

4-1-2019

Moliere: Automatic Biomedical Hypothesis Generation

Justin Sybrandt
Clemson University

Michael Shtutman
Clemson University

Ilya Safro
Clemson University

Follow this and additional works at: https://tigerprints.clemson.edu/grads_symposium

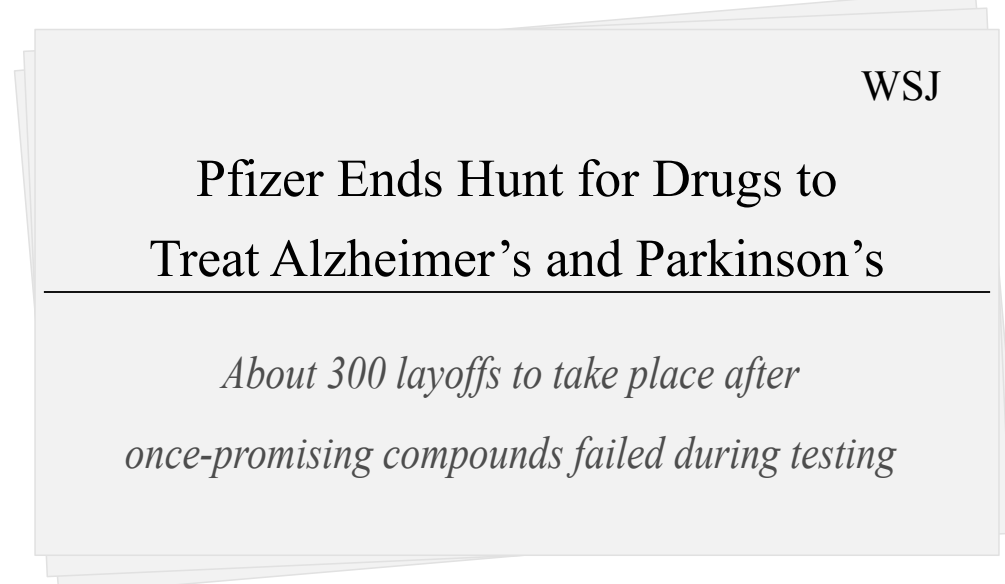
Recommended Citation

Sybrandt, Justin; Shtutman, Michael; and Safro, Ilya, "Moliere: Automatic Biomedical Hypothesis Generation" (2019). *Graduate Research and Discovery Symposium (GRADS)*. 244.
https://tigerprints.clemson.edu/grads_symposium/244

This Poster is brought to you for free and open access by the Student Works at TigerPrints. It has been accepted for inclusion in Graduate Research and Discovery Symposium (GRADS) by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

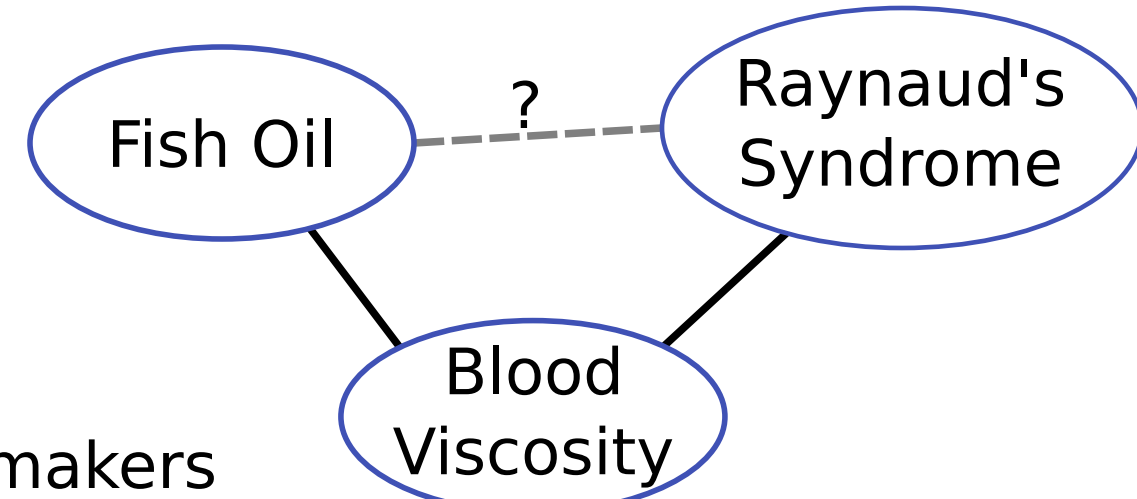
Problem

- Medical research needs early investment
- 2-4k bio papers published daily
- Missed connections lead to missed treatment options, costing lives



