Network Visualizations

#6. Simple statistics of the data sets used, e.g., number of entities, major entity attributes, etc.

The following clean up was performed based on the Julie Bass’s feedback:

-Identified projects with duplicate values and deleted projects with highest Project ID number (314, 92, 99, 90, 88, 86, 89, 85, 87, 91)

- assigned ProjectID 9999 to Project ID 1

- reassigned Ellen S Cohn projects to Helen S Cohen

- Updated blank USR\_FUNDING\_DATE\_END with the current date.

Performed crosstab analysis between USR\_PRINCIPAL\_INVESTIGATOR and number of associated parameters:

USR\_PRINCIPAL\_INVESTIGATOR (193 unique) vs USR\_AGENDA\_CATEGORY (7) (excerpt): top 5 PIs depicted in yellow



USR\_PRINCIPAL\_INVESTIGATOR vs USR\_DIAGNOSIS\_AREA(29) (excerpt): top 5 PIs depicted in yellow



USR\_PRINCIPAL\_INVESTIGATOR (193) vs ICF Categories(7) (excerpt):



- Data was sliced by the USR\_FUNDING\_DATE\_START 1982 through 2017 by seven years cumulatively

Number of project was proportionally (times 3) increasing from 18 to 54 to 149 to 406 to 755 in 2016

Properties used:

node.countProjectsperResearcher = PROJECT\_ID.count

edge.countProjectsPerDiagnosis=USR\_DIAGNOSIS\_AREA.count

Network Analysis Toolkit (NAT) was selected.

This graph claims to be directed.

Nodes: 18

Isolated nodes: 0

Node attributes present: label, countProjectsperResearcher, bipartiteType

Edges: 16

No self loops were discovered.

No parallel edges were discovered.

Edge attributes:

Did not detect any nonnumeric attributes.

Numeric attributes:

min max mean

countPr... 1 2 1.125

This network seems to be valued.

Average total degree: 1.7778

Average in degree: 0.8889

Average out degree: 0.8889

This graph is not weakly connected.

There are 3 weakly connected components. (0 isolates)

The largest connected component consists of 9 nodes.

This graph is not strongly connected.

There are 18 strongly connected components.

The largest strongly connected component consists of 1 nodes.

Density (disregarding weights): 0.0523

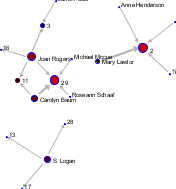
Additional Densities by Numeric Attribute

Initial analysis was performed on USR\_PRINCIPAL\_INVESTIGATOR vs USR\_DIAGNOSIS\_AREA

#7. Data analysis/visualization (algorithms) applied and resulting visualizations

Two visualization models were created and analyzed based on Network with directed edges from USR\_PRINCIPAL\_INVESTIGATOR to USR\_DIAGNOSIS\_AREA.2 visualized using

1)GUESS GEM



Analyzed time slice 1982 through 1989

Numbers represent the Diagnosis Area ID(1 through 29). Nodes were proportionally sized based on the number of projects per PI. Edges were proportionally sized by number of projects associated with the Diagnosis area

Encountered limitations of the GUESS visualization: cannot easily distinguish between PIs and Diagnosis area.

2)performed viz in GEPHI using Fruchterman Reigold view

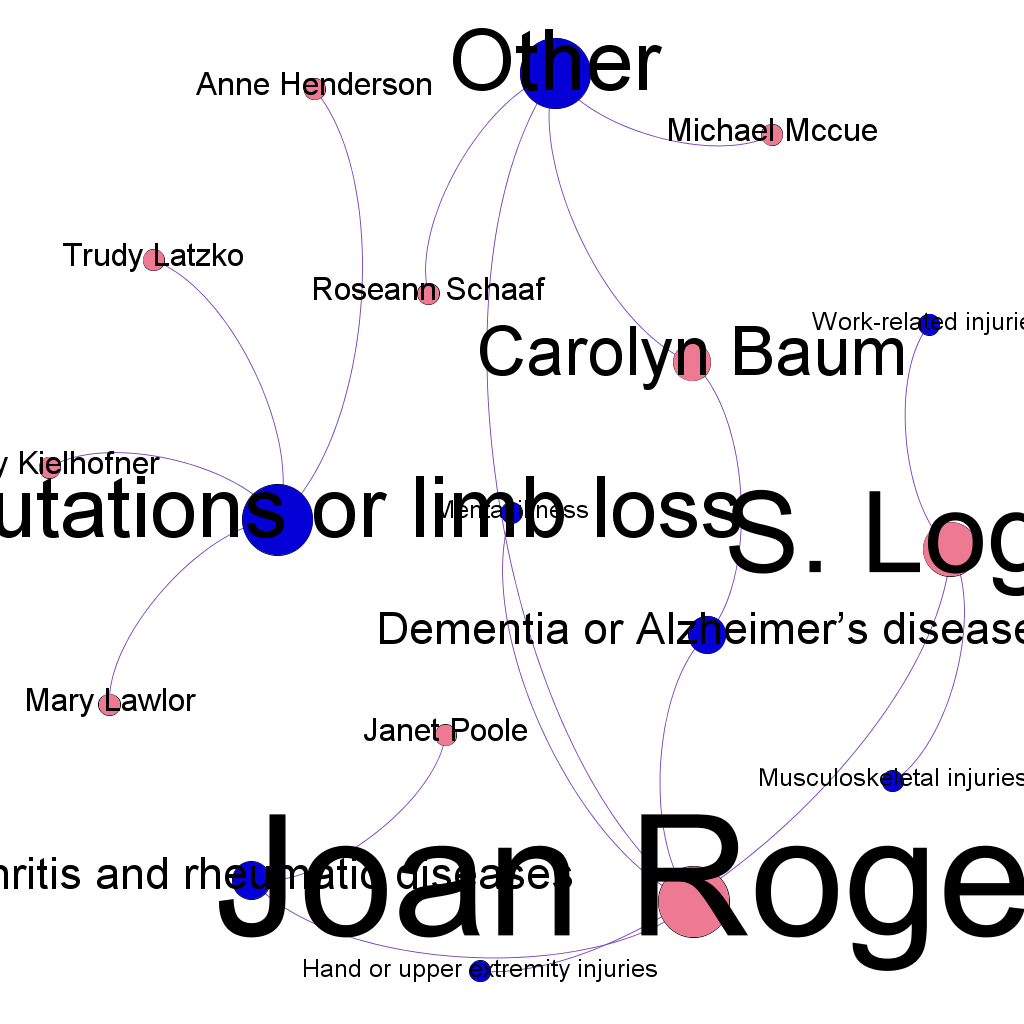
- node color by multimode

-label for Researcher sized by out-degree (number of Diagnosis per researcher)

- node sized by degree (number of projects per Diagnosis)

- edges colored by numberProjects perDiagnosisperInvestigator

Preview:default curved



#8. Discussion of key insights gained from the analysis/visualization

For the particular time slice (1982-1989) both types of visualization clearly depicted top players in field being Joan Rogers, S. Logan and Carolyn Baum (nodes or labels were sized proportionally to the number of respective projects ). In addition, Diagnosis areas 2 - Amputations or limb loss, 29-Other, 3- Arthritis and rheumatic diseases and 11- Dementia or Alzheimer’s disease were associated with the most number of researchers. Even though both algorithms correctly represented the data and the insights, the Gephi algorithm was the most flexible in capturing the types of the nodes used in the bipartite network and allowed to separate PI from diagnosis areas by color. The Gephi algorithm will be used to implement further network visualizations.