Steps of this assignment:

Step 1.

- Download and install Solr on laptop.
- Copy all the html files for Boston Globe, and index these pages using Solr.
- Create a website on my laptop with Apache Server that can deliver web pages, and run scripts.
- Write a PHP program as solr-php-client that accepts input from the user in a HTML form, and sends requests to the Solr server. (Here I modified example code from https://github.com/PTCInc/solr-php-client)
- Search in the webpage, and get top 10 result for each query.

Step 2.

- In order to use PageRank as ranking algorithm in Solr, we need to create an external file that stores rank.
- Loop each webpage in directory, and extract outgoing links using JSoup Library.(Here
 I referred code from http://www-scf.usc.edu/~csci572/2017Fall/hw4/ExtractingLinks.pdf)
- Use Boston_Global_Map.csv to create two map files, fileUrlMap and urlFileMap, and filter out resource outside the news website domain and those are not html files.
- Create a directed graph using the mapping file and extracted URLs. Vertices in the graph are webpage files and detected edge between two files represent a link between two. (Here I used edgeList format, when creating edges.)
- Load the web graph to NetworkX, and compute PageRank using parameters: "alpha=0.85, personalization=None, max_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None".
- Make the document id the same as the one presented in index.
- Name the pageRank scores file "external_pageRankFile", and put this file under the data folder of the core I created.
- There are 2 modifications of configuration files. First, add the field in the managed-schema which refers to this score. Second, define 2 listeners within the <query> element in solrconfig.xml file.
- Reload the index, or create a new core and configure it as above and leave the former one only for Lucene method.

Step 3.

- Compare the top 10 results for each query of the two ranking algorithms.

Why some pages have higher PageRank values than others?

Answer: The PageRank score of a webpage is high when it is important. The more importance the page has, the more webpages have link to it. That is why the results via PageRank are more likely to be https://www.bostonglobe.com or category pages. These pages have more in-links than others and gain a higher PageRank value.

Flow Screenshots (query = Elon Musk)

Initial page



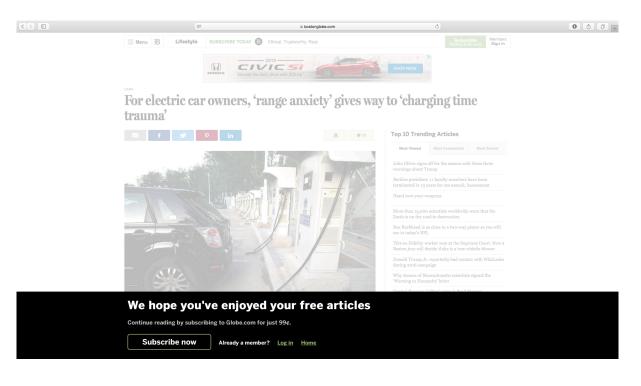
Result for Lucene(Default)



Result of PageRank



Click url https://www.bostonglobe.com/lifestyle/2017/10/05/for-electric-car-owners-range-anxiety-gives-way-charging-time-trauma/BbVjIWtQjSk0hl9nDOK8NI/story.html



