

466: Lab 3 PageRank

Daniel Wessel, Joseph Sarjeant

Implementation Overview

Our implementation of the PageRank algorithm follows the steps outlined in class. The graphs are first read into an adjacency matrix which represents all the edges from one node to another. This is implemented as a dictionary of nodes $[i]$ to list of nodes $[j]$ where each edge in the graph is represented as (i, j) . Given this description of the graph, the algorithm then calculates PageRank using an iterative approach. The initial weight of all nodes is $1/(\text{number of nodes})$. Then based on the previous PageRank values, it continually calculates a new set of values until the difference between the previous set and the current set is less than ϵ (.000001). Results are then returned to the user as a sorted array of tuples containing the node name and the PageRank value.

The user has the option to set d , ϵ , and the maximum number of iterations to run. Each of these are defined in the basic implementation of the PageRank algorithm. To improve performance and reduce memory usage, our implementation only remembers the current PageRank values (iteration i) and iteration $i-1$ PageRank values.

Results

Dataset: NCAA

Settings: $d = .85$, $\epsilon = .00001$

Output:

Team	PageRank		Team (Cont)	PageRank (Cont)
Mississippi	0.0309615373521		Fordham	0.00161719468011
Florida	0.0246997638762		UC Davis	0.0015995235808
Oklahoma	0.0158696560113		Cornell	0.00158612793728
Texas Tech	0.0155883211634		Robert Morris	0.00158489270592
Wake Forest	0.0151679891074		Florida Atlantic	0.00156102511426
Alabama	0.0138799937647		Duquesne	0.00155981791478
Virginia Tech	0.0135443028636		Jackson State	0.00151922402279
Oregon State	0.0135252870355		Southeast Missouri State	0.00151369740472
Texas	0.0129919266741		Arizona State	0.00151206389634
Vanderbilt	0.0121513105218		Portland State	0.00150259598091
Boston College	0.0114360091386		Western Illinois	0.00148836202089
James Madison	0.0114000744229		UCLA	0.00147187676694
Richmond	0.0111413811063		Virginia Military Institute	0.00144392997182
Georgia Tech	0.0111083519795		Louisiana-Lafayette	0.0014367533858
Montana	0.0106432373399		Baylor	0.00143513054573
USC	0.0105857231502		Norfolk State	0.0014337046469
Virginia	0.0104665507357		Wyoming	0.00142670950096
South Carolina	0.0103911072384		UTEP	0.0014262008833
North Carolina	0.00970769727851		Arkansas State	0.00141792963446
Florida State	0.0093324410085		Kent State	0.00141165834162
Duke	0.00915204276239		Marist	0.00139533584369
Maryland	0.00892178563477		Bucknell	0.00136509056162
Miami (FL)	0.00885567173275		San Jose State	0.00136134760756
North Carolina State	0.00881866289965		Georgia Southern	0.00133816735013

West Virginia	0.00863608196025		New Mexico	0.001329711792
Clemson	0.00828419828554		Northern Illinois	0.00132858929498
Weber State	0.00769220894527		Southern University	0.0013147677543
Georgia	0.00768320330079		Tennessee Tech	0.00128965858157
East Carolina	0.00767392005014		Hofstra	0.00128241207973
Villanova	0.00763576967633		Youngstown State	0.00128185505825
TCU	0.0075812087464		Kansas State	0.00125201384496
Pittsburgh	0.00721782224673		Michigan	0.00125106675808
Cincinnati	0.00712849781622		Eastern Michigan	0.00123879948927
Penn State	0.00706074338738		Western Carolina	0.00122412783127
LSU	0.00678759361165		Western Kentucky	0.00122204689738
Iowa	0.00665721880921		Valparaiso	0.0012213741187
Oregon	0.00575954231039		Texas Southern	0.00121254087011
California	0.0055591284632		Texas A&M	0.00120881325558
Tulsa	0.00534463918367		Florida International	0.00120544644462
Rutgers	0.00533778864829		Miami (OH)	0.00119231844437
Appalachian State	0.00516086811407		North Carolina Central	0.00119004731775
Connecticut	0.00494328052572		Purdue	0.00117072467923
Boise State	0.00494037115911		New Mexico State	0.00115505083135
Navy	0.00492306555852		Alabama State	0.0011453841961
Northwestern	0.00491565950353		Syracuse	0.00113924201622
Brown	0.00463166838755		UNLV	0.00112754137734
New Hampshire	0.00458069579335		Winston-Salem	0.00111645824197
Northern Iowa	0.00456723988707		Arkansas-Pine Bluff	0.00111308003694
Missouri	0.00443025014682		Eastern Illinois	0.00109975465843
Michigan State	0.00441591566904		Princeton	0.0010964929316
Houston	0.00433540951585		Louisiana-Monroe	0.00108907148154
Ohio State	0.00422579574623		Army	0.00107858479721
Stanford	0.00416657832087		UAB	0.0010676002905
Jacksonville	0.00410476717153		North Carolina A&T	0.00106031779936

