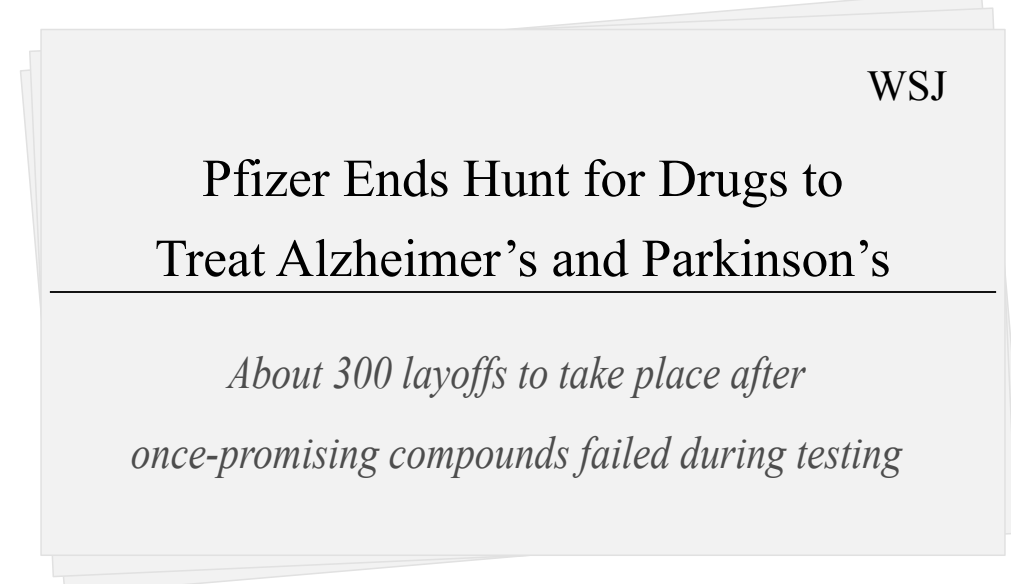


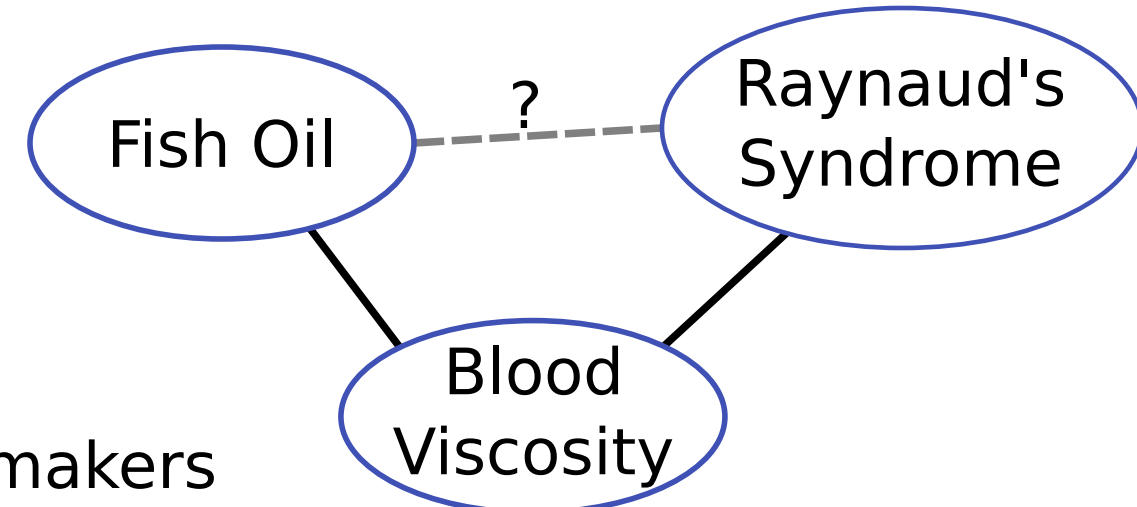
Problem

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