URLs of Top 10 Results

Query = Elon Musk

#	Results of Lucene	Results of PageRank
	https://www.bostonglobe.com/sports/bruins/2017/10/05/youn	https://www.bostonglobe.com/2014/05/04/tracking-dorm-
1	gsters-lift-bruins-victory/fsKD9FFTnXE1LQQXAS6kBM/story.html	shortfall/umq1VQjQWYrbDCH7hcOFlN/story.html
		https://www.bostonglobe.com/business/2017/09/27/jeff-
	https://www.bostonglobe.com/sports/bruins/2017/10/05/youn	bezos-has-never-lived-worked-here-but-knows-lot-about-
2	gsters-lift-bruins-victory/fsKD9FFTnXE1LQQXAS6kBM/story.html	boston/yKb9pR9zfb1SMd1MnTTTdJ/story.html
		https://www.bostonglobe.com/sports/bruins/2017/10/05/y
	https://www.bostonglobe.com/2016/09/16/freightfarms/MVvd	oungsters-lift-bruins-
3	KOWwrMSAof8cmHOliN/story.html	victory/fsKD9FFTnXE1LQQXAS6kBM/story.html
	https://www.bostonglobe.com/news/nation/2017/09/08/com	https://www.bostonglobe.com/sports/2015/06/27/boston-
	merce-secretary-ceos-were-wrong-quit-trump-councils-after-	globe-baseball-all-
4	charlottesville/0tBTfkswTtE7nflNfVNapI/story.html	scholastics/EGDePA2zBCfkT7R6kxEWAP/story.html
	https://www.bostonglobe.com/lifestyle/2017/10/05/for-	
	electric-car-owners-range-anxiety-gives-way-charging-time-	https://www.bostonglobe.com/2016/09/16/freightfarms/M
5	trauma/BbVjIWtQjSk0hl9nDOK8NI/story.html	VvdKOWwrMSAof8cmHOliN/story.html
	https://www.bostonglobe.com/lifestyle/2017/10/05/for-	https://www.bostonglobe.com/sports/high-
	electric-car-owners-range-anxiety-gives-way-charging-time-	schools/2017/06/30/boston-globe-girls-tennis-all-
6	trauma/BbVjIWtQjSk0hl9nDOK8NI/story.html	scholastics/GTzxcWey1r9d0RskjtoN5L/story.html
	https://www.bostonglobe.com/news/nation/2017/08/15/some	https://www.bostonglobe.com/sports/high-
	-corporations-have-had-enough-	schools/2017/06/30/boston-globe-preps-all-
7	trump/o2JljN2DbdkuaQYSx6HRjI/story.html	scholastics/mfVPQtq2ey9fTHuPHQec9L/story.html
	https://www.bostonglobe.com/business/2017/09/27/jeff-	https://www.bostonglobe.com/lifestyle/2017/10/05/for-
	bezos-has-never-lived-worked-here-but-knows-lot-about-	electric-car-owners-range-anxiety-gives-way-charging-time-
8	boston/yKb9pR9zfb1SMd1MnTTTdJ/story.html	trauma/BbVjIWtQjSk0hl9nDOK8NI/story.html
	https://www.bostonglobe.com/news/nation/2017/08/15/some	https://www.bostonglobe.com/news/nation/2017/08/15/s
	-corporations-have-had-enough-	ome-corporations-have-had-enough-
9	trump/o2JljN2DbdkuaQYSx6HRjI/story.html	trump/o2JljN2DbdkuaQYSx6HRjl/story.html
	https://www.bostonglobe.com/business/2017/09/27/jeff-	https://www.bostonglobe.com/sports/high-
	bezos-has-never-lived-worked-here-but-knows-lot-about-	schools/2017/06/30/boston-globe-girls-tennis-all-
10	boston/yKb9pR9zfb1SMd1MnTTTdJ/story.html	scholastics/GTzxcWey1r9d0RskjtoN5L/story.html