Northwestern State	0.0041047536248		Towson	0.0010461985298
South Florida	0.00408408865266		Temple	0.00103251013508
Central Arkansas	0.00396173060857		Akron	0.00103105260171
Arkansas	0.0037849029592		Ohio	0.00102957652612
Nebraska	0.00370504710579		Utah State	0.00101440550373
Albany	0.00357072550114		Alabama A&M	0.000998450500479
Harvard	0.00349858544		Southern Methodist	0.000953793234517
Grambling State	0.00345891629872		Citadel	0.000952296284483
Southern Illinois	0.00345783922805		Mississippi Valley State	0.000945784543147
South Carolina State	0.00344089085243		Rhodes	0.000912422598297
Kentucky	0.00340340860845		Iowa State	0.000912370148312
Oklahoma State	0.003386770759		Missouri State	0.000900133874008
Colorado	0.00337981453119		Chattanooga	0.000883757961783
Dayton	0.0033617014365		Campbell	0.000883757961783
Tennessee	0.00330038594628		Wagner	0.000872612812717
Brigham Young	0.00326597503216		Illinois State	0.00086175848219
Massachusetts	0.00324672656644		Toledo	0.000814208293538
Mississippi State	0.00324071632141		Northeastern	0.000770742508752
Holy Cross	0.00321513182689		Idaho State	0.000764333420944
Kansas	0.00318601786107		Columbia	0.000743013559916
Ball State	0.00316271507045		Austin Peay	0.000733049731737
Nevada	0.00314879852297		Georgetown	0.000705076793903
Arizona	0.00314363951014		Washington State	0.000694002669566
Colgate	0.00314234364858		Alcorn State	0.00068332083979
Liberty	0.003077980542		Edward Waters	0.000681601757226
Hawaii	0.00306939617046		Howard	0.000681601757226
Nicholls State	0.00303491488119		San Diego State	0.000669520680214
Eastern Washington	0.00302441478569		Tulane	0.000667123099254
Maine	0.00301464623234		Idaho	0.000645857294519

Buffalo	0.00293003916109		Fayetteville State	0.000622212752097
Rice	0.00290931428309		Tuskegee	0.000598947424285
Lafayette	0.00288270437558		Henderson State	0.000582831232081
Southern Miss	0.00287927955428		North Texas	0.000581580992651
Wofford	0.00284920823465		Stonehill	0.000570422117722
Yale	0.00282302975982		Northern Colorado	0.000536769134377
Texas State	0.00280050722424		Assumption	0.000477707006369
McNeese State	0.00278895759283		Angelo State	0.000477707006369
William & Mary	0.00273198779641		St. Ambrose	0.000477707006369
San Diego	0.00269842050207		Dixie State	0.000477707006369
Bowling Green	0.00262456282911		Azusa Pacific	0.000477707006369
Florida A&M	0.0026037915358		Glenville State	0.000477707006369
Furman	0.00259194596742		Delaware Valley	0.000477707006369
Wisconsin	0.00258742781086		St. Francis (PA)	0.000477707006369
Presbyterian	0.00256074806095		Shaw	0.000477707006369
Monmouth	0.00255715054059		Western Oregon	0.000477707006369
Notre Dame	0.00252657787514		West Georgia	0.000477707006369
Illinois	0.00252140119569		Faulkner	0.000477707006369
Eastern Kentucky	0.00249563809319		Stillman	0.000477707006369
Middle Tennessee State	0.00248008796927		Livingstone	0.000477707006369
Tennessee State	0.0024793853583		Merrimack	0.000477707006369
Cal Poly	0.00247359440231		Lincoln (MO)	0.000477707006369
Air Force	0.00234285900378		Pace	0.000477707006369
Elon	0.00233889803557		Dartmouth	0.000477707006369
Stony Brook	0.0023259071978		Clark Atlanta	0.000477707006369
Southeastern Louisiana	0.00231262605033		Johnson Smith	0.000477707006369
North Dakota	0.00230028279994		Webber International	0.000477707006369
Rhode Island	0.00229024175656		Indiana State	0.000477707006369
Central Connecticut State	0.00228393278462		Texas A&M-Commerce	0.000477707006369

