Report

Name : Shubha Manohar
USC Id : 9929527703

Steps followed to complete this assignment :

1) Installing Ubuntu and Solr:
   a) I had already dual booted my laptop and hence already had Ubuntu on my system.
   b) Solr Installation : Downloaded Solr 7.7.3 binary from https://www.apache.org/dyn/closer.lua/lucene/solr/7.7.3/solr-7.7.3.zip?action=download
2) Using Solr:
   a) cd into solr-7.7.3 folder and entered the command: bin/solr start
   b) Created a new core 'hw4' with the command : bin/solr create –c hw4
   c) Uncomment the following line - <!-- <copyField source="*" dest="_text_"/> --> from managed-schema file.
   d) Downloaded nytimes html pages from https://drive.google.com/drive/folders/19btQxjLM9EBEXgq2A3-EYR8L7TGXFTKh?usp=share_link.
   e) Indexed this by running bin/post -c hw4 -filetypes html ../NYTIMES/nytimes/
   f) Navigated to http://localhost:8983/solr/ , selected the core "hw4" and tried the default query (*:*) to see if everything is working fine.
   g) Uncommented the str element with name "df" in requestHandler in solrconfig.xml file and specified the default query field to be "_text_".
   h) Reloaded core by going to Core Admin on Solr UI and clicking on the "Reload" button.
3) PHP webpage creation:
   a) Installed and configured PHP in the Ubuntu System with Apache2 using the following commands:
      sudo apt install php libapache2-mod-php
      sudo apt install php-cli
      sudo apt install php-cgi
      And then restarted the apache server and checked to see if php was working.
   b) Cloned solr-php-client repository and used it in the php webpage created with html.

c) Provided radio buttons to toggle between Lucene and Pagerank (explained in the next part).
4) Working of Pagerank
   a) Installed intellij and used jsoup library jar. Wrote a Java program which used hashmap to create edgeList.ext
   b) Networkx was already downloaded as part of the Anaconda framework.
   c) Wrote a python script to generate the external_pageRankFile.txt using networkx and digraph and placed it in solr-7.7.3/server/solr/hw4/data/. The following parameters were used : alpha=0.85, personalization=None, max_iter=30, tol=1e-06, nstart=None, weight='weight',dangling=None
   d) Added pageRankFile external field in managed-schema file. Added eventListeners within the <query> element in the solrconfig.xml file.
   e) Navigated to Solr Dashboard UI -> Core Admin and clicked on "Reload".
   f) Put the value 'pageRankFile desc' in the sort field on solr UI.

Ran all the queries for both the ranking schemes.

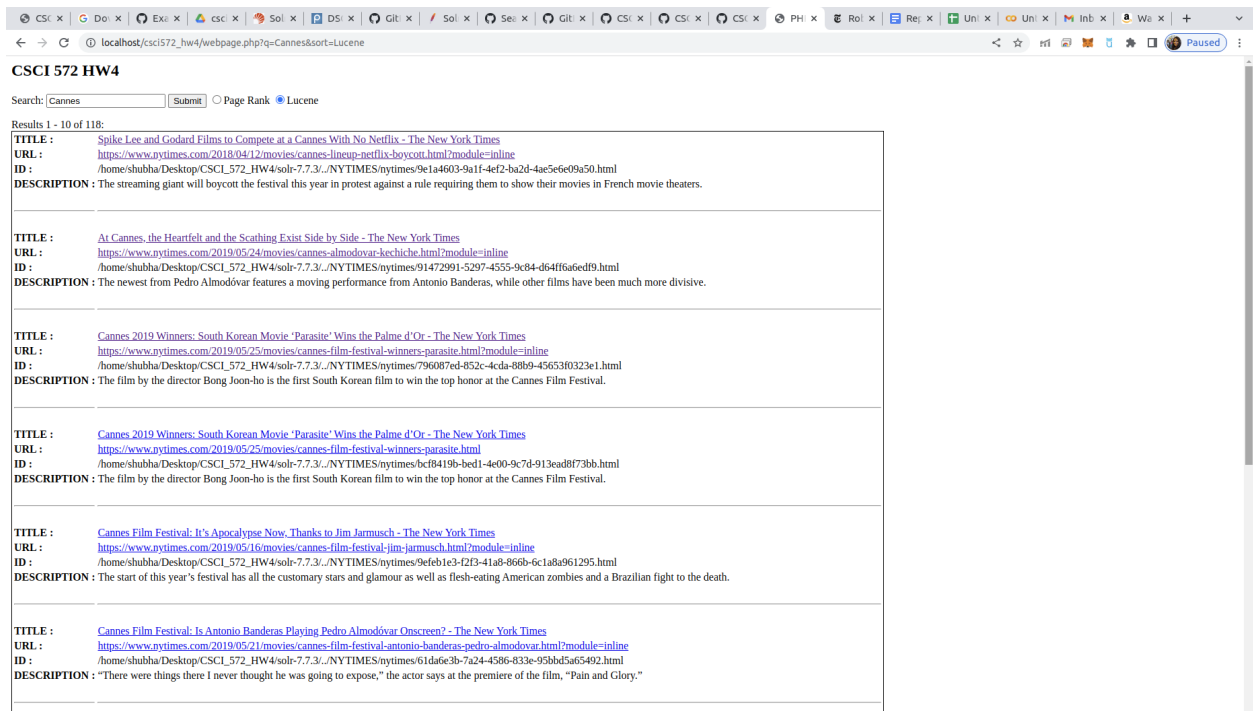# Why some pages have higher PageRank values than others :

