# Lab 7 - Page Rank

Trent Ellingsen & Thomas Wong
CPE-466-01 / CPE-466-02

## Implementation Overview:

We decided to go with an adjacency list to store the data. We used a hash table to represent the adjacency list with key having a list of outgoing nodes. We utilized a recursive definition of the following page rank algorithm.

$$pageRank(i) = (1-d) \cdot \frac{1}{|V|} + d \cdot \sum_{k=1}^{s} \frac{1}{|O_{j_k}|} \cdot pageRank(j_k).$$

To keep from infinitely recursing we kept track of how many links it would follow. After experimentation we found that limiting it to 100 was a good choice.

## Results

### State Borders

A) We ran it with pageRank stateborders.csv

B) Output

```
1 "ME" 0.008871292491901554
2 "NH" 0.00623729438837153
3 "RI" 0.00591706916539193
4 "FL" 0.0058882356745345804
5 "DC" 0.005697061561219844
6 "SC" 0.005684715410389589
7 "WA" 0.005663497393283831
8 "MV" 0.0055976997908010065
9 "DE" 0.00527457227770489
10 "NJ" 0.00527457227770489
11 "CA" 0.0051747327219218535
12 "MI" 0.005142089917146441
13 "ND" 0.005132387072717734
14 "VT" 0.005116622832758942
15 "CT" 0.005116622832758942
16 "LA" 0.005078602571821814
17 "AL" 0.005044808409702031
18 "OR" 0.004901572691965027
19 "AZ" 0.004901572691965027
20 "MS" 0.004875490386286898
21 "NM" 0.004846735815951193
22 "IN" 0.004779619498532601
23 "WI" 0.004779619498532601
24 "NC" 0.004773730254214402
25 "MT" 0.004763121245661523
26 "MN" 0.0047433697684129
27 "MA" 0.004684474724980301
28 "NY" 0.004684474724980301
29 "UT" 0.004649937672368798
30 "NV" 0.004649937672368798
31 "OH" 0.004630351989503551
32 "WV" 0.004606379627553485
33 "TX" 0.0045922595823087746
34 "KS" 0.004572393928634246
35 "PA" 0.0045686586878720525
36 "MD" 0.0045686586878720525
37 "IL" 0.004567244834338163
38 "AR" 0.004542431408901367
39 "GA" 0.00453790862353767
40 "ID" 0.004485061756524572
41 "OK" 0.004470673834930514
42 "NB" 0.004454119123535074
43 "NE" 0.004454119123535074
44 "VA" 0.004441063328959812
45 "CO" 0.0044187349352119015
46 "IA" 0.004386673326892578
47 "WY" 0.004369637119892808
48 "TN" 0.004369117742163056
49 "SD" 0.004365959195395606
50 "KY" 0.004321182298285908
51 "MO" 0.004268030922633414
```

C) We do not believe that this is accurate because ME is only directly visited by one other page, yet it is #1 in our list.

## NCAA Football

A) We ran it with pageRank NCAA_football.csv flag

B) Output

```
 1 "Northern Colorado" 0.0019668622699745 13
 2 "Wesley College" 0.0017416061762123222
 3 "North Texas" 0.0016645446042382543
 4 "Stonehill" 0.001647656898999498
 5 "Henderson State" 0.0016080555449166831
 6 "Arkansas-Monticello" 0.0016080555449166831
 7 "St. Francis (IL)" 0.0015897623269108393
 8 "Marian" 0.0015897623269108393
 9 "Fayetteville State" 0.001562503566298497
 10 "Southern Methodist" 0.0015459961554299992
 11 "Howard" 0.001539859026374712
 12 "Edward Waters" 0.001539859026374712
 13 "Washburn" 0.0015394624806800384
 14 "Tuskegee" 0.001522131845513619
 15 "Franklin" 0.0014947700613687303
 16 "Idaho State" 0.0014910722540155833
 17 "Rhodes" 0.001469906232567846
 18 "Georgetown" 0.0013877846490894467
 19 "Birmingham Southern" 0.0013874648437499997
 20 "Methodist" 0.0013874648437499997
 21 "Idaho" 0.0013681967599872566
 22 "San Diego State" 0.0013188622491348834
 23 "Iona" 0.0012105540799655322
 24 "Alcorn State" 0.0011887933974583415
 25 "Winston-Salem" 0.0011817001827262978
 26 "Austin Peay" 0.0011815917087521209
 27 "Purdue" 0.0011643788506973834
 28 "Washington State" 0.00115660750254888
 29 "Tulane" 0.0011540403316496793
 30 "Columbia" 0.001153438337145135
 31 "North Carolina Pembroke" 0.0011474219408359394
 32 "Northeastern" 0.0011406624930343499
 33 "Utah State" 0.0011357519160899222
 34 "Miami (OH)" 0.0011226810448223374
 35 "Iowa State" 0.0011225182601964944
 36 "Western Kentucky" 0.0011145869915793953
 37 "Wagner" 0.0010935562254041071
 38 "North Carolina A&T" 0.001085832998909739
 ...
```

C) This one is not accurate since Chattanooga is listed as last and Northern Colorado is listed first but both appear in the losing brackets.

