Structure of Project

Summary or Abstract:

* States the purpose, the methods, and the results of the project. (Purpose is to advance university and industrial partnerships by finding domestic untapped human capital related to specific research subject areas of interest.) (Methods utilized to obtain output were: Adapted Pure H-Index, Historical Local Citation Network Analysis, K-means Clustering on subject areas using correspondence analysis, and publication affiliation frequencies) (Results/Output show organizations tied to specific research subject areas of interest and most prolific authors (email, affiliation, papers, abstracts, keywords, and local citations) within the publications related to specific research subject areas of interest.)

Project Steps

Proof of Concept Data Acquisition

* Determined research subject areas of interest through qualitative data collection by interviewing personnel of interest. Allowed for the creation of first data set which contained 500 publications about one specific research subject area of interest. Determined that Web of Science is the best data source since it is focused most of STEM related material as oppose to Scopus and Google Scholar (which did not allow for effective data extraction (20 publications at a time only through Publish or Perish))

Proof of Concept Data Analysis

* Determined that the utilization of the bibliometrix package in R was fundamental to the analysis. Utilized this package to process test data on Theoretical Physics and produce output such as affiliations frequency and authors of interest based on adapted pure H-index. Local citation network was not able to be produced due to low local citation counts.

Data Acquisition/Validation

* Data was compiled in .bib files from Web of Science based on 4 research subject areas of interest from father organization. Files were merged through “Rbind”, ROT (Redundant, Outdated, Trivial) information was removed, Rbind was verified on key columns of interest, on-target data collection was measured by manually reading 10 abstracts per research subject area, and examining samples of size 100 publication keywords per subject and creating binomial tests to see how on target populations were for each subject area. Also removed multiple country publications to only incorporate domestic researchers per desires of father organization(customer)

Data Analysis

* Created output similar to PoC Data Analysis but going further by incorporating historical local citations network analysis plot, k-means clustering of keywords per research subject area of interest and providing more sound output regarding authors of interest by examining all publications related to the research subject areas of interest.)

Results

* Output Plots showing results for best affiliations, keywords for research subject areas, and table of most prolific authors per research subject area.)

Next Steps

* Shiny App, Data Analysis Validation: Network cross-validation by edge sampling, potential combined metric on local citations and adapted pure H-index, Publication affiliations frequency count weighted by combined metric for authors.)