PageRank calculates the rank of different web pages based on the number of links to it from other webpages (incoming links). A page that is linked to by many pages with high PageRank receives a high rank itself. Basically the pagerank formula will look at the number of inbound links, external links, and the PageRank of those links to determine authority of a webpage. Hence, some pages have higher PageRank values than others.

Screenshots :

1) Screenshot of the initial page where you enter the query in input box.
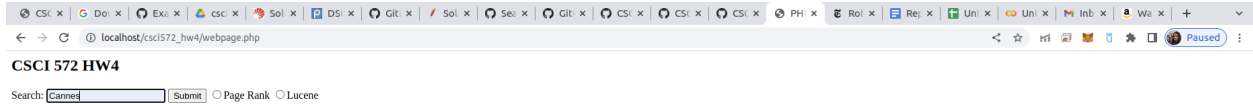
2) Screenshot of page which shows the results for Lucene(Default).

3) Screenshot of page which shows the results for PageRank.

4) Screenshot of the actual webpage that opens when clicking on one of the result's link.
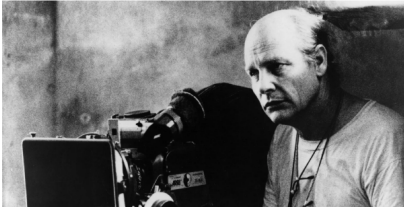
Cannes:

Lucene	Pagerank
"https://www.nytimes.com/2018/04/12/movies/cannes-lineup-netflix-boycott.html?module=inline"	"https://www.nytimes.com/2018/07/10/obituaries/robby-muller-inventive-cinematographer-is-dead-at-78.html?module=inline"
"https://www.nytimes.com/2019/05/24/movies/cannes-almodovar-kechiche.html?module=inline"	"https://www.nytimes.com/2019/05/22/movies/dicaprio-pitt.html?module=inline"
"https://www.nytimes.com/2019/05/25/movies/cannes-film-festival-winners-parasite.html?module=inline"	"https://www.nytimes.com/2019/04/25/theater/black-playwrights-theater.html?module=inline"
"https://www.nytimes.com/2019/05/25/movies/cannes-film-festival-winners-parasite.html"	"https://www.nytimes.com/2019/09/01/business/media/summer-box-office-movies.html"
"https://www.nytimes.com/2019/05/16/movies/cannes-film-festival-jim-jarmusch.html?module=inline"	"https://www.nytimes.com/2018/10/31/business/netflix-theater-streaming-release.html?module=inline"
"https://www.nytimes.com/2019/05/21/movies/cannes-film-festival-antonio-banderas-pedro-almodovar.html?module=inline"	"https://www.nytimes.com/2019/05/22/movies/dicaprio-pitt.html"
"https://www.nytimes.com/2019/05/27/movies/cannes-oscars.html"	"https://www.nytimes.com/2019/01/03/obituaries/mrinal-sen-dead.html"
"https://www.nytimes.com/2019/05/20/arts/cannes-film-festival-black-director.html?module=inline"	"https://www.nytimes.com/2019/06/03/movies/geena-davis-oscars.html"
"https://www.nytimes.com/2019/05/22/movies/quentin-tarantino-margot-robbie.html?module=inline"	"https://www.nytimes.com/2019/05/21/movies/cannes-film-festival-antonio-banderas-pedro-almodovar.html?module=inline"
"https://www.nytimes.com/2019/05/22/movies/dicaprio-pitt.html?module=inline"	"https://www.nytimes.com/2019/10/04/movies/new-york-film-festival.html?module=inline"