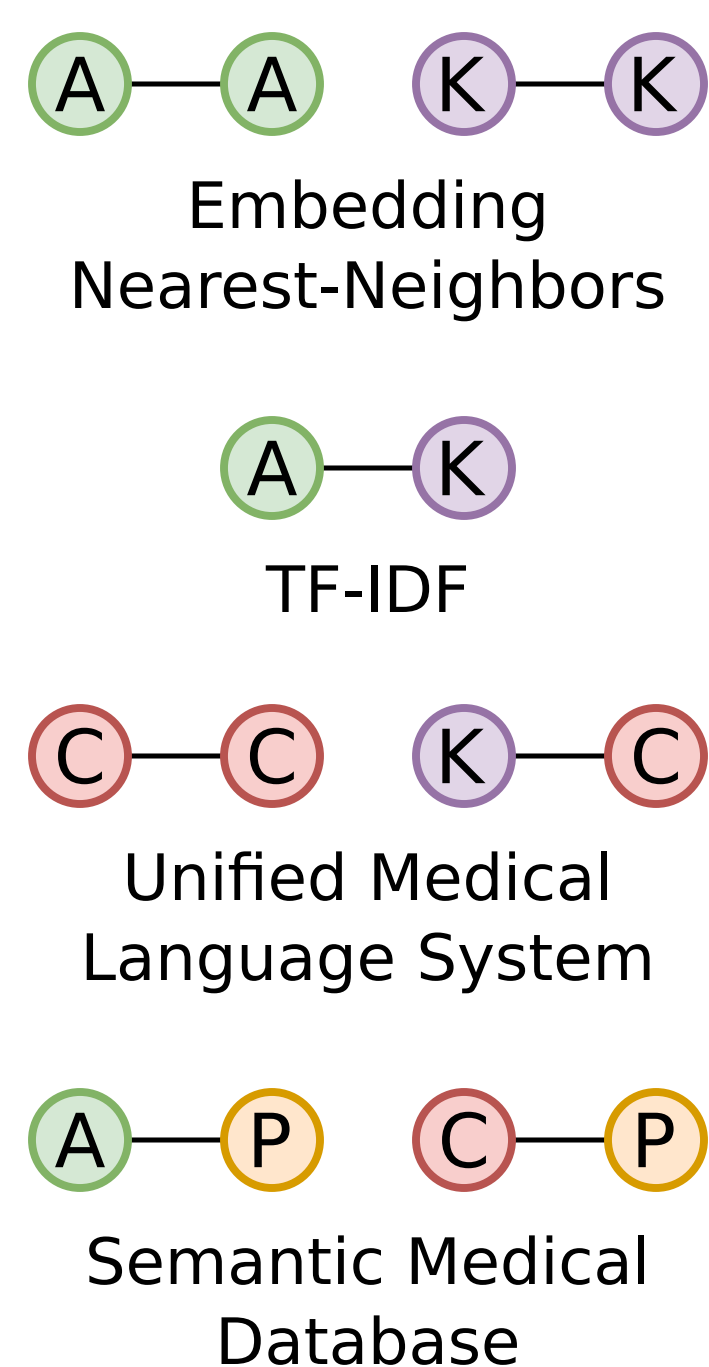
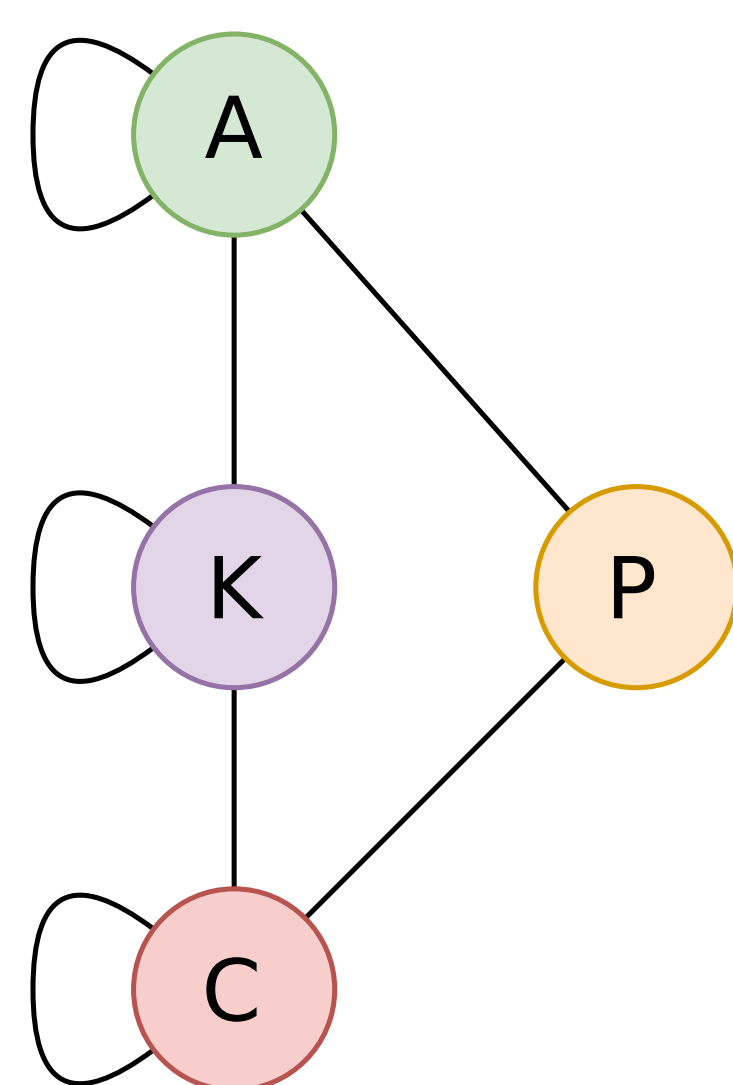
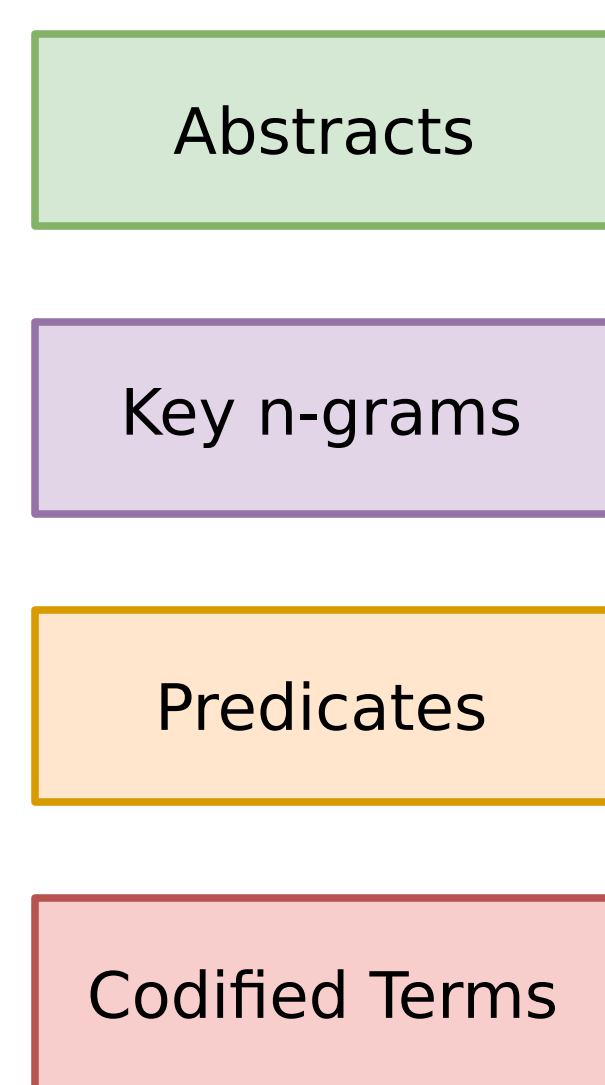
Hypothesis Generation

