CSC 691X Big Data

Project 3

Shuowen Wei

Part 1.

Top 10 page ranks are:

[[ 0.00040765]

[ 0.00057528]

[ 0.00057768]

[ 0.00094609]

[ 0.00121355]

[ 0.0012561 ]

[ 0.00161726]

[ 0.00173313]

[ 0.00190796]

[ 0.00334305]]

And the associate URLs are:

212 : http//spectacle.berkeley.edu/

3636 : http//rsv.ricoh.com/

4796 : http//www.scag.org

719 : http//www.students.ucr.edu/

1490 : http//www.uci.edu/

1489 : http//www.ucdavis.edu/

2409 : http//www.lib.uci.edu/

8052 : http//vision.berkeley.edu/VSP/index.shtml

998 : http//www.gene.com/ae/bioforum/

4392 : http//search.ucdavis.edu/

There is no Beta involved in computing this result, and there is not much in common in the names of the top 10 ranked pages.

Part 2a.

With Beta = 0.9, we have the top 10 page ranks are:

[[ 0.00089026]

[ 0.0009179 ]

[ 0.0009325 ]

[ 0.00096858]

[ 0.00100568]

[ 0.00105221]

[ 0.00111812]

[ 0.00118382]

[ 0.00186157]

[ 0.00187683]]

And the associate URLs are:

1 : http//www.berkeley.edu/

6428 : http//home.netscape.com/comprod/mirror/index.html

998 : http//www.gene.com/ae/bioforum/

4824 : http//www.berkeley.edu

18 : http//www.calacademy.org/

67 : http//www.california.edu/

2409 : http//www.lib.uci.edu/

1490 : http//www.uci.edu/

4392 : http//search.ucdavis.edu/

1489 : http//www.ucdavis.edu/

Part 2b.

With Beta = 0.8, we have the top 10 page ranks are:

[[ 0.00099161]

[ 0.00107774]

[ 0.00117408]

[ 0.0013259 ]

[ 0.00142046]

[ 0.00144726]

[ 0.00147669]

[ 0.00148769]

[ 0.00154249]

[ 0.00154345]]

And the associate URLs are:

1490 : http//www.uci.edu/

1807 : http//www.yahoo.com/

1618 : http//www.ca.gov

1 : http//www.berkeley.edu/

4824 : http//www.berkeley.edu

67 : http//www.california.edu/

2079 : http//www.linkexchange.com/

4392 : http//search.ucdavis.edu/

6428 : http//home.netscape.com/comprod/mirror/index.html

1489 : http//www.ucdavis.edu/

Some observations:

With different Beta values, we can see that the ten web pages with the highest ranking are basically the same, but their order somehow changes a lot.

Part 3.

When we use the other 9 URLs as trusted pages, and compute the span mass of the highest ranked web page we got -32.37. And this means that the highest ranked web page can also be trusted since its spam mass value is negative.