Colorado State	0.00228145012042		Culver-Stockton	0.000477707006369
Tennessee-Martin	0.00225503066388		Central Washington	0.000477707006369
North Dakota State	0.00224669770358		Southwest Baptist	0.000477707006369
Auburn	0.00222560227587		Langston	0.000477707006369
Sacred Heart	0.0021913627627		Shorter	0.000477707006369
Davidson	0.00216173777769		Lenoir-Rhyne	0.000477707006369
Minnesota	0.0021561876784		Humboldt State	0.000477707006369
Morehead State	0.00213751543817		Missouri S&T	0.000477707006369
Delaware	0.00213052744308		Central Methodist	0.000477707006369
Central Michigan	0.00209083030004		Kentucky Wesleyan	0.000477707006369
Fresno State	0.00207106587926		Delta State	0.000477707006369
Delaware State	0.00206215540053		Morehouse	0.000477707006369
Drake	0.00206001076134		Southern Virginia	0.000477707006369
Sacramento State	0.0020232452776		Tusculum	0.000477707006369
Bryant University	0.00202192912453		Chowan	0.000477707006369
Charleston Southern	0.00201645815398		Concordia University (WI)	0.000477707006369
South Dakota	0.00201373718593		Virginia State	0.000477707006369
Louisiana Tech	0.00201023148963		Upper Iowa	0.000477707006369
Western Michigan	0.00196565865992		Montana-Western	0.000477707006369
Louisville	0.00196073511848		Baker University	0.000477707006369
UCF	0.00194491806189		Western Washington	0.000477707006369
Butler	0.00194469038291		New Mexico Highlands	0.000477707006369
Lehigh	0.00193788833487		Southern Oregon	0.000477707006369
Morgan State	0.00192879333206		Albion	0.000477707006369
Marshall	0.00192793417187		East Central Oklahoma	0.000477707006369
Jacksonville State	0.00189352954637		Washington	0.000477707006369
Montana State	0.00188140187998		Concordia College	0.000477707006369
Prairie View A&M	0.00187808485949		William Penn	0.000477707006369
Indiana	0.00184017656723		Adams State	0.000477707006369
Bethune-Cookman	0.00181601462479		Wisconsin-La Crosse	0.000477707006369

Samford	0.00180238130794		North Greenville	0.000477707006369
Stephen F. Austin	0.0017975485919		Texas College	0.000477707006369
Southern Utah	0.00178182590724		St. Cloud State	0.000477707006369
Iona	0.00177466771726		Bentley College	0.000477707006369
Coastal Carolina	0.00177302686055		West Chester	0.000477707006369
Murray State	0.00176249126263		Southern Connecticut State	0.000477707006369
Memphis	0.00176026541409		Carthage	0.000477707006369
Troy	0.0017576925384		Mars Hill	0.000477707006369
Sam Houston State	0.00174976404507		Cumberland University	0.000477707006369
Gardner-Webb	0.00172990779159		Quincy	0.000477707006369
Hampton	0.00169809649698		Southeastern Oklahoma	0.000477707006369
Savannah State	0.00167913324235		Central State	0.000477707006369
South Dakota State	0.00167455518687		Benedict	0.000477707006369
Northern Arizona	0.00167389453884		Texas A&M-Kingsville	0.000477707006369
Pennsylvania	0.00165210348361		Lambuth	0.000477707006369

Observations: The PageRank resulted in a ranking where many of the Top 25 teams were among the top ranked teams. In this case, teams were ranked higher if they had a better record, where wins against high win teams had a higher effect on the rankings than a win against a lower tier team. One thing that could have improved the rankings is if the algorithm also took into consideration close versus blowout games, weighing blowouts higher than close games.

Metrics:

Iterations	54
Read Time	0:00:00.002505
Processing Time	0:00:00.552489

Dataset: States

Settings: $d = .85$ $\epsilon = .00001$

Output:

State	PageRank		State (cont)	PageRank (cont)
MA	0.0293609080234		AL	0.0176775119124
NY	0.0260865745076		IL	0.0174831598181
TN	0.0257964939787		NM	0.0171034315764
PA	0.0243949895586		NC	0.0170977026294
ID	0.0243592007782		MS	0.0170175125062
AR	0.0236463837935		IN	0.015565508318
MO	0.0235488682818		WI	0.0153798080789
KY	0.0229644644627		MT	0.0152198712632
OK	0.0221780598102		NJ	0.0149534326889
GA	0.0221327140221		MN	0.014938747679
VA	0.0218031319581		CA	0.0145070390471
NV	0.0216622318384		LA	0.014502422088
TX	0.0210591005446		DE	0.014123008056
NH	0.0209058076281		RI	0.0131774505931
MD	0.0198176939321		MI	0.0127641862412
UT	0.0192738258223		WV	0.0127330419086
OR	0.0187676370127		ND	0.0120964756542
WY	0.0187332505979		KS	0.0117890069096
CO	0.018706876499		FL	0.0105802646654
SD	0.0185402615401		WA	0.0105002418396
VT	0.0184106291885		SC	0.0104570549864
OH	0.0183945462299		DC	0.00951901634354
CT	0.0180877333473		ME	0.00898454395677
IA	0.0180383456496		NE	0.0030612244898
AZ	0.0177651769627			

Observations: Many of the top ranked states are those who are connected to other states that have a lot neighbor states. In cases like New York, a state that is only connected to 5 other states, it is ranked much higher than a state like Colorado, because its neighboring states are

also connected to a bunch of states. It also acts as a gatekeeper for the rest of New England, so all rankings and prestige has to flow through it.

Metrics:

Iterations	54
Read Time	0:00:00.000533
Processing Time	0:00:00.020292

Dataset: Karate

Settings: $d = .85$ $\epsilon = .00001$

Output:

<u>Member</u>	<u>PageRank</u>		<u>Member (Cont.)</u>	<u>PageRank</u>
34	0.100918936488		25	0.0210759859342
1	0.0969975109199		26	0.0210061457623
33	0.0716930348662		20	0.0196046506408
3	0.0570785011062		29	0.0195734323944
2	0.0528769833145		17	0.0167840856967
32	0.0371580307095		27	0.0150440034177
4	0.0358599121468		13	0.0146449205809
24	0.0315224311629		18	0.0145587046073
9	0.0297660359147		22	0.0145587046073
14	0.0295364804354		19	0.0145359662587
6	0.0291112955791		16	0.0145359662587
7	0.0291112955791		15	0.0145359662587
30	0.0262884674715		23	0.0145359662587
28	0.0256397154457		21	0.0145359662587
31	0.0245901304059		10	0.0143093833919
8	0.0244905344019		12	0.00956476343885
11	0.0219780461433			
5	0.0219780461433			

Observations: The more friends a person has, the higher their PageRank value. Ties are broken by those who are friends with people with a lot of friends.

Metrics:

Iterations	30
------------	----

Read Time	0:00:00.000271
Processing Time	0:00:00.006071

Dataset: Dolphins

Settings: d = .85 ϵ = .00001

Output:

<u>Dolphin</u>	<u>PageRank</u>		<u>Dolphin (Cont.)</u>	<u>PageRank (Cont.)</u>
Grin	0.0321443610957		DN63	0.0156430804984
Jet	0.0317283684801		PL	0.0153021039652
Trigger	0.0312992392116		Fish	0.0151083519088
Web	0.0300956026824		Oscar	0.0148457329295
SN4	0.0298752251962		Zap	0.0147678760514
Topless	0.0295140829026		DN16	0.014428162861
Scabs	0.0284229567182		Bumper	0.0133380395408
Patchback	0.0264584443319		Ripplefluke	0.0133087681747
Gallatin	0.026157096374		Knit	0.0129282506532
Beescratch	0.0246508417558		Thumper	0.0128307852747
Kringel	0.0246408428754		TSN103	0.012072547692
SN63	0.0239391538632		Mus	0.0115042903302
Feather	0.0234586768272		Notch	0.0112101987076
SN9	0.0219663088007		Zipfel	0.0110391572575
Stripes	0.021691041107		MN60	0.00986347151907
Upbang	0.0216510264811		CCL	0.00962903258121
SN100	0.020613373191		TR88	0.00887672401486
DN21	0.0200537931353		TR120	0.00882586994077
Haecksel	0.0198830079803		Wave	0.00832630420593
Jonah	0.0193954706703		TSN83	0.00818102375047
TR99	0.0192318699344		SN89	0.00776477424239
SN96	0.0176186058691		Vau	0.0074941511738
TR77	0.0173394834483		Zig	0.00619017828734
Number1	0.0171301883323		MN23	0.00541592781772

Double	0.0170982493277		Quasi	0.00541592781772
Beak	0.0169653387254		TR82	0.00526172222711
MN105	0.0169389197625		Cross	0.00507978792208
MN83	0.0169056858412		Five	0.00507978792208
Hook	0.0166267507942		Whitetip	0.00496288777414
SN90	0.0161376902039		SMN5	0.00491820566611
Shmuddel	0.0159198773727		Fork	0.00483530400041

Observations: Like the Karate dataset, Dolphins with the most associations were ranked higher than those with less associations. Those that associated themselves with other highly associative dolphins ended up with higher rankings than those that associated themselves with low association dolphins.

