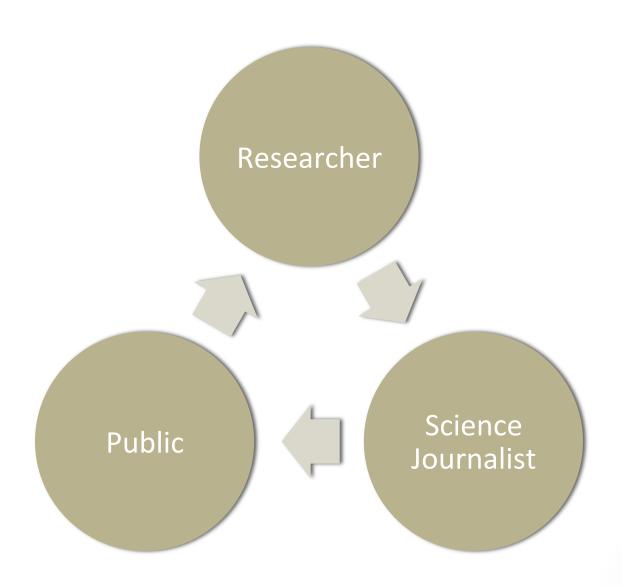
Science Surveyor Citation Network Analysis

Shuheng Gong (sg3124) EECS E6898 Project Presentation

Problem



Science Surveyor Road Map

Small Database (ACL Anthology Network) Open Platform (arXiv, PLoS One) API access database (Thomson Reuters, JSTOR, and Elsevier)

Four Dimension Evaluation

Consensus Layer

- Citation Network
- Relation with consensus

Funding Layer

- Biased Result, similar network
- Popularity Indicator

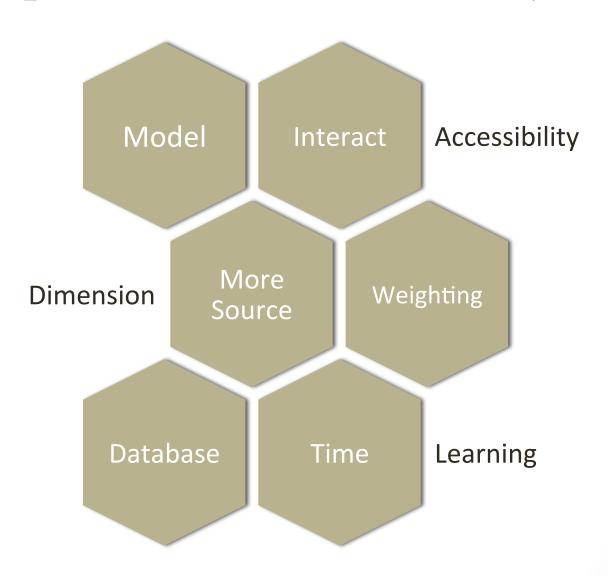
Idea Layer

- NLP Based
- Content Similarity

Temporal Layer

• Pattern Across Time

Compare to Previous Project



Evaluation Methods

Human Test

Compare cluster view and time zone view with existing citation system

Evaluate Accessibility and Result Quatliy by Science Jourlist

Computer Test

Validation in research front classification with labeled finding/region test set

Cross Validation when building trends prediction model

Preliminary Experiments

- Snap.py
- NetworkX
- Graph-tool
- igraph

Library

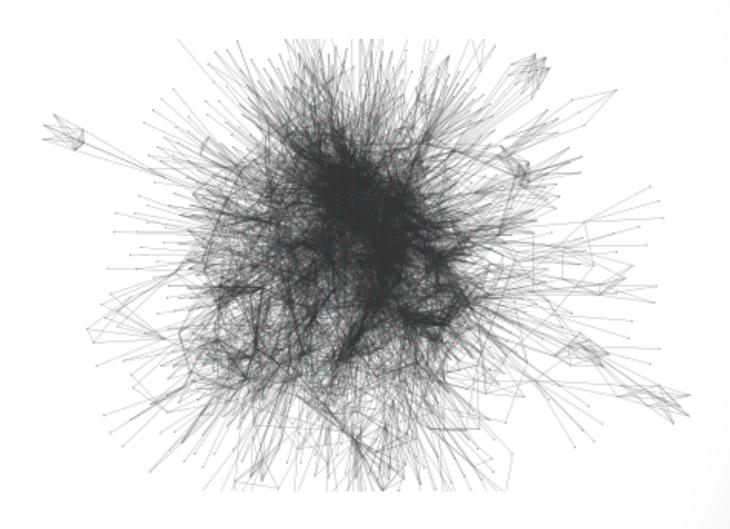
Centrality

- Degree, Closeness, Betweeness
- Eigenvector, Pagerank
- Percolation
- Communicability

- Weight
- Different Measurement for different node type
- Need Collaboration

Criterion

Visualize the Network



Result Comparison

Table.1 Centrality of Michael John Collins

Centrality	2008	2009	2010
Betweenness	0.020279099	0.020908112	0.016284481
PageRank	0.003676429	0.003630617	0.003218470

Table.2 Centrality of Susan E. Brennan

Centrality	2008	2009	2010
Betweenness	0.000755152	0.000655850	0.000232821
PageRank	0.000335703	0.000292816	0.000209016

Facility/Collaboration Needed

- Graphic Design (Web UI/UX)
- NLP (Extract Idea Network)
- Experts in the specific region (decide the criterion of scoring and centrality evaluation)
- Science journalist (Find what they want)
- Scholarly publisher (Access to the publication database)
- Press (Make it known to journalists and the public)

Thank you!

