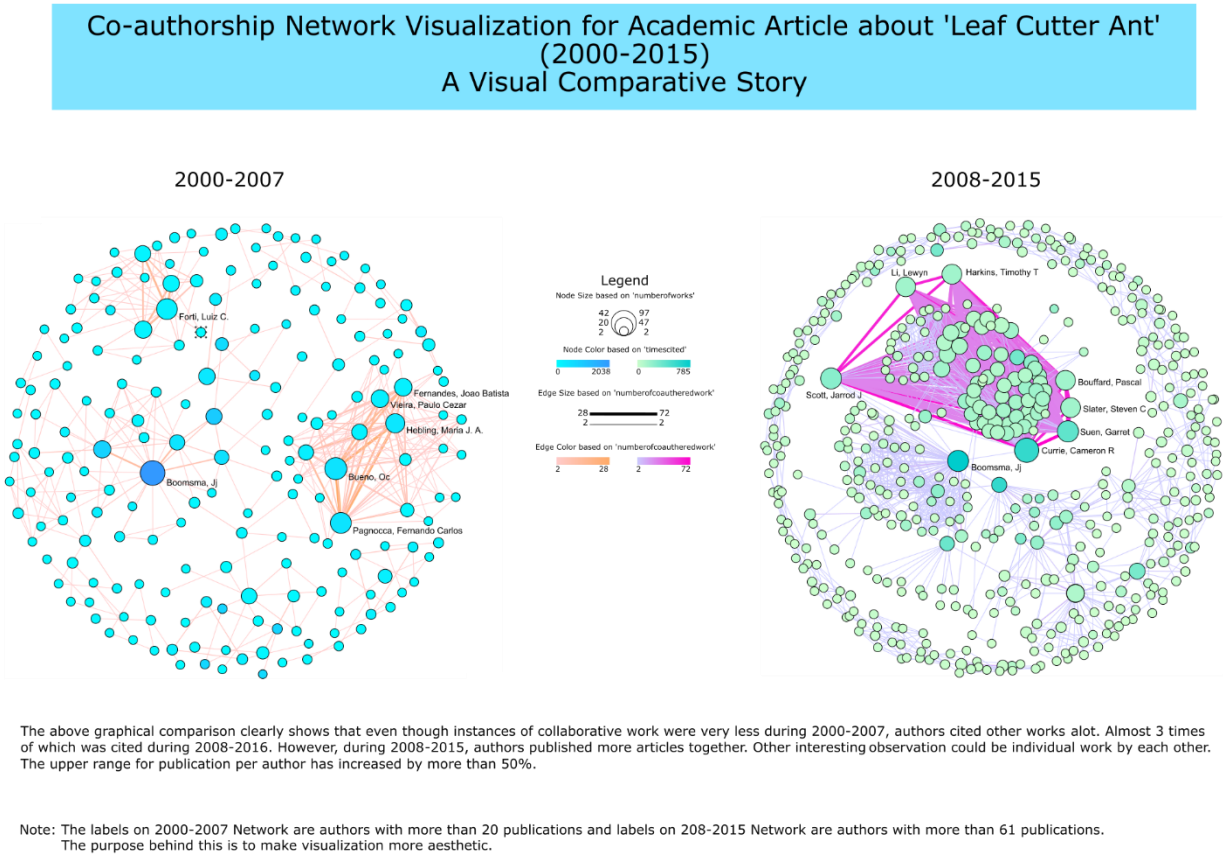
Homework 6

Samvat Rastogi
samrasto@indiana.edu

ILS-Z 637 Information Visualization

'With Whom': Networks – Approach

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



The above graphical comparison clearly shows that even though instances of collaborative work were very less during 2000-2007, authors cited other works alot. Almost 3 times of which was cited during 2008-2016. However, during 2008-2015, authors published more articles together. Other interesting observation could be individual work by each other. The upper range for publication per author has increased by more than 50%.

Note: The labels on 2000-2007 Network are authors with more than 20 publications and labels on 208-2015 Network are authors with more than 61 publications. The purpose behind this is to make visualization more aesthetic.

