

Timothy Asp
Caleb Carlton
CPE 466 - Lab 7
Page Rank

Run Instructions:

For directed data, run: `python pageRank.py <datasource> directed`
For undirected data, run: `python pageRank.py <datasource> undirected`

Implementation: Generates a directed or undirected graph of the data, then runs the PageRank algorithm, iterating over every node checking the neighbors (undirected) and out-edges (directed). We used NetworkX (<http://networkx.lanl.gov/>) for our graphs which is a great python library for creating and maintaining these graphs. For D, which is the probability that the user will randomly go to a different page, we chose .85 which is what the book had as the probability for D. The graph is iterated over 10 times, to calculate the rank best.

State Borders

Analysis: The top ranked states are all high density area states, meaning they have lots of state borders, which makes sense and matches expected output.

MO	:0.0315971844476
KY	:0.0311331684216
TN	:0.0310935756102
MA	:0.0287725140929
PA	:0.0269173607527
MD	:0.0267222576413
GA	:0.0262810787353
NY	:0.0261652680999
SD	:0.0259332059378
WY	:0.0256935355929
IA	:0.0255032593163
CO	:0.0252932994762
VA	:0.0250814382815
ID	:0.0250734427792
AR	:0.023278072507
OK	:0.0223321781313
NV	:0.0216708088742
NB	:0.0214511655309
NE	:0.0214511655309
OH	:0.021122132435
WV	:0.02097603798
TX	:0.0203753104517
NH	:0.0203612457806
UT	:0.0201041909315
IL	:0.0194874330536
OR	:0.0186831020097
KS	:0.0182382396929
AL	:0.0180729949833
VT	:0.0180524967929
NC	:0.017964064708
CT	:0.0177488522188
AZ	:0.0175795169308
MS	:0.0171642829093
IN	:0.0170252621532
WI	:0.0167143241873
MT	:0.0164648690288
MN	:0.0164414868817
NM	:0.0163676118695
NJ	:0.0154237955317
DE	:0.0149027237957
CA	:0.014328872014
MI	:0.0137016010819
LA	:0.0133453258507

ND	:0.0130759232999
RI	:0.0128664771188
FL	:0.0104997265484
SC	:0.0104816930419
WA	:0.0104611056458
DC	:0.0102730766081
MV	:0.0100317563706
ME	:0.00871019610841

Political Blogs

855	:0.0123982365159
155	:0.0102073178341
963	:0.0085996555082
1051	:0.00779787367556
641	:0.00740159936356
55	:0.00708484611118
1245	:0.0059611211307
1153	:0.00569707706033
1437	:0.00562599374973
729	:0.00547134446998
1112	:0.00527363212751
1479	:0.00487418572979
1041	:0.00469773070457
454	:0.00469567338062
1101	:0.00464578664097
363	:0.00447834018621
323	:0.00435790083927
1000	:0.0038614062333
798	:0.00381999172434
642	:0.00370102086354
741	:0.0036547261639
979	:0.00362174829003
434	:0.0036180870679
1179	:0.00350056696695
493	:0.00346775923273
512	:0.0034349511101
880	:0.00342661996956
180	:0.00341641339595
996	:0.00340638815647
1270	:0.00340087077945
387	:0.00339655223689
1461	:0.00332533847476
878	:0.00324778609618
756	:0.0032359049905
1306	:0.00315517498621
144	:0.00312457924806
483	:0.00309610699993
535	:0.00308475291018
99	:0.00307278802084
1330	:0.00301907244338
189	:0.00301478979974
1209	:0.00300901159605
980	:0.00299508763393
170	:0.00298107564567
1463	:0.00283380952926
762	:0.00280848085965
856	:0.00279802241224
644	:0.00276932320703
1384	:0.00276752568959
547	:0.0027426663457
514	:0.00273565583007
72	:0.00273238941209
919	:0.00273076502953
664	:0.00272610535733
687	:0.00270819350753
150	:0.00269966673637
524	:0.00268026659323
826	:0.00266091439199
119	:0.00265691688035
297	:0.00264045909771
935	:0.00263408868513

892 :0.00261845274543
... trimmed

NCAA

Analysis: A quick look through the 2009 season results show that most of the teams ranked the highest we're the division leaders and winning teams.