Metrics:

Iterations	54
Read Time	0:00:00.000533
Processing Time	0:00:00.020292

Dataset: Les Miserables

Settings: d = .85 ϵ = .00001

Output:

Character	PageRank		Character (cont)	PageRank (cont)
Valjean	0.0754302258377		Brujon	0.0118666500268
Myriel	0.0427794352555		Fauchelevant	0.011638055639
Gavroche	0.0357672482605		MmeHucheloup	0.0106897948472
Marius	0.0308948877057		MlleBaptistine	0.010277194684
Javert	0.0303027351708		MmeMagloire	0.010277194684
Thenardier	0.0279265079537		Simplice	0.00907364967458
Fantine	0.0270227036076		LtGillenormand	0.00871359278354
Enjolras	0.0218819819161		MmeBurgon	0.00780557285298
Cosette	0.0206112109513		Pontmercy	0.00736809167723
MmeThenardier	0.0195011279263		Woman2	0.00683686459215
Bossuet	0.0189594803129		Toussaint	0.00683686459215
Courfeyrac	0.0185783883716		Anzelma	0.00631353466764
Eponine	0.0177938908875		MotherInnocent	0.00620213104166
Mabeuf	0.0174779776753		MmePontmercy	0.00601013068824
Joly	0.0171998261916		Child2	0.00579124632079
Bahorel	0.0171998261916		Child1	0.00579124632079
Babet	0.0166918282753		Napoleon	0.00558433310127
Gueulemer	0.0166918282753		OldMan	0.00558433310127
Claguesous	0.0165610103296		CountessDeLo	0.00558433310127
MlleGillenormand	0.0162602034955		Geborand	0.00558433310127
Feuilly	0.0158920775992		Champtercier	0.00558433310127
Combeferre	0.0158920775992		Cravatte	0.00558433310127
Tholomyes	0.0156474170779		Count	0.00558433310127
Bamatabois	0.0155762882752		Perpetue	0.00540748938804
Montparnasse	0.0151709193692		Magnon	0.00527122134243
Gillenormand	0.0149574700877		Jondrette	0.00526541929201
Grantaire	0.0144566250256		Marguerite	0.00526033005641

Prouvaire	0.0131458410625		Woman1	0.0052441802215
Fameuil	0.012618195713		BaronessT	0.00514645519523
Blacheville	0.012618195713		Gribier	0.00442113943729
Dahlia	0.012618195713		MlleVaubois	0.00392250516571
Zephine	0.012618195713		Gervais	0.00372904355241
Listolier	0.012618195713		Scaufflaire	0.00372904355241
Favourite	0.012618195713		MmeDeR	0.00372904355241
Champmathieu	0.0124246821995		Labarre	0.00372904355241
Chenildieu	0.0124246821995		Isabeau	0.00372904355241
Cochepaille	0.0124246821995		Boulatruelle	0.00343164743867
Brevet	0.0124246821995		MotherPlutarch	0.00329862193547
Judge	0.0124246821995			

Observations: Interestingly, while the dataset for the most part had similar behavior to Karate and Dolphins, the character Myriel had only 10 associations in the whole book, but she ended up with the second highest rank.

Metrics:

Iterations	42
Read Time	0:00:00.000957
Processing Time	0:00:00.037883

Dataset: Political Blogs

Settings: d = .85 ϵ = .00001

Output:

ID	PageRank		ID (Cont)	PageRank (cont)
155	0.0134621593419		210	0.00285781386941
55	0.0114250475602		1330	0.00281835046535
1051	0.00947134438524		826	0.00274657106886
855	0.00937133775696		72	0.00270994561049
641	0.00932852341795		405	0.00269241905555
1153	0.00818483988972		687	0.00268591996385
963	0.00803589245098		644	0.00265382172838
729	0.00791181805002		483	0.0026458054234
1245	0.00670309125435		1429	0.00257636248792
323	0.00638995228272		810	0.0025636865174
1112	0.00636075041502		363	0.0025270928379
1461	0.0053811625633		23	0.00251752483421
1306	0.00525150474738		1000	0.00248118825446
1463	0.00512221913997		564	0.00246906484695
1179	0.00506988809935		664	0.00233510670281
1041	0.0050243781697		563	0.00233434781524
1437	0.00483464472713		726	0.00232541761891
535	0.00467320941817		119	0.0023103226728
180	0.00417685084368		490	0.00230696149563
642	0.00413040815797		754	0.0022420883677
756	0.0040332265523		27	0.00222096293753
301	0.00399468615958		1478	0.00221929954797
297	0.0038069802365		1232	0.00221473212106
1479	0.00353527540692		885	0.00219931973968
1159	0.00351967322304		99	0.00218086074169
1270	0.00347944499279		892	0.00213336234183
1293	0.00344908038872		189	0.00211445780351