Query = Star Wars

# F		
# 1	Results of Lucene	Results of PageRank
	https://www.bostonglobe.com/arts/movies/2015/12/15/the	
_	-force-with-star-wars-the-force-	
1 2	awakens/KyOQ1W4zeMyVDncrlirp5M/story.html	https://www.bostonglobe.com
<u> </u>	https://www.bostonglobe.com/arts/movies/2015/12/15/the	
_	-force-with-star-wars-the-force-	
2 2	awakens/KyOQ1W4zeMyVDncrlirp5M/story.html	https://www.bostonglobe.com
<u> </u>	https://www.bostonglobe.com/metro/2015/12/17/first-line-	
<u>f</u>	for-the-force-fan-stakes-out-spot-for-latest-star-wars-	
3 <u>f</u>	film/QBYM4R1QwbCoA7c7cBs8wO/story.html	https://www.bostonglobe.com/sports
<u> </u>	https://www.bostonglobe.com/lifestyle/style/2015/12/18/fo	
<u>r</u>	r-star-wars-die-hards-hour-marathon-daunting-but-	
4 6	enjoyable-task/W08Q3d2TeXWCQzDalZaZsI/story.html	https://www.bostonglobe.com
ŀ	https://www.bostonglobe.com/arts/2015/12/12/starwarsge	
5 <u>r</u>	nerations/PLDMidc6USG536AfUaPA7N/story.html	https://www.bostonglobe.com/business
<u>ł</u>	https://www.bostonglobe.com/lifestyle/style/2015/12/12/st	
<u>a</u>	ar-wars-sneakers-not-just-for-	
6 <u>l</u>	kids/XnHidQOeClvVNY94HfptQJ/story.html	https://www.bostonglobe.com/arts
<u> </u>	https://www.bostonglobe.com/metro/2015/12/17/charlie-	
<u>k</u>	baker-not-fan-new-star-wars-	
	flicks/UX6q3i6wY5F6AluydftuTI/story.html	https://www.bostonglobe.com/business
<u> </u>	https://www.bostonglobe.com/arts/movies/2015/12/12/get	
<u>t</u>	ting-latest-scoop-scoops-star-	
8 \	wars/1CDNVgBUSzNeG9FPW50yJP/story.html	https://www.bostonglobe.com
<u> </u>	https://www.bostonglobe.com/arts/movies/2015/12/12/get	
<u>t</u>	ting-latest-scoop-scoops-star-	
9 \	wars/1CDNVgBUSzNeG9FPW50yJP/story.html	https://www.bostonglobe.com/arts
<u>I</u>	https://www.bostonglobe.com/arts/movies/2015/12/12/osc	
<u> </u>	ar-isaac-talks-about-acting-new-star-wars-	
10 <u>r</u>	movie/Z85IkTprnFbJz9wT7PoygM/story.html	https://www.bostonglobe.com/sports

Query = North Korea

#	Results of Lucene	Results of PageRank
	https://www.bostonglobe.com/news/world/2017/09/08/kor	
	ea-braces-for-another-possible-korea-missile-	
1	test/wrfYUFe7CEgGvKbcg41ZbN/story.html	_
	https://www.bostonglobe.com/news/world/2017/08/09/tru	
	mp-threatens-fire-and-fury-against-north-korea-	
2	endangers/sSaFxbsoTVeQ3syE8gJvgO/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/news/world/2017/08/09/tru	
	mp-threatens-fire-and-fury-against-north-korea-	
3	endangers/sSaFxbsoTVeQ3syE8gJvgO/story.html	https://www.bostonglobe.com/metro
	https://www.bostonglobe.com/opinion/2017/08/09/plan-for-	
	north-korea-before-this-spirals-out-	
4	control/Q23bbp1BbDPlTxdvYxv3LK/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/opinion/2017/08/09/plan-for-	
	north-korea-before-this-spirals-out-	
5	control/Q23bbp1BbDPlTxdvYxv3LK/story.html	https://www.bostonglobe.com/lifestyle/crossword
	https://www.bostonglobe.com/news/world/2017/09/25/nort	
	h-korea-minister-says-can-shoot-down-	
6	bombers/8TtjEtXdimYcIXEArVHu5K/story.html	https://www.bostonglobe.com/opinion
	https://www.bostonglobe.com/news/politics/2017/09/26/no	
	rth-korea-seeks-help-from-republican-analysts-what-with-	
7	trump/Iro675CGFxHP3wqY4ZbzhK/story.html	https://www.bostonglobe.com/sports
	https://www.bostonglobe.com/news/politics/2017/09/26/no	
	rth-korea-seeks-help-from-republican-analysts-what-with-	
8	trump/Iro675CGFxHP3wqY4ZbzhK/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/news/nation/2017/09/03/tru	
	mp-says-north-korea-very-hostile-and-	
9	dangerous/SoR0T7tl6i3MdNpwHonu9M/story.html	https://www.bostonglobe.com/lifestyle
	https://www.bostonglobe.com/news/nation/2017/09/03/tru	
	mp-says-north-korea-very-hostile-and-	
10	dangerous/SoR0T7tl6i3MdNpwHonu9M/story.html	https://www.bostonglobe.com/lifestyle