Congress:

Lucene	Pagerank
"https://www.nytimes.com/interactive/2019/us/politics/trump-impeachment-congress-list.html"	"https://www.nytimes.com/2019/09/29/opinion/trump-impeachment.html"
"https://www.nytimes.com/2019/11/05/magazine/congress-president-impeachment.html?module=inline"	"https://www.nytimes.com/2019/09/12/us/politics/ukraine-aid-white-house.html?module=inline"
"https://www.nytimes.com/2019/11/05/magazine/congress-president-impeachment.html"	"https://www.nytimes.com/2019/07/23/technology/justice-department-tech-antitrust.html?module=inline"
"https://www.nytimes.com/2019/07/25/us/politics/mueller-impeachment.html"	"https://www.nytimes.com/2019/10/23/us/politics/trump-deep-state-impeachment.html"
"https://www.nytimes.com/2019/07/26/us/politics/donald-trump-impeachment.html?rref=collection%2Fbyline%2Fnicholas-fandos&action=click&contentCollection=undefined&region=stream&module=inline&version=latest&contentPlacement=9&pgtype=collection"	"https://www.nytimes.com/2018/02/26/world/asia/xi-jinping-thought-explained-a-new-ideology-for-a-new-era.html?module=inline"
"https://www.nytimes.com/2018/09/17/us/politics/bloomberg-president-2020-democrat.html?module=inline"	"https://www.nytimes.com/interactive/2018/10/23/us/metoo-replacements.html?module=inline"
"https://www.nytimes.com/2018/11/21/obituaries/james-billington-dead.html"	"https://www.nytimes.com/2019/10/31/us/politics/katie-hill-speech.html"
"https://www.nytimes.com/2019/09/07/us/politics/congress-assault-weapons-ban.html"	"https://www.nytimes.com/2019/09/12/us/politics/ukraine-aid-white-house.html"
"https://www.nytimes.com/2019/08/07/us/politics/foreign-aid-freeze-congress.html?module=inline"	"https://www.nytimes.com/2019/11/02/sunday-review/data-protection-privacy.html?searchResultPosition=1&module=inline"
"https://www.nytimes.com/1973/05/13/archives/constitutional-ervin-ah-thi-nk-the-senate-is-determined-to-recover.html?module=in	"https://www.nytimes.com/video/world/americas/100000006042079/how-the-drone-attack-on-maduro-unfolded-in-venezuela.html

line"	?playlistId=100000005638006"

Democrats:

Lucene	Pagerank
"https://www.nytimes.com/2018/09/04/us/politics/ayanna-pressley-massachusetts.html?module=inline"	"https://www.nytimes.com/2019/09/29/opinion/trump-impeachment.html"
"https://www.nytimes.com/2018/11/07/us/politics/election-history-firsts-blackburn-pressley.html?module=inline"	"https://www.nytimes.com/2019/09/12/us/politics/ukraine-aid-white-house.html?module=inline"
"https://www.nytimes.com/2019/10/16/us/politics/democrats-impeachment.html?module=inline"	"https://www.nytimes.com/video/us/100000006417858/genetic-genealogy-alabama-cold-case.html?playlistId=100000004327932"
"https://www.nytimes.com/2019/09/29/us/politics/pelosi-trump-impeachment.html"	"https://www.nytimes.com/2019/10/23/us/politics/trump-deep-state-impeachment.html"
"https://www.nytimes.com/2018/01/23/us/politics/impeach-trump-democrats-tom-steyer.html"	"https://www.nytimes.com/interactive/2018/10/23/us/metoo-replacements.html?module=inline"
"https://www.nytimes.com/2019/09/24/us/politics/trump-democrats.html"	"https://www.nytimes.com/2019/10/31/us/politics/katie-hill-speech.html"
"https://www.nytimes.com/2019/10/28/us/politics/trump-impeachment-subpoena-house.html"	"https://www.nytimes.com/2019/09/12/us/politics/ukraine-aid-white-house.html"
"https://www.nytimes.com/issue/magazine/2019/06/25/the-61619-issue"	"https://www.nytimes.com/2019/10/24/opinion/democrats-neocons.html"
"https://www.nytimes.com/2019/11/05/opinion/trump-democrats-2020.html"	"https://www.nytimes.com/video/world/americas/100000006042079/how-the-drone-attack-on-maduro-unfolded-in-venezuela.html?playlistId=100000005638006"
"https://www.nytimes.com/2019/09/11/opinion/black-voters-democrats-2020-election.html?module=inline"	"https://www.nytimes.com/video/us/100000006789008/elijah-cummings-funeral.html?playlistId=100000002797598"