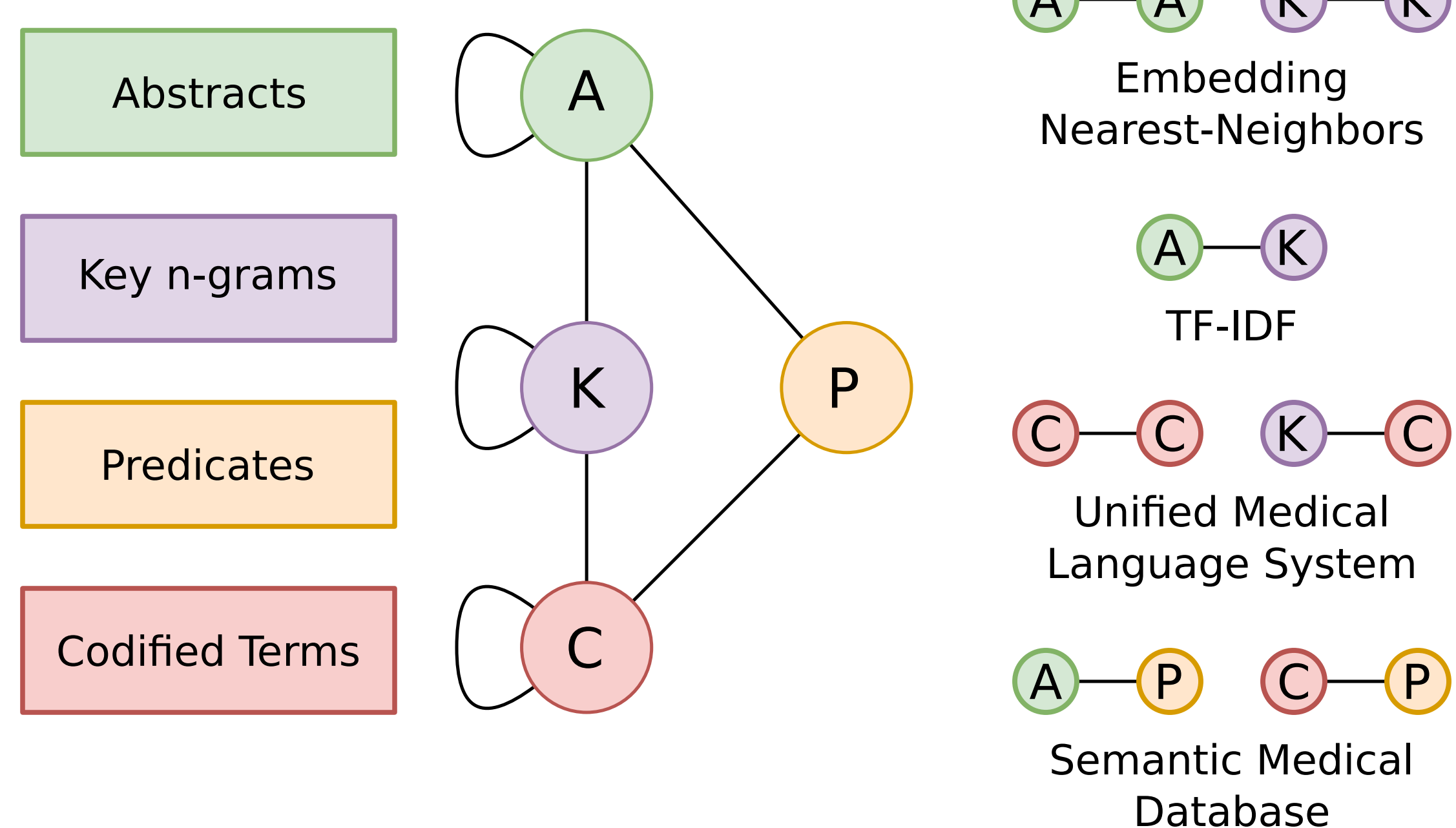
Hypothesis Generation

- Understand all available research
- Identify missed implicit connections
- Provide early information to decision makers