Query = LA Dodgers

ш	Results of Lucene	Possible of Possible
#		Results of PageRank
	https://www.bostonglobe.com/sports/2017/10/07/mlb-	
	boxscore-dodgers-	
1	diamondbacks/TAdQkAqY6OqncGtJjZyG8O/story.html	https://www.bostonglobe.com/arts
_	https://www.bostonglobe.com/sports/2012/08/26/gonzo/Av	
2	zuwi4ZrRfOBdQKd1UFGP/story.html	https://www.bostonglobe.com/arts
	https://www.bostonglobe.com/lifestyle/names/2017/09/28/	
	carl-crawford-boston-carried-hate-for-that-city-for-long-	
3	time/SKvbZfN6QwarHEy24f965H/story.html	https://www.bostonglobe.com/lifestyle/names
	https://www.bostonglobe.com/lifestyle/names/2017/09/28/	
	carl-crawford-boston-carried-hate-for-that-city-for-long-	
4	time/SKvbZfN6QwarHEy24f965H/story.html	https://www.bostonglobe.com/lifestyle/names
		https://www.bostonglobe.com/arts/2017/10/06/seuss-
		controversy-illuminates-cultural-
5		clash/9bg6jN1x3Dzsk5yT1dv8XP/story.html
		https://www.bostonglobe.com/arts/2017/10/06/seuss-
		controversy-illuminates-cultural-
6		clash/9bg6jN1x3Dzsk5yT1dv8XP/story.html
7		https://www.bostonglobe.com/arts
	https://www.bostonglobe.com/sports/2017/09/30/baseball-	https://www.bostonglobe.com/news/bigpicture/2017/04/12/
8	standings/G6K9mjYu6AokJmy36PkhcL/story.html	animal-expressions/jqzgvgqvk5GqncG2shUP1O/story.html
	https://www.bostonglobe.com/sports/redsox/2017/09/16/in	animal-expressions/jqzgvgqvk5dqftcdzsff0f 10/stofy.fttfff
	dians-have-central-title-	
9	grasp/gBRKvUTjvx5LirE6i6HviO/story.html	https://www.bostonglobe.com/arts
3	https://www.bostonglobe.com/sports/redsox/2017/10/04/di	ittps.//www.bostofigione.com/arts
		https://www.hostonglobo.com/nows/highisture/2017/04/12/
10	amondbacks-advance-after-wild-one-over-	https://www.bostonglobe.com/news/bigpicture/2017/04/12/
10	rockies/ZXhz3EJeboHWpmrFA4UcRP/story.html	animal-expressions/jqzgvgqvk5GqncG2shUP1O/story.html

Query = Puerto Rico

Ì		
#	Results of Lucene	Results of PageRank
	https://www.bostonglobe.com/news/politics/2017/10/03/six	
	-things-trump-probably-shouldn-have-said-about-puerto-rico-	
1	while-puerto-rico/z2A6pU5lozLPkJdMzSY3HP/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/news/politics/2017/10/03/six	
	-things-trump-probably-shouldn-have-said-about-puerto-rico-	
2	while-puerto-rico/z2A6pU5lozLPkJdMzSY3HP/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/news/politics/2017/10/03/six	
	-things-trump-probably-shouldn-have-said-about-puerto-rico-	
3	while-puerto-rico/z2A6pU5lozLPkJdMzSY3HP/story.html	https://www.bostonglobe.com/business
	https://www.bostonglobe.com/news/politics/2017/10/03/six	
	-things-trump-probably-shouldn-have-said-about-puerto-rico-	
4	while-puerto-rico/z2A6pU5lozLPkJdMzSY3HP/story.html	https://www.bostonglobe.com/business
	https://www.bostonglobe.com/news/politics/2017/10/03/six	
	-things-trump-probably-shouldn-have-said-about-puerto-rico-	
5	while-puerto-rico/z2A6pU5lozLPkJdMzSY3HP/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/news/politics/2017/10/03/six	
	-things-trump-probably-shouldn-have-said-about-puerto-rico-	
6	while-puerto-rico/z2A6pU5lozLPkJdMzSY3HP/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/opinion/editorials/2017/09/2	
	6/where-urgency-puerto-rico-	
7	response/oFxAJIj9duwxiaxVGK0LIL/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/opinion/editorials/2017/09/2	
	6/where-urgency-puerto-rico-	
8	response/oFxAJIj9duwxiaxVGK0LIL/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/opinion/editorials/2017/09/2	
	3/don-let-puerto-rico-	
9	languish/k7iPpkMGa54CDFd65Iwh4H/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/opinion/editorials/2017/09/2	
	3/don-let-puerto-rico-	
10	languish/k7iPpkMGa54CDFd65Iwh4H/story.html	https://www.bostonglobe.com