## Karate

A) We ran it with pageRank karate.csv

B) Output

```
1 12 0.010702278277374936
2 27 0.008686816781654156
3 17 0.00861311250562092
4 10 0.008317518531660132
5 19 0.008289313458214178
6 16 0.008289313458214178
7 15 0.008289313458214178
8 23 0.008289313458214178
9 21 0.008289313458214178
10 18 0.00824819796221688
11 13 0.00824819796221688
12 22 0.00824819796221688
13 25 0.007848441714411538
14 26 0.007674934249138413
15 29 0.007476384903459695
16 11 0.007430171190497527
17 5 0.007430171190497527
18 20 0.007430171190497527
19 30 0.007204730098618516
20 31 0.0070667660143209875
21 28 0.007055818089359477
22 24 0.007026712087470451
23 8 0.00702115780463785
24 7 0.00702115780463785
25 6 0.00702115780463785
26 14 0.006775749773122045
27 9 0.006775749773122045
28 32 0.006612144418778174
29 4 0.006612144418778174
30 2 0.006339468828205057
31 3 0.006284933710090434
32 33 0.0062146844611590405
33 34 0.006114172180459822
34 1 0.006112279738874364
```

## Dolphins

A) We ran it with pageRank dolphins.csv

B) Output

```
1 "Zig" 0.006508118092486805
2 "Fork" 0.006123752146561403
3 "Five" 0.006123752146561403
4 "Cross" 0.006123752146561403
5 "TR82" 0.006008678259434428
6 "Whitetip"
0.006004848671276325
7 "SMN5" 0.005998898451769422
8 "MN23" 0.0059816484990582
9 "Quasi" 0.0059816484990582
10 "Wave" 0.00474536484919868
11 "TR88" 0.004673922485408157
12 "Vau" 0.004622728745368304
13 "SN89" 0.004615642698703314
14 "TSN83" 0.004610845554093384
15 "TR120" 0.004610845554093384
16 "Zipfel" 0.004230417118396959
17 "CCL" 0.004177560180046976
18 "Ripplefluke"
0.0041478191004334945
19 "MN60" 0.004136234917415112
20 "Notch" 0.004121733435600881
21 "Mus" 0.004112162402911873
22 "Thumper"
0.003967167677507396
23 "Bumper" 0.003933989154015877
24 "DN16" 0.0039052191640347985
25 "Knit" 0.003885654915410338
26 "TSN103"
0.0038647458168614536
27 "Zap" 0.0038483355197542536
28 "DN63" 0.0038014555342536093
29 "PL" 0.0037879331136398387
30 "Fish" 0.0037687163097243747
```

C) Although Zig shows up twice and the last one Grin shows up multiple times, Grin is last and Zig is first.

## Lesmis

A) We ran it with pageRank lesmis.csv

B) Output

```
1 "Jondrette"                          11 "MlleVaubois"
0.006807704788284647                   0.0048777035267345715
2 "MmeBurgon"                          12 "MotherPlutarch"
0.005291735030937844                   0.004769998827022276
3 "Gribier" 0.005042185694701694       13 "Boulatruelle"
4 "Champtercier"                       0.004723946971042979
0.004930813416711781                   14 "MmeDeR" 0.004696999789233911
5 "Count" 0.004930813416711781         15 "Isabeau"
6 "Cravatte"                           0.004696999789233911
0.004930813416711781                   16 "Scaufflaire"
7 "Geborand"                           0.004696999789233911
0.004930813416711781                   17 "Labarre"
8 "OldMan" 0.004930813416711781        0.004696999789233911
9 "CountessDeLo"                       18 "Gervais"
0.004930813416711781                   0.004696999789233911
10 "Napoleon"                          19 "MotherInnocent"
0.004930813416711781                   0.003800313626571626
```

C) Jondrette shows up only twice, while Valjean shows up many times. But Jondrette is first and Valjean is last.

## Polblogs

A) We ran it with pageRank polblogs.csv flag

B) Output

```
1 325 9.389671361502346E-4          16 364 3.8780240233694677E-4
2 1159 9.389671361502346E-4         17 820 3.8259492711944967E-4
3 1293 9.389671361502346E-4         18 382 3.777234038523594E-4
4 1260 9.389671361502346E-4         19 1455 3.760304591254712E-4
5 1259 9.389671361502346E-4         20 908 3.744139339696812E-4
6 821 4.922002522985143E-4          21 766 3.734741784037558E-4
7 1291 4.8692898449934964E-4        22 208 3.6426616381977667E-4
8 757 4.6026125680554016E-4         23 1340 3.6416096139231676E-4
9 400 4.555015081928903E-4          24 926 3.6416096139231676E-4
10 380 4.2482245735866596E-4        25 1405 3.6416096139231676E-4
11 1490 4.054552269379823E-4        26 1279 3.6416096139231676E-4
12 33 3.961316468289802E-4          27 1247 3.6416096139231676E-4
13 424 3.955813637773306E-4         28 230 3.6416096139231676E-4
14 858 3.9361738831360726E-4        29 216 3.6416096139231676E-4
15 235 3.8780240233694677E-4        30 833 3.6416096139231676E-4
```

## Overall Summary

The idea of a good page rank would be one that had a lot of links pointing to and from the page.  For our algorithm however, it did not produce good rankings for the data sets since some data only showed up twice, yet they were the top ranked.  We know that the probability of a page to be visited is increased when there are more links to it.  It worked better with undirected graphs versus directed ones.

## Appendix

README