Collaboration in Research

Blaine Farrell

Goal

Finding the impact of collaboration on an author's rank, what drives an increase in an entity's rank, and the differences in impact between each field of study.

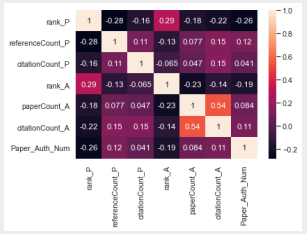
Background

- Social Network Analysis:
  - Nodes - each individual research author
  - Edges (Ties) - co-author relations between researchers
  - Importance of Structure - similar to covalent bonds in Chemistry
- Academic Social Network
  - Networks for academic entities
- Scholarly Big Data:
  - Large amount of data pertaining to academia

Measurements

- Paper Author Number – number of authors in a paper
- Between - the number of shortest paths between nodes that pass through a node
- Eigenvector - measure of influence of a node
- Closeness - sum of the length of the shortest paths between the node and all other nodes in the graph
- Number of Unique Co-Authors – number of coauthors that an author has worked with
- Entity Rank – probability of an entity being important

Paper Analysis

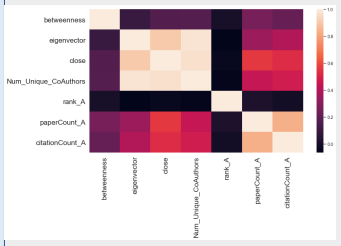


Field of Study	
Art	0.176302
Biology	0.351038
Business	0.301841
Chemistry	0.357608
Computer science	0.318231
Economics	0.253002
Engineering	0.355423
science	0.321036
Geography	0.241203
Geology	0.376647
History	0.159304
Materials science	0.39057
Mathematics	0.325191
Medicine	0.352145
Philosophy	0.12752
Physics	0.32342
Political science	0.250773
Psychology	0.27023
Sociology	0.248287

Paper Rank Correlations

rank_P	-0.260685
referenceCount_P	0.123457
citationCount_P	0.041057
rank_A	-0.186841
paperCount_A	0.083905
citationCount_A	0.112847

Author Analysis



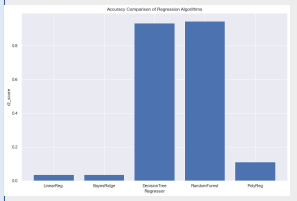
Fields of Study	
Art	-0.10828
Biology	-0.07696
Business	0.333928
Chemistry	-0.09585
Computer science	-0.0187
Economics	-0.0525
Engineering	-0.01664
Environmental science	-0.08433
Geography	-0.07197
Geology	-0.1115
History	-0.03289
Materials science	-0.01221
Mathematics	-0.02265
Medicine	-0.13311
Philosophy	-0.06857
Physics	-0.01317
Political science	0.025074
Psychology	0.036448
Sociology	-0.06763

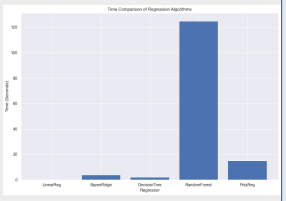
Author Rank Correlations

degree	-0.046171
betweenness	0.040615
eigenvector	-0.052977
close	-0.030472
Num_Unique_CoAuthors	-0.046171
rank_A	1.000000
paperCount_A	0.025924
citationCount_A	-0.007920
Cite_per_P	-0.174996

Machine Learning Algorithms

- Linear Regression - uses residuals to optimize a line of best fit for the given data
- Bayesian Ridge - conditional model where the mean of a variable is described by a linear combination of other variables
- Decision Tree Regression - uses a tree-like model of decisions to either predict the target value
- Random Forest Regression - combines the output of multiple decision trees to reach a single result
- Polynomial Regression - relationship between the independent & dependent variables is modelled as an nth degree polynomial





Feature Importance