The above visualization is based on the dataset provided in the homework description in Canvas.

1. The Process:

   The Dataset was mostly preprocessed. Therefore there is **NO PROCESSING SCRIPT FILE**. So I was able to do Time slicing from Preprocessing > Temporal > Slice Table by Time. Below are the parameters I used to slice the data into 2 disjoint parts of 8 years (2000-2007 to 2008-2015).
   Input Parameters:
   To Time: 2015

Slice Into: years
Week Starts On: Sunday
Date/Time Column: Publication Year
From Time: 2000
How Many?: 8
Align With Calendar: false
Date/Time Format: yyyy
Cumulative?: false

2. After slicing it, I got two data time slices. For each time slice, I performed the below steps:
    a. Extracted Co-occurrence Network graph with below parameters and property files provided in the problem statement:

        Extract Co-Occurrence Network was selected.
        Implementer(s): Timothy Kelley
        Integrator(s): Timothy Kelley
        Documentation: http://wiki.cns.iu.edu/display/CISHELL/Extract+Co-Occurrence+Network
        Input Parameters:
        Aggregation Function File: C:/Users/samva/OneDrive - indiana.edu/Assignments/Information Visualization/HW6/WK6-WoS-CoAuthorship.properties
        Text Delimiter: |
        Column Name: Authors

    b. Preprocessed data slices to extract nodes based on the number of works with a minimum value of 2. And for edges, extraction was done on basis of coauthored work with a minimum value of 2. Further, isolated nodes were removed.

        Data Slice 1:
        Input Parameters:
        Below?: false
        Numeric Attribute: numberofworks
        Extract from this number: 1.0
        .....
        Extract Edges Above or Below Value was selected.
        Author(s): Thomas G. Smith
        Implementer(s): Thomas G. Smith
        Integrator(s): Thomas G. Smith
        Documentation:http://wiki.cns.iu.edu/display/CISHELL/Extract+Edges+Above+or+Below+Value

        Input Parameters:
        Below?: false

Numeric Attribute: numberofcoauthoredworks
Extract from this number: 1.0
.....
Delete Isolates was selected.
Implementer(s): Patrick Phillips
Integrator(s): Patrick Phillips
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Delete+Isolates
Removed 33 isolate nodes.

Data Slice 2:
Input Parameters:
Below?: false
Numeric Attribute: numberofworks
Extract from this number: 1.0
.....
Extract Edges Above or Below Value was selected.
Author(s): Thomas G. Smith
Implementer(s): Thomas G. Smith
Integrator(s): Thomas G. Smith
Documentation:
http://wiki.cns.iu.edu/display/CISHELL/Extract+Edges+Above+or+Below+Value

Input Parameters:
Below?: false
Numeric Attribute: numberofcoauthoredworks
Extract from this number: 1.0
.....
Delete Isolates was selected.
Implementer(s): Patrick Phillips
Integrator(s): Patrick Phillips
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Delete+Isolates
Removed 49 isolate nodes.

c. Performed NAT analysis with below results for both slices:
Data Slice 1:
This graph claims to be undirected.

Nodes: 196
Isolated nodes: 0
Node attributes present: label, timescited, cited2000, numberofworks

Edges: 369
No self loops were discovered.
No parallel edges were discovered.

Edge attributes:

Did not detect any nonnumeric attributes.

Numeric attributes:

	min	max	mean
numbero...	2	28	3.65583
weight	2	28	3.65583

This network seems to be valued.

Average degree: 3.7653
This graph is not weakly connected.
There are 35 weakly connected components. (0 isolates)
The largest connected component consists of 43 nodes.
Did not calculate strong connectedness because this graph was not directed.

Density (disregarding weights): 0.0193
Additional Densities by Numeric Attribute

Data Slice 2:
This graph claims to be undirected.

Nodes: 501
Isolated nodes: 0
Node attributes present: label, timescited, cited2000, numberofworks

Edges: 2371
No self loops were discovered.
No parallel edges were discovered.