Patriot Movement:

Lucene	Pagerank
"https://www.nytimes.com/2019/08/14/opinion/joe-walsh-trump-primary.html"	"https://www.nytimes.com/interactive/2019/02/21/magazine/autism-office-design.html"
"https://www.nytimes.com/1976/08/27/archives/lewis-michaux-92-dies-ran-bookstore-in-harlem.html?module=inline"	"https://www.nytimes.com/2019/10/04/world/asia/hong-kong-emergency-powers.html"
"https://www.nytimes.com/2019/05/15/us/politics/us-iraq-embassy-evacuation.html?module=inline"	"https://www.nytimes.com/2018/05/04/books/junot-diaz-accusations.html?module=inline"
"https://www.nytimes.com/2019/01/07/obituaries/moshe-arens-dead.html"	"https://www.nytimes.com/2018/03/29/opinion/vietnam-war-necessary.html"
"https://www.nytimes.com/2019/10/27/obituaries/john-conyers-jr-dead.html"	"https://www.nytimes.com/interactive/2018/10/23/us/metoo-replacements.html?module=inline"
"https://www.nytimes.com/2018/12/29/obituaries/shehu-shagari-dead.html"	"https://www.nytimes.com/2013/03/03/arts/dance/jean-butler-riverdance-star-makes-her-own-work-hurry.html?module=inline"
"https://www.nytimes.com/2019/10/22/sports/football/jets-patriots.html"	"https://www.nytimes.com/2019/10/31/us/politics/katie-hill-speech.html"
"https://www.nytimes.com/by/jaclyn-peiser"	"https://www.nytimes.com/2019/02/25/movies/oscars-critics-green-book.html?module=inline"
"https://www.nytimes.com/2019/09/23/opinion/democrats-republicans.html"	"https://www.nytimes.com/2019/09/18/opinion/male-care-giver.html"
"https://www.nytimes.com/2019/10/30/opinion/alexander-vindman-trump-ukraine.html"	"https://www.nytimes.com/video/world/americas/100000006042079/how-the-drone-attack-on-maduro-unfolded-in-venezuela.html?playlistId=100000005638006"

Republicans:

Lucene	Pagerank
"https://www.nytimes.com/by/catie-edmondson"	"https://www.nytimes.com/2018/11/16/us/politics/presidential-medal-freedom-adelson.html?module=inline"
"https://www.nytimes.com/2019/10/24/us/politics/senate-republicans-trump.html"	"https://www.nytimes.com/2019/09/29/opinion/trump-impeachment.html"
"https://www.nytimes.com/interactive/2019/08/14/magazine/republicans-racism-african-americans.html"	"https://www.nytimes.com/2019/09/12/us/politics/ukraine-aid-white-house.html?module=inline"
"https://www.nytimes.com/2019/10/31/us/politics/house-impeachment-partisan-vote.html"	"https://www.nytimes.com/2019/10/23/us/politics/trump-deep-state-impeachment.html"
"https://www.nytimes.com/2019/11/04/us/politics/trump-republicans-impeachment.html?module=inline"	"https://www.nytimes.com/2018/03/29/opinion/vietnam-war-necessary.html"
"https://www.nytimes.com/2019/11/04/us/politics/trump-republicans-impeachment.html"	"https://www.nytimes.com/2019/10/31/us/politics/katie-hill-speech.html"
"https://www.nytimes.com/2019/10/04/us/politics/republicans-trump-defense.html"	"https://www.nytimes.com/2019/09/12/us/politics/ukraine-aid-white-house.html"
"https://www.nytimes.com/2019/09/30/opinion/republicans-trump-impeachment.html?module=inline"	"https://www.nytimes.com/2019/10/24/opinion/democrats-neocons.html"
"https://www.nytimes.com/2019/11/09/us/politics/republican-strategy-impeachment-trump.html"	"https://www.nytimes.com/video/us/100000006789008/elijah-cummings-funeral.html?playlistId=100000002797598"
"https://www.nytimes.com/2019/11/06/us/trump-senate-republicans-courts.html?module=inline"	"https://www.nytimes.com/2019/10/16/world/middleeast/trump-erdogan-turkey-syria-kurds.html"