Query = Hurricane Harvey

	query – Hurricane Harvey		
#	Results of Lucene	Results of PageRank	
	https://www.bostonglobe.com/news/nation/2017/09/08/ho		
	<u>use-seals-deal-hurricane-</u>	https://www.bostonglobe.com/1970/01/13/terms-	
1	relief/mgMI2SyD8kUIb4ewoFSBII/story.html	service/XpUbEbX2aS9pkwUa7aOKTI/story.html	
	https://www.bostonglobe.com/news/bigpicture/2017/08/28/		
	hurricane-harvey-devastates-southeast-	https://www.bostonglobe.com/1970/01/05/term-	
2	texas/ZWm7Ln0xffqQqpglzffYxL/story.html	purchase/RfO4zG3drUPKl8BRZNKlDP/story.html	
	https://www.bostonglobe.com/news/bigpicture/2017/08/28/		
	hurricane-harvey-devastates-southeast-	https://www.bostonglobe.com/1970/01/13/privacy-	
3	texas/ZWm7Ln0xffqQqpglzffYxL/story.html	policy/LZx4MmGtzRGBg8CkDCladI/story.html	
	https://www.bostonglobe.com/news/bigpicture/2017/08/28/		
	hurricane-harvey-devastates-southeast-		
4	texas/ZWm7Ln0xffqQqpglzffYxL/story.html	https://www.bostonglobe.com/newsletters	
	https://www.bostonglobe.com/news/bigpicture/2017/08/28/		
	hurricane-harvey-devastates-southeast-		
5	texas/ZWm7Ln0xffqQqpglzffYxL/story.html	https://www.bostonglobe.com/tools/help/stafflist	
	https://www.bostonglobe.com/news/bigpicture/2017/08/28/		
	hurricane-harvey-devastates-southeast-		
6	texas/ZWm7Ln0xffqQqpglzffYxL/story.html	https://www.bostonglobe.com/tools/help/stafflist	
	https://www.bostonglobe.com/news/bigpicture/2017/08/28/		
	hurricane-harvey-devastates-southeast-		
7	texas/ZWm7Ln0xffqQqpglzffYxL/story.html	https://www.bostonglobe.com/newsletters	
	https://www.bostonglobe.com/news/bigpicture/2017/08/28/		
	hurricane-harvey-devastates-southeast-	https://www.bostonglobe.com/1970/01/05/term-	
8	texas/ZWm7Ln0xffqQqpglzffYxL/story.html	purchase/RfO4zG3drUPKl8BRZNKlDP/story.html	
	https://www.bostonglobe.com/news/bigpicture/2017/08/28/		
	hurricane-harvey-devastates-southeast-	https://www.bostonglobe.com/1970/01/13/terms-	
9	texas/ZWm7Ln0xffqQqpglzffYxL/story.html	service/XpUbEbX2aS9pkwUa7aOKTI/story.html	
	https://www.bostonglobe.com/news/bigpicture/2017/08/28/		
	hurricane-harvey-devastates-southeast-	https://www.bostonglobe.com/1970/01/13/privacy-	
10	texas/ZWm7Ln0xffqQqpglzffYxL/story.html	policy/LZx4MmGtzRGBg8CkDCladI/story.html	

