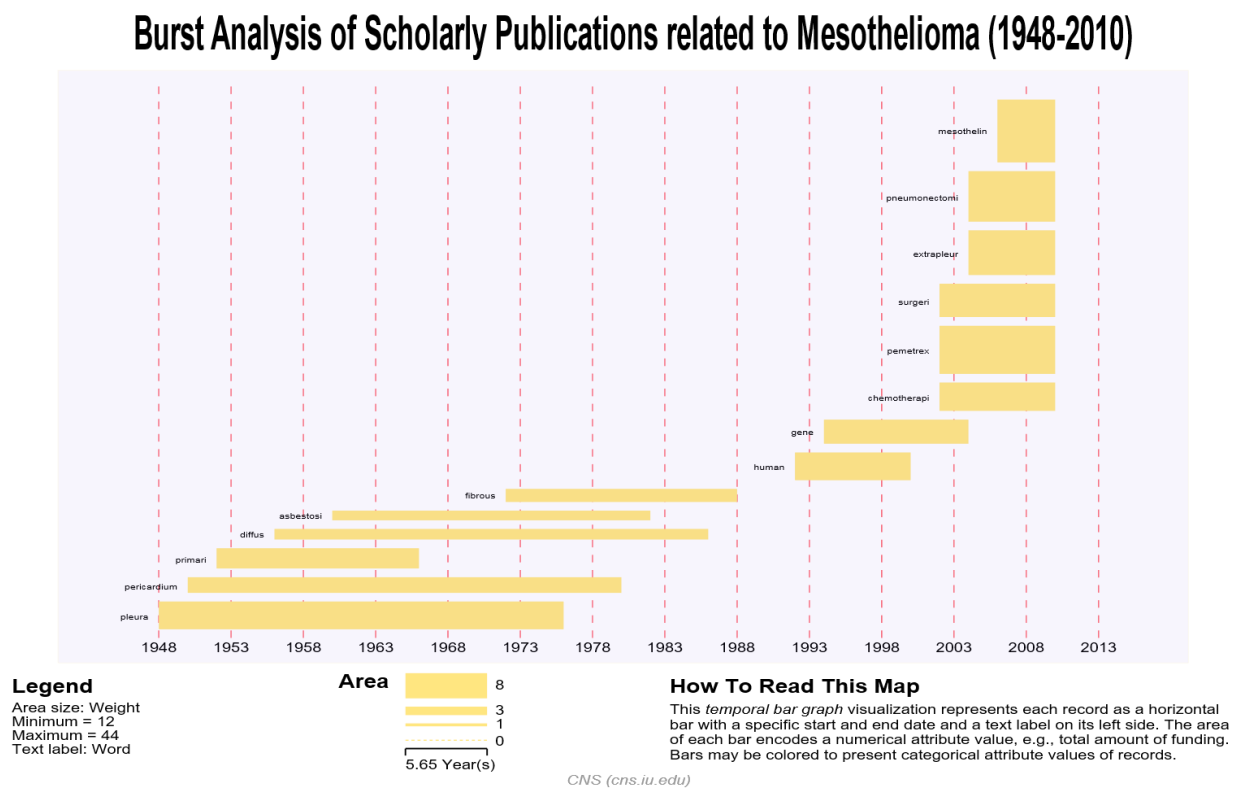
Homework 2

Samvat Rastogi
samrasto@indiana.edu

ILS-Z 637 Information Visualization

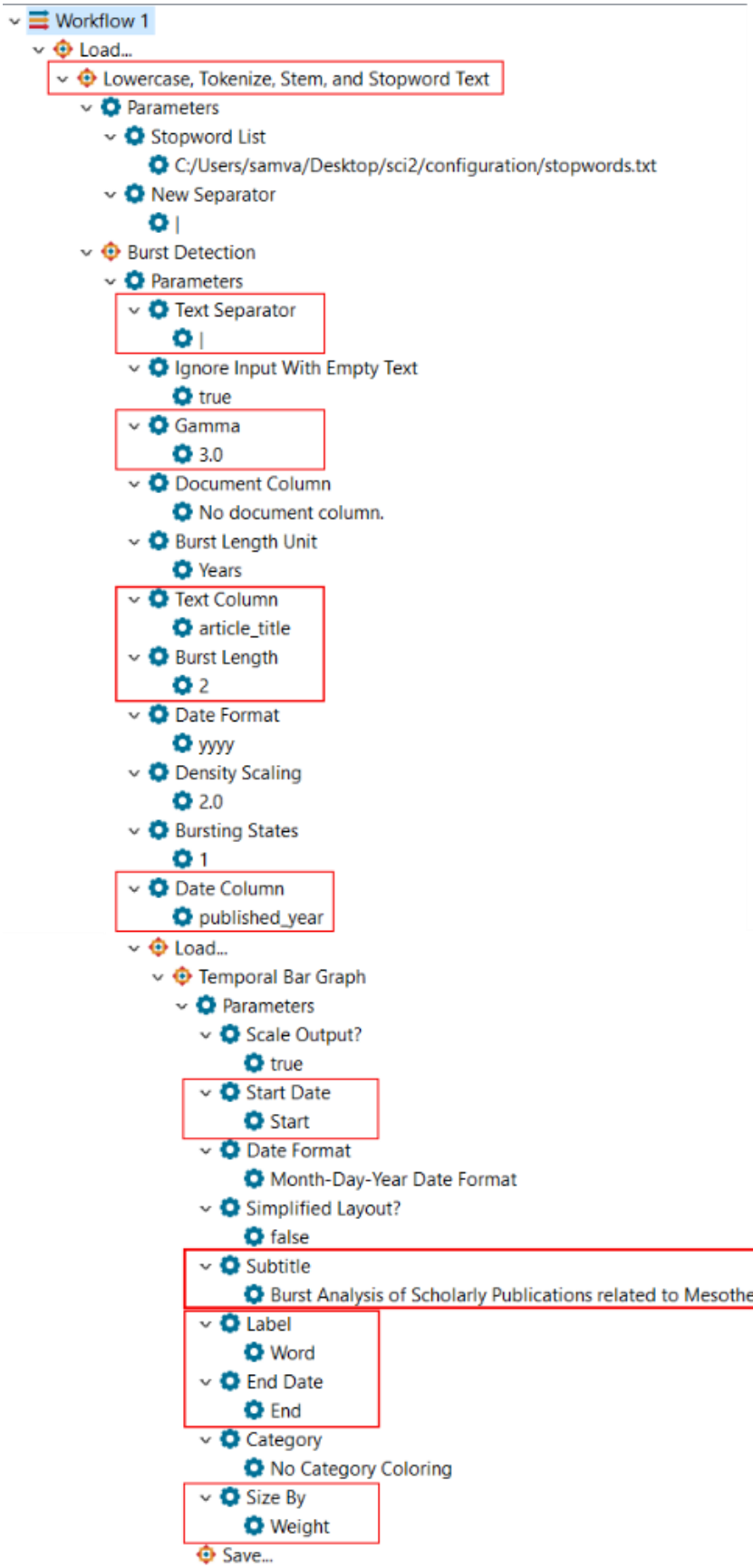
Making Temporal Burst Analysis Visualization – Approach