Boise State	:0.00733262361219
Utah	:0.00722435930833
Montana	:0.00709947017054
Tulsa	:0.00670055195551
Ball State	:0.0063758368291
Texas	:0.00636546801443
Rice	:0.00636420852951
Alabama	:0.00625447664103
Florida	:0.00624166939591
TCU	:0.00612222447911
Texas Tech	:0.00602273860572
Penn State	:0.00583129342948
Cincinnati	:0.00575753270947
Oklahoma	:0.0056904527237
Richmond	:0.00547645775694
East Carolina	:0.00529388406925
Missouri	:0.00524749922454
Ohio State	:0.00522310227913
Brigham Young	:0.00514457541395
Weber State	:0.00513022880521
Nebraska	:0.00504668897937
USC	:0.00499134298746
New Hampshire	:0.00498805759706
Georgia	:0.00496366497189
Michigan State	:0.00492355290029
Iowa	:0.00491360116607
Western Michigan	:0.00488826180424
James Madison	:0.00488409341977
Houston	:0.00487147333886
Northwestern	:0.00485804367807
Troy	:0.00481302701871
Pittsburgh	:0.00480537805717
Louisiana Tech	:0.00476687389604
Air Force	:0.00472795714934
Oklahoma State	:0.00472453412691
Buffalo	:0.00469183049703
West Virginia	:0.00468412914986
Virginia Tech	:0.00467792833048
Villanova	:0.00461018702304
Mississippi	:0.00460722692705
LSU	:0.00458586681751
South Carolina State	:0.00457632203178
Navy	:0.00455469420055
Oregon	:0.00455031908469
Boston College	:0.00453703396878
Fresno State	:0.00444681070456
Rutgers	:0.00440682553538
Florida Atlantic	:0.00439888351631
Connecticut	:0.00437901660454
Florida A&M	:0.00437700673944
Kansas	:0.00437053073582
Southern Miss	:0.00434165382392
Colorado State	:0.00433892942647
Nevada	:0.00432803496027
South Florida	:0.00431247388258
Florida State	:0.00430781518633
Appalachian State	:0.0042855078864
Maine	:0.00427719960866
San Jose State	:0.00421017077414
Georgia Tech	:0.00417210375934
Central Michigan	:0.00416730758325
Wake Forest	:0.00404176127288
Minnesota	:0.00399261481722
Bethune-Cookman	:0.00395345951099

Louisiana-Lafayette	:0.00394932125891
Montana State	:0.00392668285699
California	:0.0038692041512
Wisconsin	:0.00385875981679
Memphis	:0.00382804688824
Maryland	:0.00380882019124
North Carolina	:0.0038077233882
Notre Dame	:0.00380022455306
Kentucky	:0.0037939751987
Vanderbilt	:0.00378971204221
Arizona	:0.0037893865312
Oregon State	:0.00377213824036
Cal Poly	:0.00374115053075
Hawaii	:0.0037386301266
Eastern Washington	:0.00369986078233
South Carolina	:0.00366865310273
Arkansas State	:0.0036312635554
Wofford	:0.00362089389909
Miami (FL)	:0.00361898361937
Northern Arizona	:0.00356914337631
Bowling Green	:0.00355155219836
UTEP	:0.0034868669963
William & Mary	:0.00345594533534
Colgate	:0.00343346141159
Northern Iowa	:0.00341207809157
Clemson	:0.00340172846317
Central Arkansas	:0.00339275926671
Kansas State	:0.00336454093013
Florida International	:0.00330978664396
Elon	:0.00330964240795
Morgan State	:0.00329645553639
Harvard	:0.00324203805793
Grambling State	:0.00319995545647

Les Miserables

Analysis: The main characters are ranked the highest, which is as expected

Valjean	:0.0758308901974
Myriel	:0.0429505779458
Gavroche	:0.0359160666477
Marius	:0.0310172618335
Javert	:0.0304815194381
Thenardier	:0.0280510981415
Fantine	:0.027191326266
Enjolras	:0.0219890272928
Cosette	:0.0207014337229
MmeThenardier	:0.0196084909871
Bossuet	:0.0190434224611
Courfeyrac	:0.0186615760305
Eponine	:0.0178862752016
Mabeuf	:0.0175458955297
Joly	:0.0172827540076
Bahorel	:0.0172689775792
Babet	:0.0167910676219
Gueulemer	:0.0167728629274
Claquesous	:0.0166377902786
MlleGillenormand	:0.0163128064923
Feuilly	:0.0159647347833
Combeferre	:0.0159634323963
Tholomyes	:0.0157308656902
Bamatabois	:0.0156663859666
Montparnasse	:0.0152477377714
Gillenormand	:0.0150159848507
Grantaire	:0.014524331992
Prouvaire	:0.0131993961441
Favourite	:0.0126897550776
Fameuil	:0.0126870828847
Dahlia	:0.0126820677133
Blacheville	:0.0126796581368
Zephine	:0.0126773324066
Listolier	:0.0126750883243