Query = iPhone X

	<u>, </u>	
#	Results of Lucene	Results of PageRank
	https://www.bostonglobe.com/business/2017/09/21/should-	
	you-queue-for-new-iphone-wait-	
1	for/MZLL85K3mhl2iGVYuy8ofN/story.html	_
	https://www.bostonglobe.com/business/2017/09/21/should-	
	you-queue-for-new-iphone-wait-	
2	for/MZLL85K3mhl2iGVYuy8ofN/story.html	_
	https://www.bostonglobe.com/ideas/2017/09/12/apple-	
	world-iphone-doesn-need-look-like-	https://www.bostonglobe.com/1970/01/13/terms-
3	iphone/C8EFHm2BFe6eV9oeCudAKO/story.html	service/XpUbEbX2aS9pkwUa7aOKTI/story.html
	https://www.bostonglobe.com/ideas/2017/09/12/apple-	
	world-iphone-doesn-need-look-like-	https://www.bostonglobe.com/1970/01/05/term-
4	iphone/C8EFHm2BFe6eV9oeCudAKO/story.html	purchase/RfO4zG3drUPKl8BRZNKlDP/story.html
	https://www.bostonglobe.com/ideas/2017/09/12/apple-	
	world-iphone-doesn-need-look-like-	https://www.bostonglobe.com/1970/01/13/privacy-
5	iphone/C8EFHm2BFe6eV9oeCudAKO/story.html	policy/LZx4MmGtzRGBg8CkDCladl/story.html
	https://www.bostonglobe.com/business/2017/09/12/super-	
	new-iphone-may-not-quite-super-enough-for-	
6	<pre>price/cVm5OkTOwWHIRtNtCORVDJ/story.html</pre>	https://www.bostonglobe.com/newsletters
	https://www.bostonglobe.com/business/2017/09/12/super-	
	new-iphone-may-not-quite-super-enough-for-	
7	<pre>price/cVm5OkTOwWHIRtNtCORVDJ/story.html</pre>	https://www.bostonglobe.com/tools/help/stafflist
	https://www.bostonglobe.com/business/2017/09/12/super-	
	new-iphone-may-not-quite-super-enough-for-	
8	price/cVm5OkTOwWHIRtNtCORVDJ/story.html	https://www.bostonglobe.com/tools/help/stafflist
	https://www.bostonglobe.com/business/2017/09/12/super-	
	new-iphone-may-not-quite-super-enough-for-	
9	<pre>price/cVm5OkTOwWHIRtNtCORVDJ/story.html</pre>	https://www.bostonglobe.com/newsletters
		https://www.bostonglobe.com/1970/01/05/term-
10	_	purchase/RfO4zG3drUPKl8BRZNKIDP/story.html

Query = Paris Climate Deal

#	Results of Lucene	Results of PageRank
		https://www.bostonglobe.com/1970/01/13/privacy-
1		policy/LZx4MmGtzRGBg8CkDCladI/story.html
	https://www.bostonglobe.com/news/politics/2017/06/02/jo	
	hn-kerry-calls-trump-climate-change-announcement-day-	
	craven-ignorance-and-	https://www.bostonglobe.com/1970/01/13/privacy-
2	cynicism/GWBwSaQwm24lsxfX5jHmWM/story.html	policy/LZx4MmGtzRGBg8CkDCladI/story.html
	https://www.bostonglobe.com/news/politics/2017/06/02/jo	
	hn-kerry-calls-trump-climate-change-announcement-day-	
	<u>craven-ignorance-and-</u>	
3	cynicism/GWBwSaQwm24lsxfX5jHmWM/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/magazine/2015/07/22/energ	
	<u>y-secretary-ernie-moniz-cabinet-star-</u>	
4	born/PsqmP1TairoTYQ1OSbt3uJ/story.html	https://www.bostonglobe.com/news/bigpicture
	https://www.bostonglobe.com/business/2017/06/14/stata-	
5	slowing-down-hardly/xBhdDp7fFMfsxpjCzGunQJ/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/metro/2017/06/02/climate/q	
6	ELM7JPNpKnORMMqkzi2XJ/story.html	https://www.bostonglobe.com/opinion
	https://www.bostonglobe.com/metro/2017/06/02/climate/q	
7	ELM7JPNpKnORMMqkzi2XJ/story.html	https://www.bostonglobe.com/sports
	https://www.bostonglobe.com/metro/2017/06/02/climate/q	
8	ELM7JPNpKnORMMqkzi2XJ/story.html	https://www.bostonglobe.com
	https://www.bostonglobe.com/metro/2017/07/04/united-	
	church-christ-approves-emergency-resolution-climate-	
9	change/pgzKhYbm3UEyJcxA21GPAI/story.html	https://www.bostonglobe.com/news/politics
	https://www.bostonglobe.com/metro/2017/07/04/united-	
	church-christ-approves-emergency-resolution-climate-	
10	change/pgzKhYbm3UEyJcxA21GPAI/story.html	https://www.bostonglobe.com/business

Overlap graph