Edge attributes:

Did not detect any nonnumeric attributes.

Numeric attributes:

	min	max	mean
numbero...	2	72	16.92661
weight	2	72	16.92661

This network seems to be valued.

Average degree: 9.4651
This graph is not weakly connected.
There are 53 weakly connected components. (0 isolates)
The largest connected component consists of 275 nodes.
Did not calculate strong connectedness because this graph was not directed.

Density (disregarding weights): 0.0189
Additional Densities by Numeric Attribute

d. Created Kamada-Kawai formatted network visualization in GUESS.

3. Further, I performed 2 more analysis, Betweenness Centrality and Node Degree and Strength with 'numberofcoauthoredworks' as a parameter.

Node Betweenness Centrality was selected.
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Node+Betweenness+Centrality
Input Parameters:
Weight: numberofcoauthoredworks
Centrality Attribute Name: betweenness_centrality
Saved: C:\Users\samva\OneDrive - indiana.edu\Assignments\Information Visualization\HW6\Betweenness Centrality_2000_2007.nwb

Node Betweenness Centrality was selected.
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Node+Betweenness+Centrality
Input Parameters:
Weight: numberofcoauthoredworks
Centrality Attribute Name: betweenness_centrality
Saved: C:\Users\samva\OneDrive - indiana.edu\Assignments\Information Visualization\HW6\Betweenness Centrality_2008_2015.nwb

Degree && Strength was selected.
Implementer(s): Duygu Balcan
Integrator(s): Russell Duhon
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Degree+%26+Strength
Input Parameters:
Weight Attribute: numberofcoauthoredworks
Number of nodes : 196
Number of edges : 369
Saved: C:\Users\samva\OneDrive - indiana.edu\Assignments\Information Visualization\HW6\Node_Degree_and_Strength_2000_2007.nwb

Degree && Strength was selected.
Implementer(s): Duygu Balcan
Integrator(s): Russell Duhon
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Degree+%26+Strength
Input Parameters:
Weight Attribute: numberofcoauthoredworks

Number of nodes : 501
Number of edges : 2371
Saved: C:\Users\samva\OneDrive - indiana.edu\Assignments\Information Visualization\HW6\Node_Degree_and_Strength_2008_2015.nwb

4. Number of Analysis done: 3
    a. NAT analysis : Outputs: (2000-2007, 2008-2015)
    b. Node Betweenness Centrality : Outputs: (2000-2007, 2008-2015)
    c. Degree && Strength : Outputs: (2000-2007, 2008-2015)

5. Analysis and Insights:
From the comparative study of the graph, we can clearly infer that number of publications about articles related to Leaf Cutter Ants has increased drastically and so does the number of authors writing about it. There were 196 authors who wrote about Leaf Cutter Ants at least twice during 2000-2007. This number increased to 501 authors for 2008-2015.The visualization also shows that upper limit of the number of articles published has increased from 42 to 97. That's more than 100% increase. The average number of collaborative work has increased from 3.66 to 16.93. However, the number of citations has decreased to 38.5% from Period 2000-2007 to Period 2008-2015.

I decided to used Disjoint Time Slices because as per the problem statement, I have to a comparative study of only 'ONE' Scholarly word over a time period. Moreover, Disjoint times slices give the acute details about how things have changes in another time frame.
For example, the most influential authors of 2000-2007 Era were Boomsma JJ, Bueno Oc, and Forti Luiz C and of 2008-2015 Era were Currie Cameron R, Suen Garret, and Boomsma JJ. Moreover, it can also help to compare details about one author in two different time periods. For example, Boomsma JJ, during 2000-2007 was a highly cited author and had the highest number of publications. While during 2008-2015, Currie, Cameron R had the highest number of publications.

NOTE: The labels on 2000-2007 Network are authors with more than 20 publications and labels on 208-2015 Network are authors with more than 61 publications. The purpose behind this is to make the visualization more aesthetic.
The year 2016 was not considered because including it will have disturbed the uniformity of time frames. Though there were only two records for the year 2016, but a small variation in time range could have lead different analysis and results.