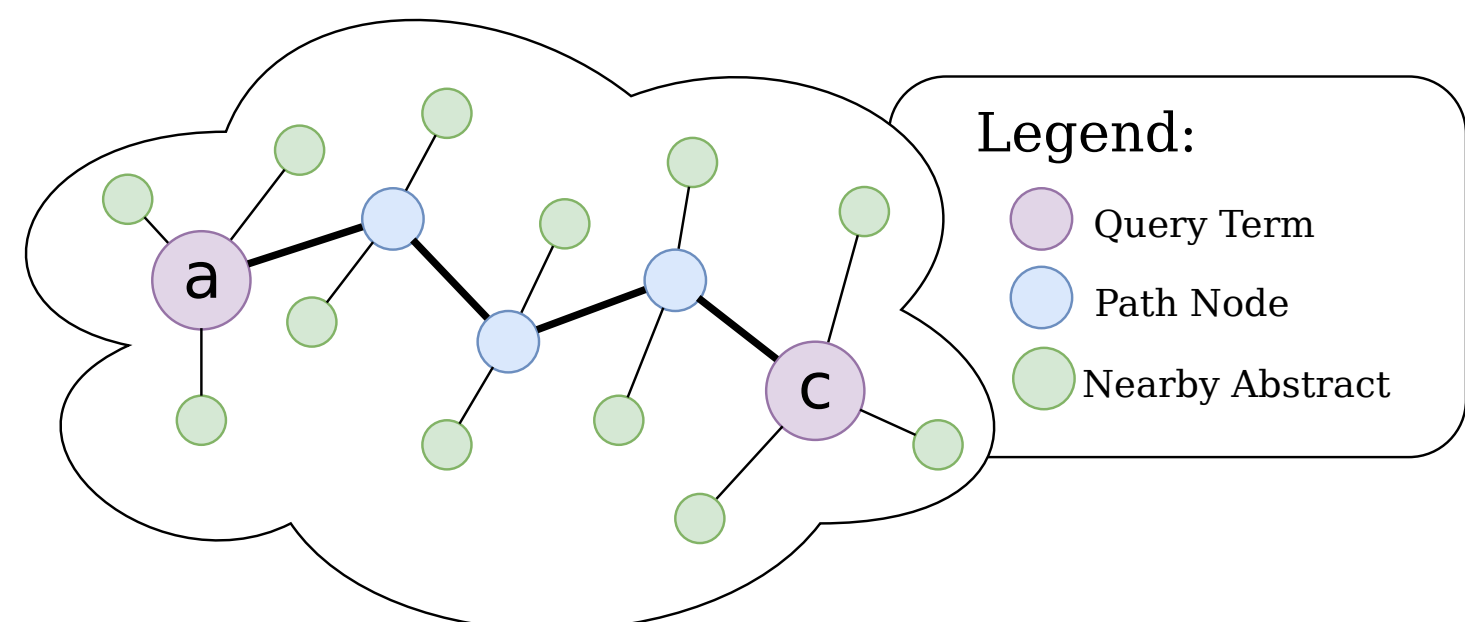


Generating Hypotheses Automatically

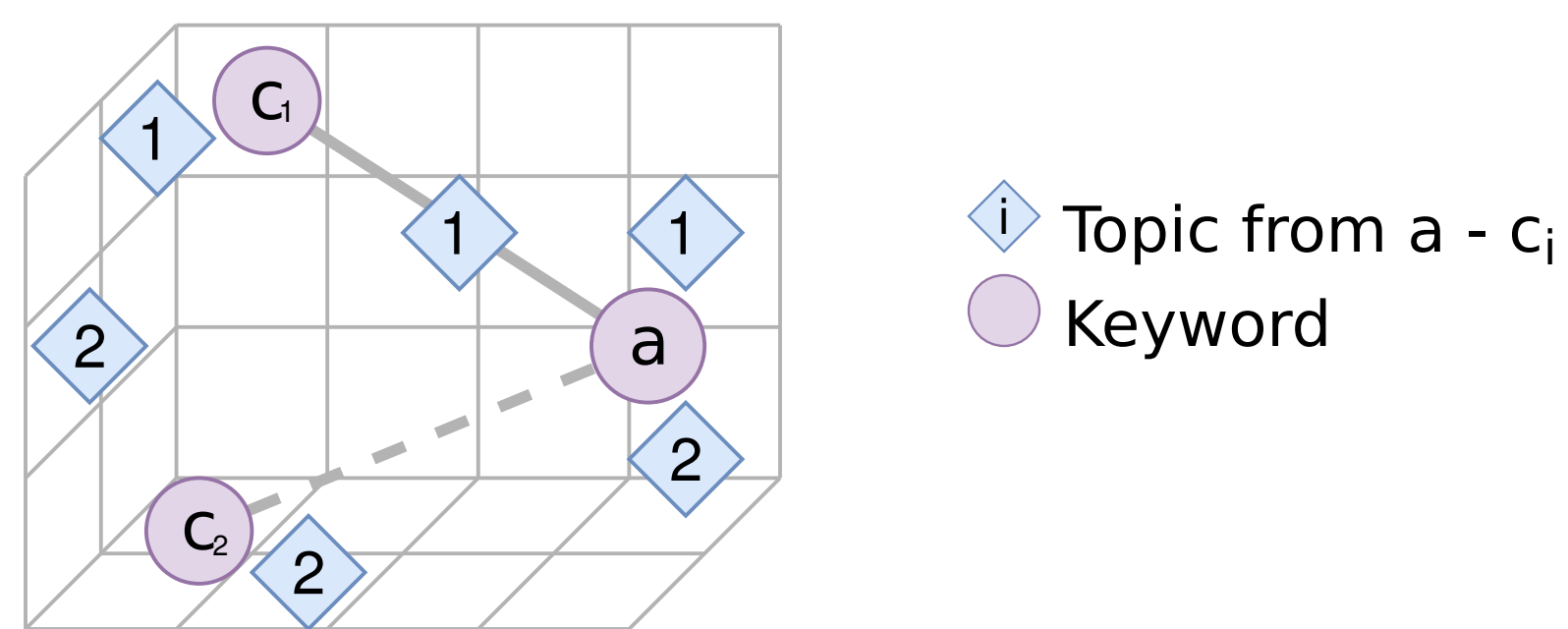
- Construct network with NLM sources