Cochepaille	:0.0125074398829
Chenildieu	:0.0125030179402
Judge	:0.0124920149995
Champpathieu	:0.0124884594437
Brevet	:0.0124851057688
Brujon	:0.0119298962133
Fauchelevant	:0.0116775047428
MmeHucheloup	:0.0107290724039
MlleBaptistine	:0.0103332200281
MmeMagloire	:0.0103247490448
Simplice	:0.00911030673281
LtGillenormand	:0.00874898560852
MmeBurgon	:0.00782036048117
Pontmercy	:0.0073937575872
Toussaint	:0.00686488659601
Woman2	:0.00686488659601
Anzelma	:0.00633806764306
MotherInnocent	:0.00621997328001
MmePontmercy	:0.00602616566625
Child1	:0.0058066429878
Child2	:0.00580537558791
Champfercier	:0.00560655179602
Geborand	:0.00560655179602
Napoleon	:0.00560655179602
Count	:0.00560655179602
OldMan	:0.00559885107345
Cravatte	:0.00559885107345
CountessDeLo	:0.00559885107345
Perpetue	:0.00542796568708
Magnon	:0.00528924458944
Marguerite	:0.0052793453439
Jondrette	:0.00527170515255
Woman1	:0.00526257949406
BaronessT	:0.00516084129531
Gribier	:0.0044295217059
MlleVaubois	:0.00393126768321
Scaufflaire	:0.00373850352216
Labarre	:0.00373850352216
Isabeau	:0.00373850352216
Gervais	:0.00373850352216
MmeDeR	:0.00373850352216
Boulatruelle	:0.00343826653682
MotherPlutarch	:0.00330387114808

Karate

Analysis: Looking through the data, the top people the most connections, thus making them the highest ranked, so it's as expected

34	:0.100779659381
1	:0.097005456951
33	:0.071576888392
3	:0.0570213307483
2	:0.0528547897183
32	:0.037117301531
4	:0.0358484502685
24	:0.0314579138303
9	:0.0297284975846
14	:0.0295173382004
7	:0.0291367362044
6	:0.0291334696727
30	:0.0262477118839
28	:0.0256009815829
31	:0.0245624539154
8	:0.0244823644259
5	:0.0219971434271
11	:0.0219885658829
25	:0.0210407262397
26	:0.0209763705902
20	:0.0195946521061
29	:0.0195468185624

17	:0.0167941834548
27	:0.0150190551952
13	:0.0146437100611
22	:0.0145596783316
18	:0.0145570208576
19	:0.0145207772694
21	:0.0145118570135
23	:0.0145118570135
15	:0.0145118570135
16	:0.0145118570135
10	:0.014293551594
12	:0.0095651796064

Dolphins

Grin	:0.032098769619
Jet	:0.0318806892333
Trigger	:0.0312662964391
Web	:0.0302903715199
SN4	:0.0298427560105
Topless	:0.0294801891168
Scabs	:0.0283966864331
Patchback	:0.0264319121028
Gallatin	:0.0263223854408
Beescratch	:0.0247553683412
Kringel	:0.0246260790508
SN63	:0.0239144788977
Feather	:0.0235998207139
SN9	:0.0219581962847
Upbang	:0.0217432490668
Stripes	:0.0216651111975
SN100	:0.0206241003106
DN21	:0.0201591774591
Haecksel	:0.019855361239
Jonah	:0.0193677489256
TR99	:0.019209987275
SN96	:0.0176136632619
TR77	:0.0173381999006
Number1	:0.0171934420037
Double	:0.0170864915312
Beak	:0.0169533238355
MN105	:0.0169216400508
MN83	:0.0168828868947
Hook	:0.0166108587996
SN90	:0.0162182764868
Shmuddel	:0.015898717753
DN63	:0.0156906891579
PL	:0.0153202288699
Fish	:0.0150996337083
Oscar	:0.0148566792738
Zap	:0.0147590831571
DN16	:0.0145192871451
Ripplefluke	:0.0133626577398
Bumper	:0.0133280233838
Knit	:0.0129680844496
Thumper	:0.0128173211756
TSN103	:0.0120603851553
Mus	:0.0115560295913
Notch	:0.0112467062498
Zipfel	:0.0110318043671
MN60	:0.00985621187565
CCL	:0.00962034429327
TR88	:0.00887000341
TR120	:0.00881850953712
Wave	:0.00836058683041
TSN83	:0.00817461979539
SN89	:0.00778446718663
Vau	:0.00748623884361
Zig	:0.00620544119831
Quasi	:0.0054370288389

MN23	:0.00543030882186
TR82	:0.00528011214892
Cross	:0.00507523069316
Five	:0.00507523069316
Whitetip	:0.0049602682216
SMN5	:0.00491419391454
Fork	:0.00483181375568

Conclusion

PageRank worked well for most of the datasets, but it worked best for datasets with Directed graphs. The results seemed less relevant for undirected graphs because each edge specifies a two way relation between two nodes. Directed graphs are more specific in that a link to a page does not necessarily imply that page links back. For this reason we conclude that pageRank is much more useful on directed graphs. Also, at first we only iterated over the graph once, but realizing that we need to recalibrate the links after every run, we ran it over the graph multiple times, which yielded much better results.