Senate:

Lucene	Pagerank
"https://www.nytimes.com/2019/10/01/us/p	"https://www.nytimes.com/2018/11/16/us/p

olitics/senate-trump-impeachment.html"	olitics/presidential-medal-freedom-adelson.html?module=inline"
"https://www.nytimes.com/by/carl-hulse"	"https://www.nytimes.com/2019/09/29/opinion/trump-impeachment.html"
"https://www.nytimes.com/2019/10/26/us/trump-senate-presidency.html"	"https://www.nytimes.com/2019/09/12/us/politics/ukraine-aid-white-house.html?module=inline"
"https://www.nytimes.com/2019/08/03/us/politics/senate-votes-mcconnell.html?module=inline"	"https://www.nytimes.com/2019/07/23/technology/justice-department-tech-antitrust.html?module=inline"
"https://www.nytimes.com/2019/08/22/us/politics/john-hickenlooper-senate-2020.html"	"https://www.nytimes.com/2019/10/23/us/politics/trump-deep-state-impeachment.html"
"https://www.nytimes.com/2019/11/08/us/impeachment-senate-democrats.html"	"https://www.nytimes.com/interactive/2018/10/23/us/metoo-replacements.html?module=inline"
"https://www.nytimes.com/2019/10/24/us/politics/senate-republicans-trump.html"	"https://www.nytimes.com/2019/09/12/us/politics/ukraine-aid-white-house.html"
"https://www.nytimes.com/2019/08/31/us/politics/senate-race-2020.html"	"https://www.nytimes.com/2019/10/24/learning/impeachment-inquiry-teachable-moment.html"
"https://www.nytimes.com/2019/10/18/us/politics/mcconnell-impeachment-senate.html"	"https://www.nytimes.com/2019/10/16/world/middleeast/trump-erdogan-turkey-syria-kurds.html"
"https://www.nytimes.com/interactive/2016/us/politics/donald-trump-administration.html"	"https://www.nytimes.com/issue/todayspaper/2018/04/08/todays-new-york-times?module=inline"

Olympics 2020:

Lucene	Pagerank
"https://www.nytimes.com/2019/10/01/travel/what-you-need-to-know-the-tokyo-2020-olympics.html"	"https://www.nytimes.com/2013/06/05/opinion/global/xi-jinpings-chinese-dream.html?module=inline"

"https://www.nytimes.com/2018/02/15/learning/should-technology-in-sports-be-limited.html?module=inline"	"https://www.nytimes.com/2017/04/10/us/politics/trumps-company-settles-lawsuit-with-a-second-celebrity-chef.html?module=inline"
"https://www.nytimes.com/2018/02/15/learning/should-technology-in-sports-be-limited.html?rref=collection%2Fcolumn%2Flearning-student-opinion&module=inline"	"https://www.nytimes.com/2018/07/10/obituaries/robby-muller-inventive-cinematographer-is-dead-at-78.html?module=inline"
"https://www.nytimes.com/2017/12/05/sports/olympics/ioc-russia-winter-olympics.html?module=inline"	"https://www.nytimes.com/2019/10/03/arts/music/kim-shattuck-muffs-dead.html"
"https://www.nytimes.com/video/sports/olympics/100000005750317/johnny-weir-tara-lipinski-figure-skating.html"	"https://www.nytimes.com/2018/03/23/t-magazine/foundrae-jewelry-store.html"
"https://www.nytimes.com/video/sports/olympics/100000005750317/johnny-weir-tara-lipinski-figure-skating.html?playlistId=100000005719964"	"https://www.nytimes.com/2018/12/18/reader-center/crossword-puzzle-explainer.html"
"https://www.nytimes.com/video/sports/olympics/100000005752444/south-korea-speedskaters.html?playlistId=100000005719964"	"https://www.nytimes.com/2013/11/02/your-money/jon-huntsman-srs-early-calling-to-philanthropy.html?module=inline"
"https://www.nytimes.com/video/sports/olympics/100000005752444/south-korea-speedskaters.html"	"https://www.nytimes.com/2019/05/13/t-magazine/greece-ithaca-visit.html?module=inline"
"https://www.nytimes.com/2018/04/23/sports/breakdancing-olympics.html?module=inline"	"https://www.nytimes.com/2018/11/16/us/politics/presidential-medal-freedom-adelson.html?module=inline"
"https://www.nytimes.com/video/sports/football/100000005827420/nfl-cheerleaders-and-the-gender-gap.html?playlistId=1194811622289"	"https://www.nytimes.com/1998/01/24/us/the-president-under-fire-bill-clinton-accusations-and-explanations.html?module=inline"