- Describe potential connections through shortest-paths

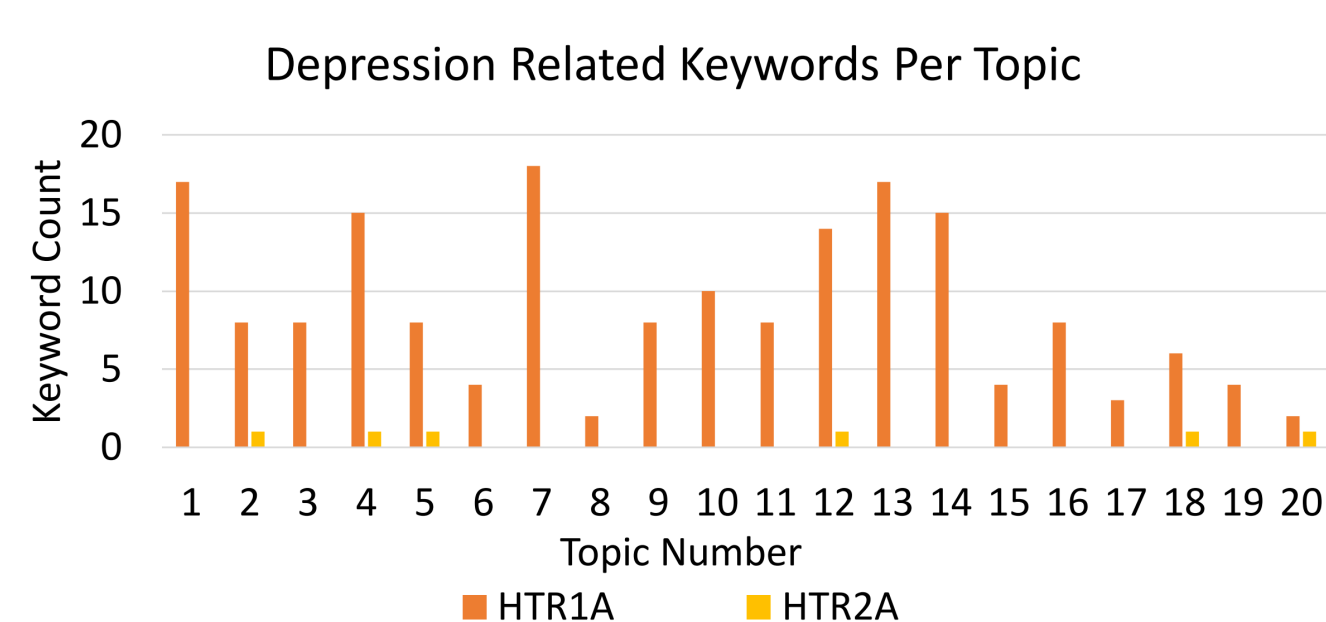
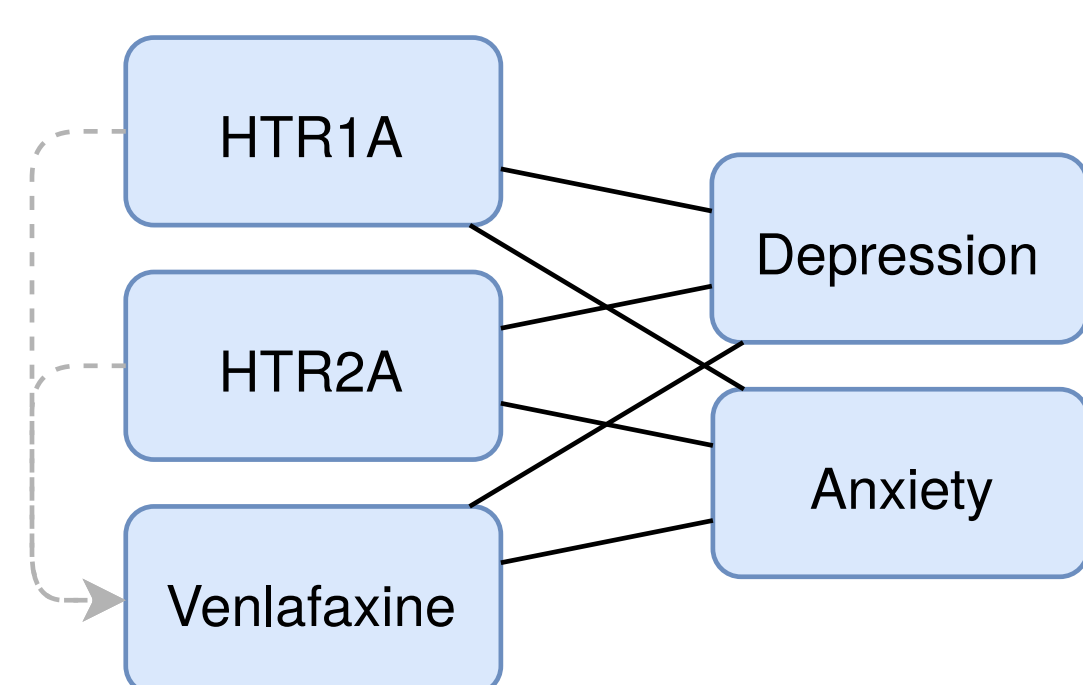