878	0.00340957464669		396	0.00208166666918
1101	0.00331284194648		172	0.00205532326281
741	0.00330339563414		919	0.00201437847146
1317	0.00328081076834		1223	0.00199806700522
493	0.00318627704974		35	0.00199237278745
434	0.00311799722319		677	0.00197028824041
170	0.0030446072607		394	0.00196761851114
547	0.00299165814252		941	0.00196242668574
1209	0.00296664407161		1164	0.00196108206703
1055	0.00290178032808		441	0.00196093308847
623	0.00288095258977		1260	0.00193661971831
150	0.0028578981322			

Observations: Similar to Karate and Dolphins, those with more associations ended up with higher PageRank, especially if they were associated with other highly associated people.

Metrics:

Iterations	52
Read Time	0:00:00.021057
Processing Time	0:00:06.992229

Dataset: Wiki Vote

Settings: d = .85, ϵ = .00001

Output:

ID	PageRank		ID (Cont)	PageRank (Cont)
4037	0.00224023984247		1549	0.000603686323251
15	0.00178933698414		4666	0.0005893030168
6634	0.00174412633246		993	0.000583622050968
2398	0.00126845359785		3084	0.000575471582561
2237	0.0012139845892		3562	0.000572898363926
4191	0.00110274577877		5123	0.000566425919438
5254	0.00104548844875		2958	0.000563531859915
2328	0.000991592264497		2576	0.000554345648061
1297	0.000946167293592		922	0.000539763189476
4335	0.000941754322895		5022	0.000538735486184
6946	0.000879356284144		1855	0.000535590357149
3352	0.000867450928267		3897	0.000534415083473
2654	0.000859195613032		2651	0.000534253906039
762	0.000847120444127		4110	0.000533164588814
737	0.000845900049678		282	0.000522632710421
3089	0.000827605508132		4600	0.000520513573227
28	0.000821183459603		1031	0.000519907935572
2535	0.000810193386897		4687	0.000519543803422
3334	0.000805699645309		4536	0.000517618550653
665	0.000798939920379		2871	0.00051738206388
6774	0.000785239476816		2746	0.000517295385738
7092	0.000778794213173		3443	0.000516739883515
2565	0.000754262723921		1385	0.000515135766035
5484	0.000750296291042		3568	0.000511394831851
8042	0.000718105869222		5543	0.000510130742774
4310	0.000710899321837		4400	0.00050725062599
5423	0.000689322131134		2485	0.000500621632307

1211	0.000689299760584		3976	0.000492554193653
3456	0.000688954581928		3238	0.000481714786874
2657	0.000663826064955		2323	0.000479807821657
5404	0.000663225163847		6784	0.000477547488866
5233	0.000662170466033		3034	0.000476038449879
4712	0.00065249655531		3192	0.00047402903885
4828	0.000632432323699		2859	0.000471164368634
5079	0.000631720849807		6006	0.000470607464866
4261	0.000625198826223		6327	0.000470021886322
6914	0.000612923806057		825	0.000468696199428
3459	0.000611372063317		3898	0.000467178516451
2285	0.000603758094546		4099	0.000463788967573

Observations: Most of the top results are the ones that occur most often as both from nodes and to nodes in the graph. There are exceptions such as node 6634 (rank 2) which appear less frequently but are still high up in the graph because they are connected to many high ranking nodes.

Metrics:

Iterations	26
Read Time	0:00:00.121556
Processing Time	0:01:55.436720

Dataset: P2P-Gnutella05

Settings: d = .85, ϵ = .00001

Output:

