
Hub: [('Bread', 0.3063039711484006), ('Sandwich', 0.2362261989540766), ('Juice', 0.20968187918560605), ('Tea', 0.20856611669523903), ('Cake',
0.20239639280840085), ('Hot chocolate', 0.19517967979239265), ('Pastry', 0.19105958471835818), ('Medialuna', 0.1806954010607278), ('Soup',
0.1804798270556596), ('Toast', 0.17445415214774937), ('Focaccia', 0.17445415214774937), ('Duck egg', 0.17445415214774937), ('Scandinavian',
0.17445415214774937), ('Bowl Nic Pitt', 0.17445415214774937), ('The Nomad', 0.17445415214774937), ('Jammie Dodgers', 0.17445415214774937), ('Extra
Salami or Feta', 0.17445415214774937), ('Salad', 0.17445415214774937), ('Lemon and coconut', 0.17445415214774937), ('Pick and Mix Bowls',
0.17445415214774937), ('Nomad bag', 0.17445415214774937), ('Scone', 0.17445415214774937), ('Cookies', 0.17445415214774937), ('Empanadas',
0.17445415214774937), ('Tacos Fajita', 0.17445415214774937), ('Bread Pudding', 0.17445415214774937), ('Smoothies', 0.17445415214774937), ('Basket',
0.17445415214774937), ('Afternoon with the baker', 0.036111265970969524), ('Muesli', 0.027942240660651475), ('Spanish Brunch',
0.027942240660651475), ('Brownie', 0.025942939237171 61), ('Chicken Stew', 0.013983988257084138), ('Hack the stack', 0.007242448869525318), ('Hearty
Seasonal', 0.00674153938755914), ('Vegan Feast', 0.0062412489129784286), ('Coke', 3.200360256925238e-16), ('Valentine s card',
3.200360256925238e-16), ('Tartine', 3.200360256925238e-16), ('Coffee', 3.200360256925238e-16), ('Baguette', 3.200360256925238e-16), ('Keeping It
Local', 0.0), ('Eggs', 0.0), ('Muffin', 0.0), ('Tshirt', 0.0), ('Argentina Night', 0.0), ('Truffles', 0.0), ('Frittata', 0.0), ('Brioche and
salami', 0.0), ('Crisps', 0.0), ('Jam', 0.0), ('Vegan mincepie', 0.0), ('Art Tray', 0.0), ('Ella s Kitchen Pouches', 0.0)]
```

- PageRank



```
The page rank is
 [('Coffee', 0.0719950407378087), ('Keeping It Local', 0.06397370496194216), ('Tea', 0.015469277414742714), ('Cake', 0.011446041901791135),
('Argentina Night', 0.008341332295309653), ('Sandwich', 0.006684472222222222), ('Cookies', 0.006319444444444444), ('Bread', 0.005925925925925925 9265),
('Art Tray', 0.005897435185185185), ('Tshirt', 0.005138888888888889), ('Scandinavian', 0.005138888888888889), ('Truffles', 0.005138888888888889),
('Pick and Mix Bowls', 0.005138888888888889), ('Juice', 0.004462037037037037), ('Crisps', 0.004198228125), ('Muffin', 0.003958333333333334), ('Jammie
Dodgers', 0.003536324074074074), ('Ella s Kitchen Pouches', 0.003536324074074074), ('Eggs', 0.0032814814814814816), ('Frittata',
0.0032814814814814816), ('Scone', 0.0032814814814814816), ('Brioche and salami', 0.0032814814814814816), ('Jam', 0.0032814814814814816), ('Vegan
mincepie', 0.0032814814814814816), ('Hot chocolate', 0.002777777777777778), ('Brownie', 0.002777777777777778), ('Toast', 0.002777777777777778),
('Focaccia', 0.002777777777777778), ('Coke', 0.002777777777777778), ('Valentine s card', 0.002777777777777778), ('Duck egg', 0.002777777777777778),
('Tartine', 0.002777777777777778), ('Bowl Nic Pitt', 0.002777777777777778), ('The Nomad', 0.002777777777777778), ('Chicken Stew',
0.002777777777777778), ('Extra Salami or Feta', 0.002777777777777778), ('Salad', 0.002777777777777778), ('Vegan Feast', 0.002777777777777778),
('Muesli', 0.002777777777777778), ('Lemon and coconut', 0.002777777777777778), ('Soup', 0.002777777777777778), ('Nomad bag', 0.002777777777777778),
('Hack the stack', 0.002777777777777778), ('Baguette', 0.002777777777777778), ('Afternoon with the baker', 0.002777777777777778), ('Empanadas',
0.002777777777777778), ('Hearty Seasonal', 0.002777777777777778), ('Spanish Brunch', 0.002777777777777778), ('Pastry', 0.002777777777777778), ('Tacos
Fajita', 0.002777777777777778), ('Bread Pudding', 0.002777777777777778), ('Medialuna', 0.002777777777777778), ('Smoothies', 0.002777777777777778),
('Basket', 0.002777777777777778)]
```

Discussion:

- Can link analysis algorithms really find the “important” pages from Web?

I think it is hard to get the important pages by using only one algorithm from web. For example, in graph_1 from authority and hub it is hard to see that Node 6 is the last reference but after we check the pagerank we can see the answer. In this way, we can see what website people stop at the end and it might be the important pages. In the other hand, we might find many pages which people just press by accident.

- Any new idea about the link analysis algorithm?

I think these three algorithms consider previous and next nodes. But in the real world, what if there is a malicious website which keep showing up then it will become an important webpage. Therefore, I think maybe there’s a way to calculate the bad website or low score nodes and split into real bad webpage and miss-calculated bad webpages then add as minus points in the algorithm so that we can have more confidential results.

- What is the effect of “C” parameter in SimRank?

C is like confidence level in the formula. It means how fast it decay.

- What you learned from this project?

In this chapter I learned many common approaches to link analysis and also implement three ways with examples. Even I got the answers there are still many relationships between data need to be observed. For hub and authority, it considers how many pages are linked to them or how many pages they link to. For pagerank, I think it considers the history status like after how many pages before it. And simrank considers who links to them. I think they cover different perspective. However, how to know that the pages we are calculating are all important pages will be another difficult works.