- Generate topic model from cloud and apply metrics



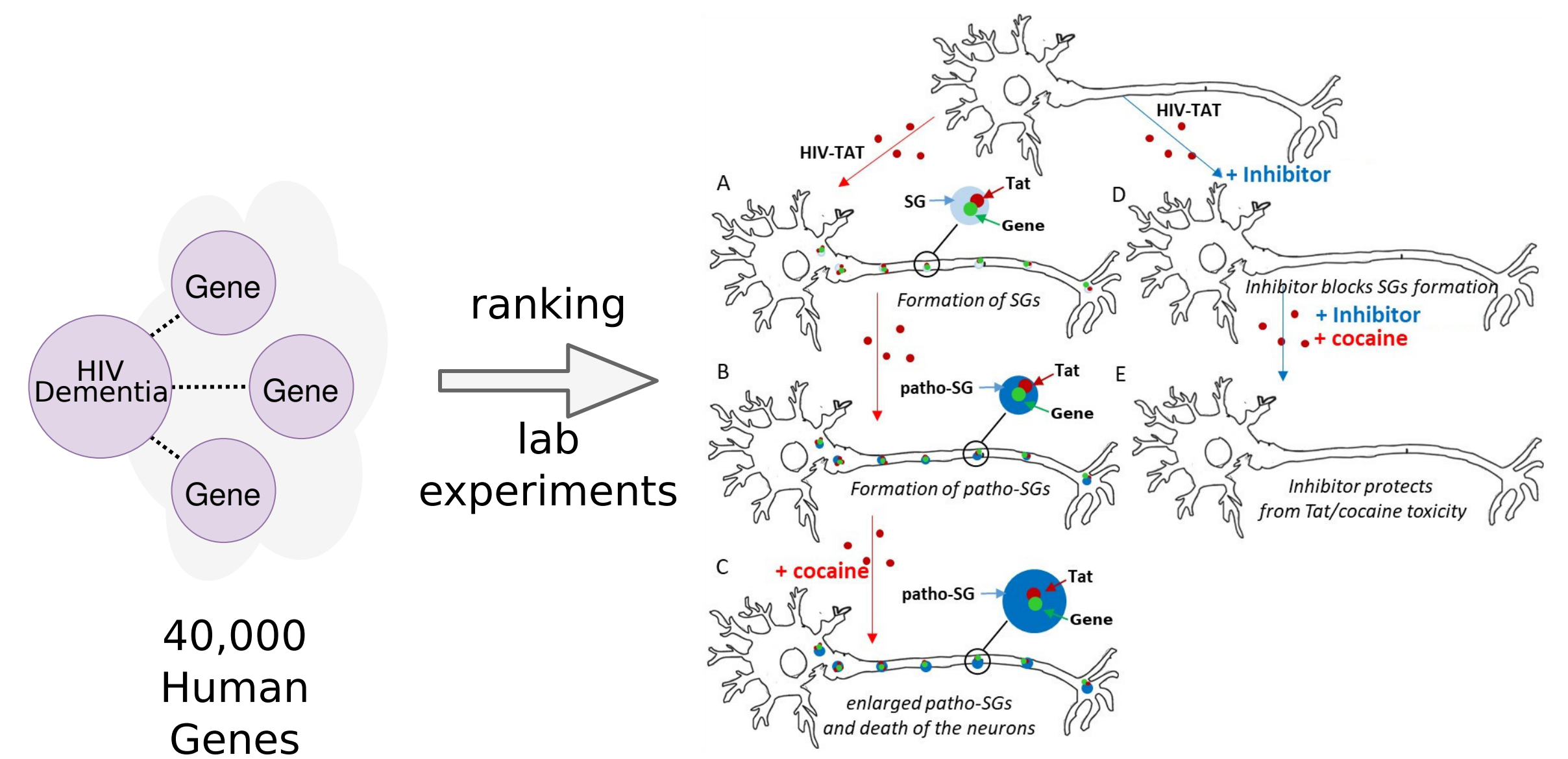
HTR[12]A & Venlafaxine

- Initial validation example
- Rediscover existing HTR1A connection
- Detect lack of HTR2A connection
- Uses pre-2010 data



Discovering New Treatments

- 25-33% of HIV+ patients develop HIV-Associated Dementia
- Apply MOLIERE to identify treatment options
- Rank all human genes & create candidate set
- Discover treatment option via lab experiments