Stock:

Lucene	Pagerank
"https://www.nytimes.com/interactive/2017/10/04/us/bump-stock-las-vegas-gun.html?module=inline"	"https://www.nytimes.com/2017/03/29/business/chase-ads-youtube-fake-news-offensive-videos.html?module=inline"
"https://www.nytimes.com/2019/05/02/business/tesla-stock-fundraising.html?module=inline"	"https://www.nytimes.com/interactive/2018/10/23/us/metoo-replacements.html?module=inline"
"https://www.nytimes.com/2018/12/16/technology/tech-workers-company-stock-shareholder-activism.html?module=inline"	"https://www.nytimes.com/2019/11/02/sunday-review/data-protection-privacy.html?searchResultPosition=1&module=inline"
"https://www.nytimes.com/2019/05/10/technology/uber-stock-price-ipo.html?module=inline"	"https://www.nytimes.com/2019/11/04/us/politics/elizabeth-warren-wall-street.html?module=inline"
"https://www.nytimes.com/2019/11/04/business/under-armour-stock-investigation.html"	"https://www.nytimes.com/2019/11/01/business/economy/jobs-report.html?module=inline"
"https://www.nytimes.com/2019/09/20/technology/airbnb-employees-ipo-payouts.html?module=inline"	"https://www.nytimes.com/2019/09/26/business/tech-ipo-market.html?module=inline"
"https://www.nytimes.com/2019/10/24/style/rebag-clair-handbag-stock-market.html"	"https://www.nytimes.com/section/upshot"
"https://www.nytimes.com/2019/05/30/technology/uber-stock-earnings.html?module=inline"	"https://www.nytimes.com/issue/todayspaper/2018/02/11/todays-new-york-times?module=inline"
"https://www.nytimes.com/2018/11/12/obituaries/david-pearson-nascar-dead.html"	"https://www.nytimes.com/2019/11/09/nyregion/airbnb-new-york-nycha.html"
"https://www.nytimes.com/2019/04/03/business/tesla-sales-elon-musk.html?module=inline"	"https://www.nytimes.com/2014/10/24/arts/design/hans-haacke-gets-establishment-nod-of-approval.html?module=inline"

Virus:

Lucene	Pagerank
"https://www.nytimes.com/2019/11/09/scie	"https://www.nytimes.com/2013/06/05/opini

nce/seals-distemper.html"	on/global/xi-jinpings-chinese-dream.html?module=inline"
"https://www.nytimes.com/2019/10/25/health/predict-usaid-viruses.html"	"https://www.nytimes.com/2019/11/07/business/gilead-truvada-hiv-lawsuit.html"
"https://www.nytimes.com/2019/10/28/us/politics/kay-hagan-dead.html"	"https://www.nytimes.com/2018/11/12/opinion/russia-meddling-disinformation-fake-news-elections.html"
"https://www.nytimes.com/2019/02/05/us/politics/trump-hiv-aids-plan.html?module=inline"	"https://www.nytimes.com/2019/09/16/health/ebola-congo-azar-fauci.html"
"https://www.nytimes.com/by/donald-g-mcneil-jr"	"https://www.nytimes.com/video/opinion/100000005790489/i-hacked-an-election-so-can-the-russians.html?playlistId=1194811622299"
"https://www.nytimes.com/by/tariro-mzezewa"	"https://www.nytimes.com/2019/10/21/science/loudest-bird-bellbird.html"
"https://www.nytimes.com/2019/01/23/neediest-cases/daughter-cmv.html"	"https://www.nytimes.com/2019/09/03/t-magazine/walter-van-beirendonck.html"
"https://www.nytimes.com/by/denise-grady"	"https://www.nytimes.com/2019/10/08/style/dating-while-ob-gyn.html"
"https://www.nytimes.com/2019/09/27/well/live/when-is-the-best-time-to-get-your-flu-shot.html"	"https://www.nytimes.com/by/christine-hauser"
"https://www.nytimes.com/2019/11/06/health/crispr-cancer-leukemia.html"	"https://www.nytimes.com/by/carl-zimmer"

Overlap Graph :

Overlap graph