This document provides the other essential information about the temporal visualization which I have made. Below is a snapshot of visualization. A high-resolution copy of same can be found here.



The above visualization is based on the dataset provided in the homework description in Canvas. The workflow to create this visualization has been captured using Sci2 Workflow manager. Please note that all the parameter values are available in the following screenshot. All the important parameters have been highlighted by Red boxes. Two major parameters which have a great impact on data are:

1. Gamma: Set at 3.0
2. Burst Length: 2 years


Workflow 1
Load...
Lowercase, Tokenize, Stem, and Stopword Text
Parameters
Stopword List
C:/Users/samva/Desktop/sci2/configuration/stopwords.txt
New Separator
|
Burst Detection
Parameters
Text Separator
|
Ignore Input With Empty Text
true
Gamma
3.0
Document Column
No document column.
Burst Length Unit
Years
Text Column
article_title
Burst Length
2
Date Format
yyyy
Density Scaling
2.0
Bursting States
1
Date Column
published_year
Load...
Temporal Bar Graph
Parameters
Scale Output?
true
Start Date
Start
Date Format
Month-Day-Year Date Format
Simplified Layout?
false
Subtitle
Burst Analysis of Scholarly Publications related to Mesothelioma (1948-2010)
Label
Word
End Date
End
Category
No Category Coloring
Size By
Weight
Save...

After the burst analysis, I filtered out the words which do not have any specialized connection to a cancer research paper. The words highlighted in Red are NOT specific to the given cancer research as I believe they are more generic in nature.

Burst Analysis Output-csv

Burst Analysis Output-xlsx

Word	Level	Weight	Length	Start	End
surgeri	1	15.111373	5	2002	
primari	1	16.080032	8	1952	1966
pneumonectomi	1	17.164862	4	2004	
human	1	12.568955	5	1992	2000
chemotherapi	1	12.848056	5	2002	
pemetrex	1	21.779227	5	2002	
mesothelin	1	14.240798	3	2006	
pericardium	1	26.574274	16	1950	1980
extrapleur	1	15.16887	4	2004	
author	1	24.027117	5	1974	1982
transl	1	24.027117	5	1974	1982
fibrous	1	11.871329	9	1972	1988
case	1	32.923898	22	1932	1974
gene	1	13.656039	6	1994	2004
pleura	1	43.957899	23	1932	1976
asbestosi	1	11.910517	12	1960	1982
report	1	12.288506	18	1932	1966
diffus	1	18.024226	16	1956	1986

Now according to the problem, the burst analysis visualization should range from 1948 to 2010. So, I used 2010 to fill in the missing values in “End” column. Also, the row, partially highlighted in gold, has a start date from 1932. So I changed it to 1948, changing “Length” to 15 as

$$23 = \left(\frac{1976-1932}{2}\right) + 1$$

Similarly,

$$15 = \left(\frac{1976-1948}{2}\right) + 1$$

Below is the processed output on which the visualization is based.

Burst Analysis Output – Processed

Word	Level	Weight	Length	Start	End
surgeri	1	15.11137	5	2002	2010
primari	1	16.08003	8	1952	1966
pneumonectomi	1	17.16486	4	2004	2010
human	1	12.56895	5	1992	2000
chemotherapi	1	12.84806	5	2002	2010
pemetrex	1	21.77923	5	2002	2010
mesothelin	1	14.2408	3	2006	2010
pericardium	1	26.57427	16	1950	1980
extrapleur	1	15.16887	4	2004	2010
fibrous	1	11.87133	9	1972	1988
gene	1	13.65604	6	1994	2004
pleura	1	43.9579	15	1948	1976
asbestosi	1	11.91052	12	1960	1982
diffus	1	18.02423	16	1956	1986

Hence, a total of 14 words can be visualized on a temporal graph with Gamma parameter of 3 and Burst length of 2.