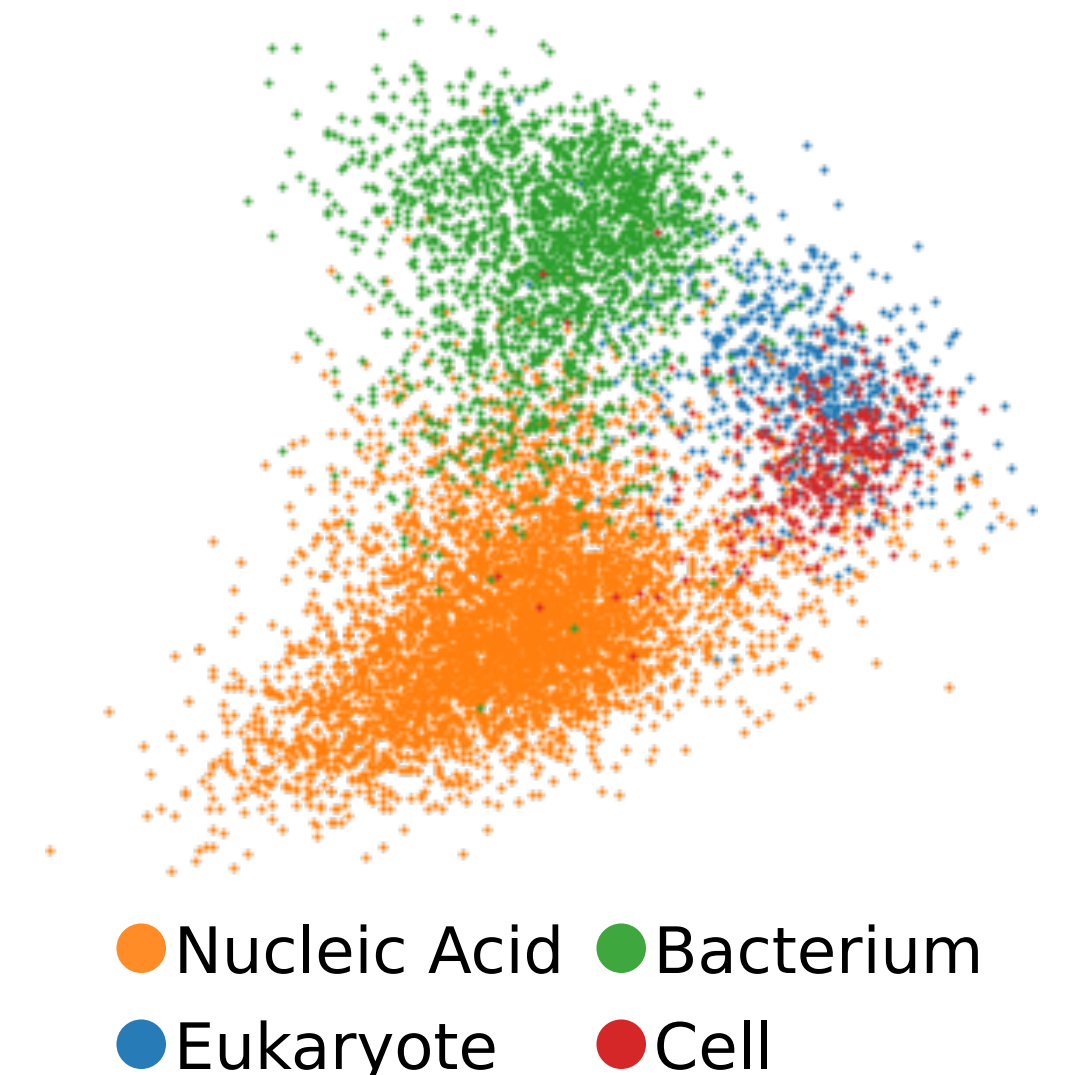


Ranking Hypotheses

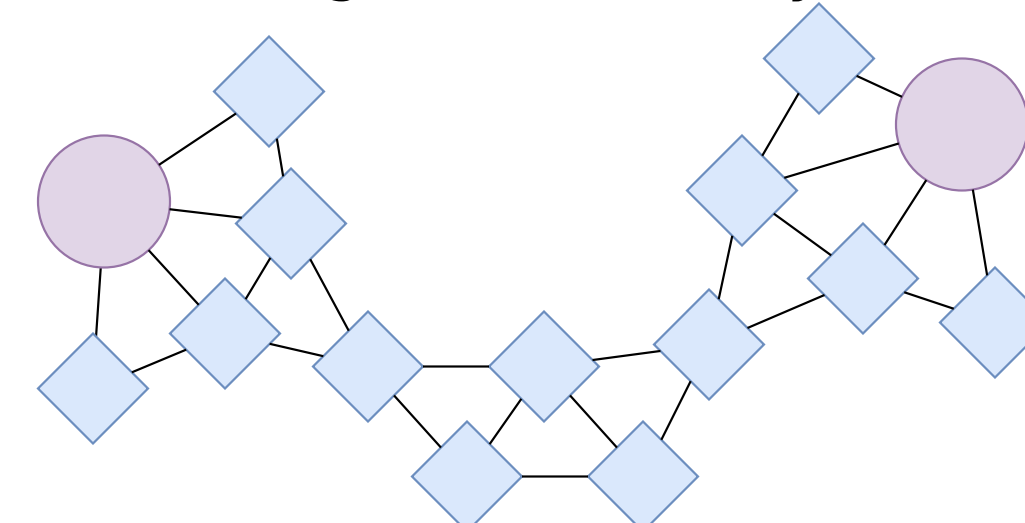
- Quantify connections through heuristics
- Derived via embeddings & topic network