ID	PageRank		ID (Cont)	PageRank (Cont)
1676	0.000594632613349		393	0.000281993881117
1020	0.000581917285492		854	0.000280421134888
386	0.000555532445409		1277	0.000279906324686
222	0.000550145326768		1364	0.000267752349206
227	0.000534748393854		1363	0.000261139580926
388	0.00052843001897		3078	0.000247209964448
688	0.000505900994616		509	0.000246858459965
226	0.000495643835612		1198	0.000245170523122
842	0.000494640910419		150	0.000237178250242
876	0.000489631657135		273	0.000235687659515
223	0.000457648017018		1226	0.000228567708453
31	0.000455226072277		2022	0.000217754942409
391	0.000450235563669		207	0.000209951733411
279	0.000441078919425		1069	0.000206113090043
271	0.000439360134774		686	0.000199646396843
225	0.000436539238615		2817	0.000198477364529
277	0.000435740030726		459	0.000192415778507
274	0.000431148595927		305	0.000185272333852
272	0.000426944075892		2432	0.000180452627924
887	0.000423608985667		1424	0.00017780971641
278	0.000419631040988		676	0.000174799411015
229	0.000418585365139		900	0.000170946455808
47	0.000397362205257		2329	0.000168812674548
541	0.000394654881513		947	0.000168549602693
221	0.000391117805768		5130	0.000167131642191
230	0.000389785723736		1113	0.000166677221707
385	0.000383696048151		2754	0.000166032603217

821	0.000376979831126		3538	0.000165458996193
276	0.000376450815505		3869	0.000164938434666
999	0.000371950098039		350	0.000163748128219
275	0.000365948251481		325	0.00016212171748
48	0.000341288182341		952	0.000161414332438
693	0.000338576351364		1777	0.000159303482075
387	0.000338214337479		3370	0.000158732087151
392	0.000321909505126		2663	0.000158232152178
224	0.000321045440131		2327	0.000157306795719
1086	0.000306351487718		3574	0.000156848062835
3079	0.000305856595448		2493	0.000156360356007
841	0.000287199011295		38	0.000154774722856

Observations: The top ten in the page rank list had very similar occurrences in the initial graph (between 80 and 90). The only exception is node 388 which has only 69 occurrences. This is because 388 receives prestige from other very prestigious nodes.

Metrics:

Iterations	16
Read Time	0:00:00.037654
Processing Time	0:00:26.123625

Dataset: SlashdotZoo

Settings: d = .85, ϵ = .00001

Output:

ID	PageRank		ID (cont)	PageRank (Cont)
75	0.00325271881106		38568	0.000375301506663
43	0.00318665700119		10451	0.000372371898225
749	0.00292657480737		51	0.000366495460423
184	0.00193417747129		875	0.000363257736676
163	0.00114293966498		743	0.000350307266892
1810	0.00101662322912		4720	0.00034794091268
57	0.00094101475211		27	0.000344803217921
34	0.000919553624982		44	0.000340542778847
74	0.000885743799455		32889	0.00033721423147
15	0.000866538130735		2354	0.000336309785555
53	0.000852354436866		170	0.000332657684625
50	0.000841662767528		1455	0.000331835176917
1808	0.000834473980376		106	0.000328709873774
1832	0.000754372321289		950	0.000322611592055
877	0.000684793625318		81	0.000322446245904
3335	0.00067720456636		190	0.000320321202995
1116	0.000645042546628		2	0.000318292573821
1240	0.000622969053155		2066	0.000306544962063
1397	0.000598730626596		1730	0.000305313055575
13382	0.000551869392368		2935	0.000300377337913
47	0.000536405250578		1358	0.000296232356724
945	0.000529691507802		1556	0.000296115468545
46	0.000516450777448		1052	0.000291457521401
17	0.000487565762048		1070	0.000288993125421
1491	0.00047811584754		618	0.000288838919838
670	0.000450326469988		0	0.000288106941206
1838	0.000436186182967		1686	0.000288106404328

1850	0.000433554667175		981	0.000287933835695
1981	0.000431472670902		52	0.000286094154768
523	0.000430287737445		676	0.000284764391267
1803	0.000426311152207		4057	0.00028128860559
165	0.000416691127287		45	0.000280936544201
1300	0.000410492451483		1265	0.00027967774207
2516	0.000401094055385		1342	0.000274994063291
2113	0.000393442284956		738	0.000274755195084
522	0.000392806440255		813	0.000273977270097
1887	0.000388639052708		809	0.000271505187579
1807	0.000383735695868		1785	0.000270245790457

Observations: This data set is fairly different from the previous two because the number of occurrences in the graph drops off significantly as you go down the pagerank list. The node with the highest pagerank has 2546 occurrences, the fifth highest pagerank has 894 occurrences. This is shown in the pagerank as the drop over that same interval is from $\sim .003$ to $\sim .001$.

Metrics:

Iterations	55
Read Time	0:00:00.578758
Processing Time	6:36:52.268528

Dataset: Amazon Product

Settings: $d = .85$, $\epsilon = .00001$

Output: We left it running for 25 hours and it didn't converge. Unfortunately we were anticipating it finishing so we weren't printing out intermittent statistics so we don't know how close it was to converging.