- Example Metrics:

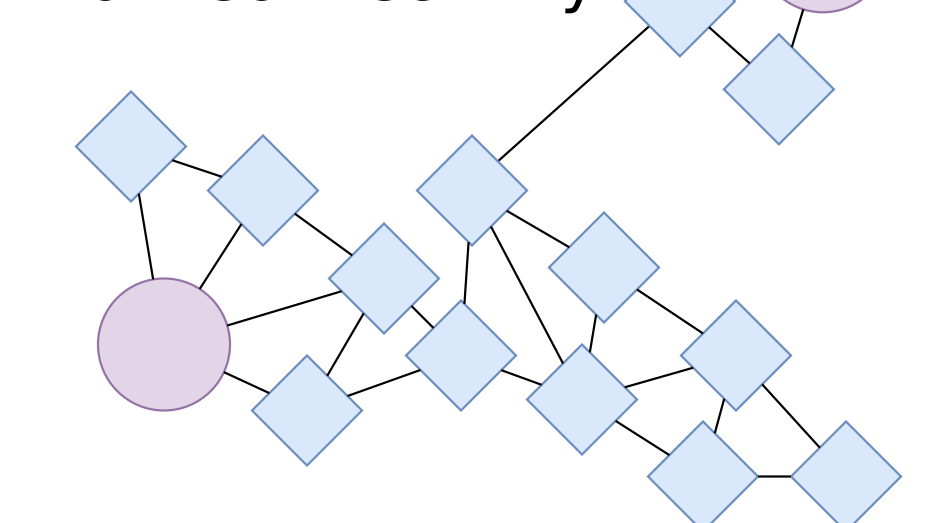
- Distance between terms
- Distance to topic centroids
- Correlation w.r.t. topic model
- Betweenness of topic network
- Topic network clustering coef.
- Polynomial combination



Published Connection High Connectivity

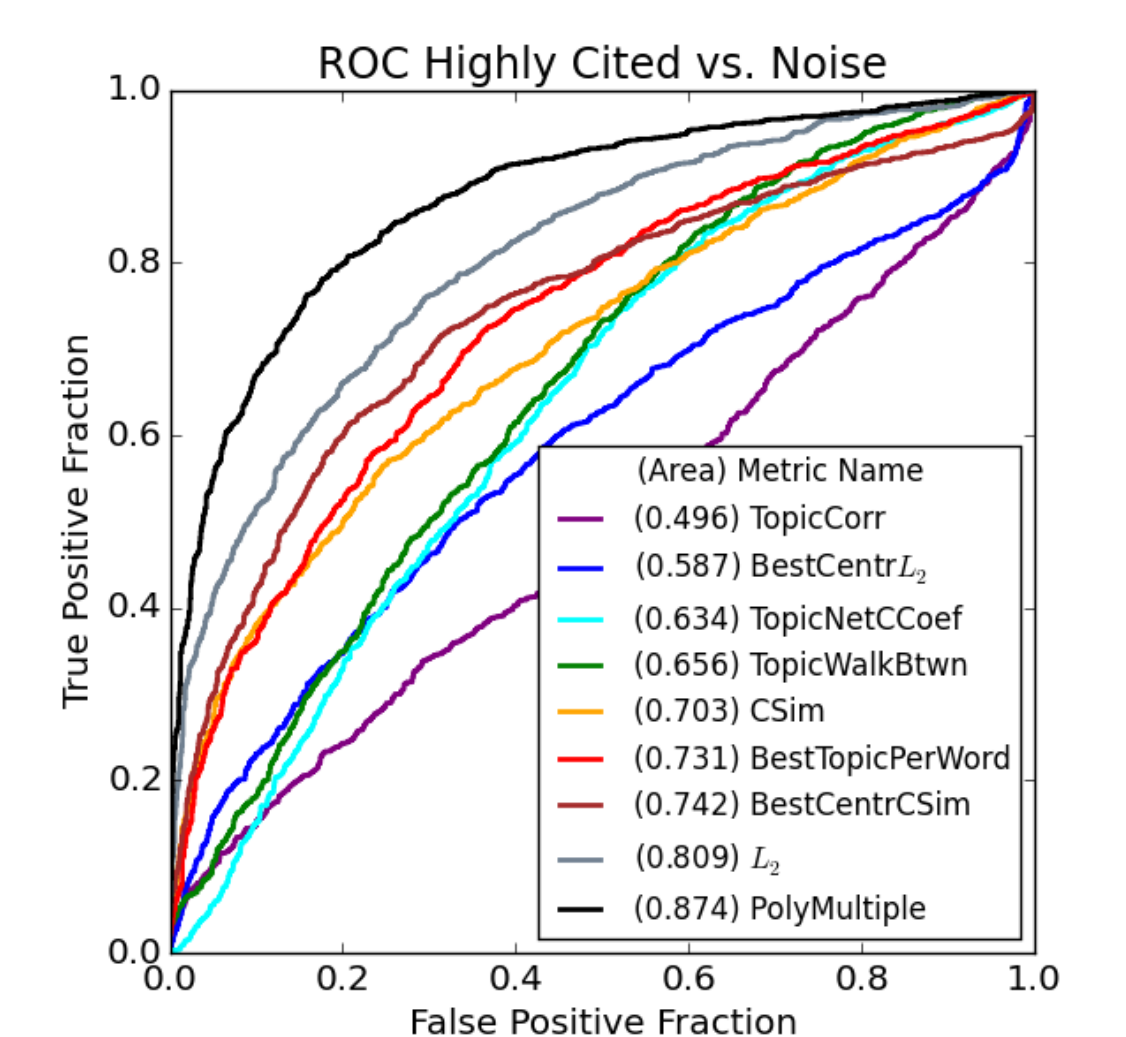
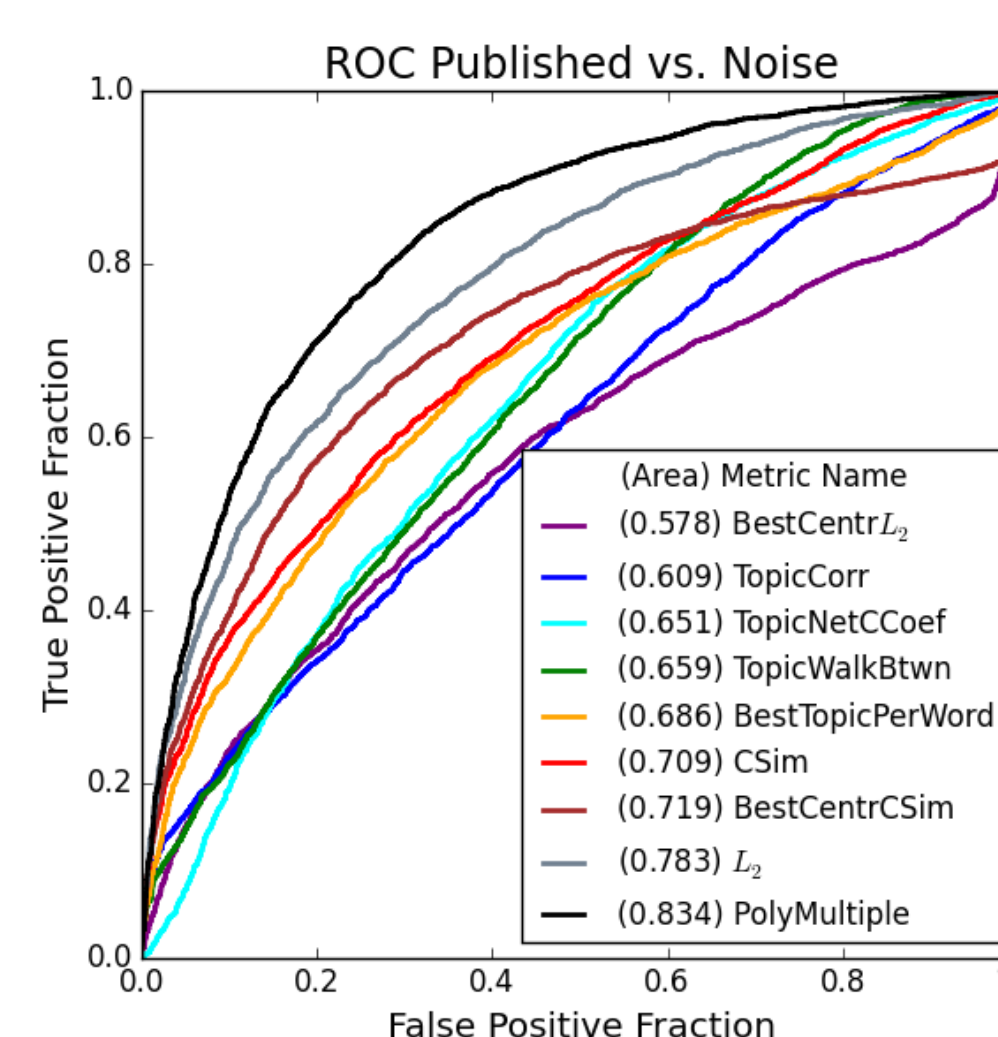


Noise Connection Low Connectivity



Large-Scale Validation

- Evaluate system without expert input
- Holdout experiment with historical data
- Evaluate thousands of potential hypotheses
- Rank via above metrics



See more here