Dataset: LiveJournal

Settings: $d = .85$, $\epsilon = .00001$

Output: A single iteration of LiveJournal did not complete after 6 hours so we decided to not try to run it.

Overall Summary

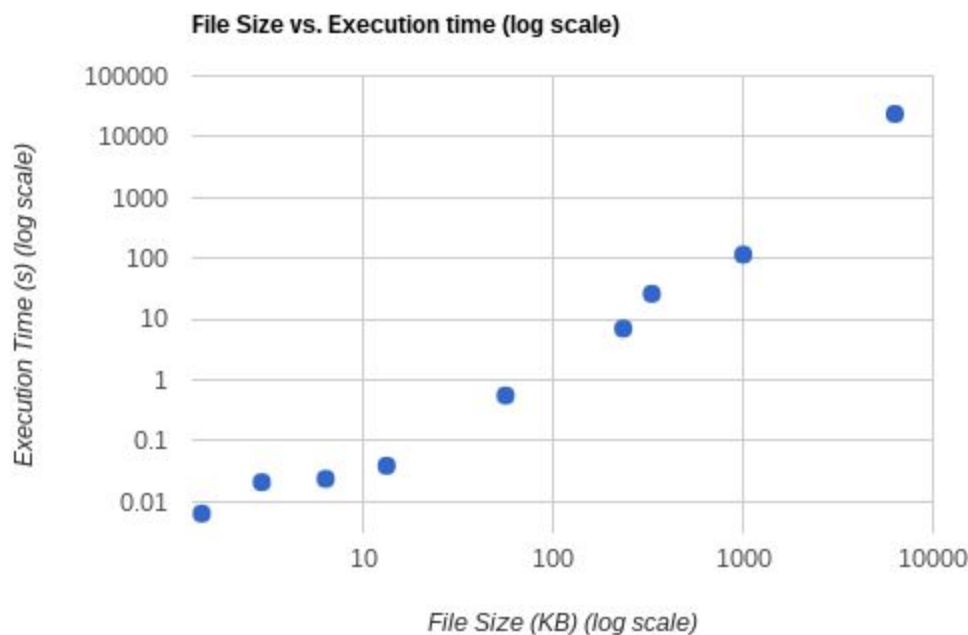
For the most part the nodes that appear in the graph (regardless of as the from node or the to node) have a higher pagerank. This obviously has some exceptions in cases where the node has few edges, but the connected nodes are connected to many other nodes. This skews the page rank to be higher (e.g. New York in stateborders).

Weighted graphs also did not come out to be exactly as expected. If we were to take into account weighted edges it would remove some of the skew towards giving nodes with many edges a higher pagerank because certain nodes could have few highly weighted edges and their pagerank would reflect that fact. For example in the NCAA graph, it would have been beneficial to take into account the score differential between teams. This would have granted teams who outscored their opponents by more a higher pagerank than teams who only won by a small margin. Also, in the Slashdot-zoo graph we would have accounted for the 'sign' of the edge and adjust the pagerank accordingly. This would cause node 1810 (pagerank #6) which has many edges of sign -1 to have a lower pagerank and possibly show up lower in the ranking.

Performance Evaluation

Due to our implementation of the PageRank algorithm, we found that our system had an exponential behavior as the file size increased. While our system could handle the smaller files easily, the runtime and the memory usage of the files increased rapidly on the SNAP datasets, as exhibited by the following graph. For the largest two files, Amazon and LiveJournal, our system ran at close to 8 GB of RAM and close to max CPU usage for over 24 hours without converging. The LiveJournal dataset appeared to not even complete one round of the PageRank calculations in the time that it ran. As a result, we were unable to complete the calculations for the last two datasets.

Evidently, our system could have benefited from both parallelization and a better language choice. Python is not optimized for high performance issues such as this, so a more powerful language would have been better suited.



*Graph does not include the two largest data sets because we could not finish processing them.

Appendix

Files

main.py : The driver of the file.

Graph.py : Builds an adjacency list from a graph file. (.txt or .csv)

pageRank.py : Given an adjacency list from Graph.py compute the PageRank for each node in the graph.

Flags

-p prints the Adjacency list after it has been constructed.

-d the provided file is a directed graph

Usage

python main.py filename.txt filename2.csv ... filenameN.txt

This will create a graph from the provided files and compute the pagerank of the constructed graph. Each graph's pagerank will be computed and output into a text file with -result appended to it.

e.g. filename.txt -> filename-result.txt