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

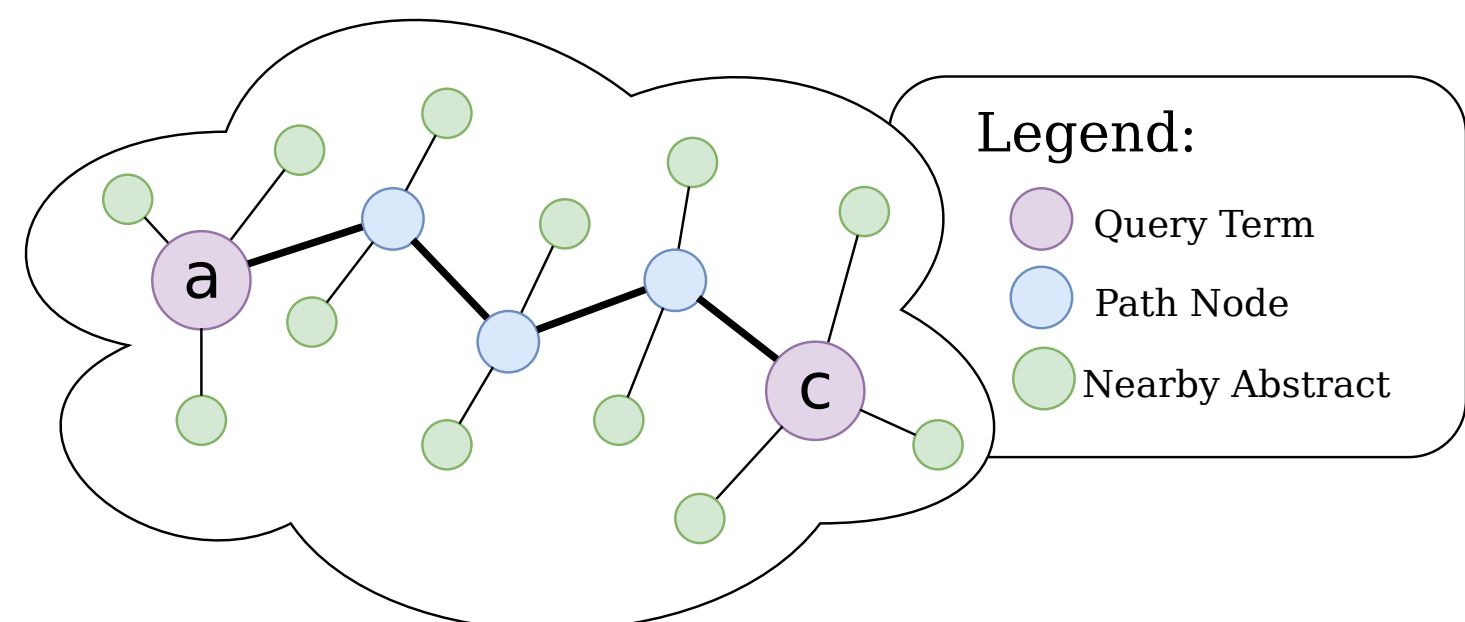


Generating Hypotheses Automatically

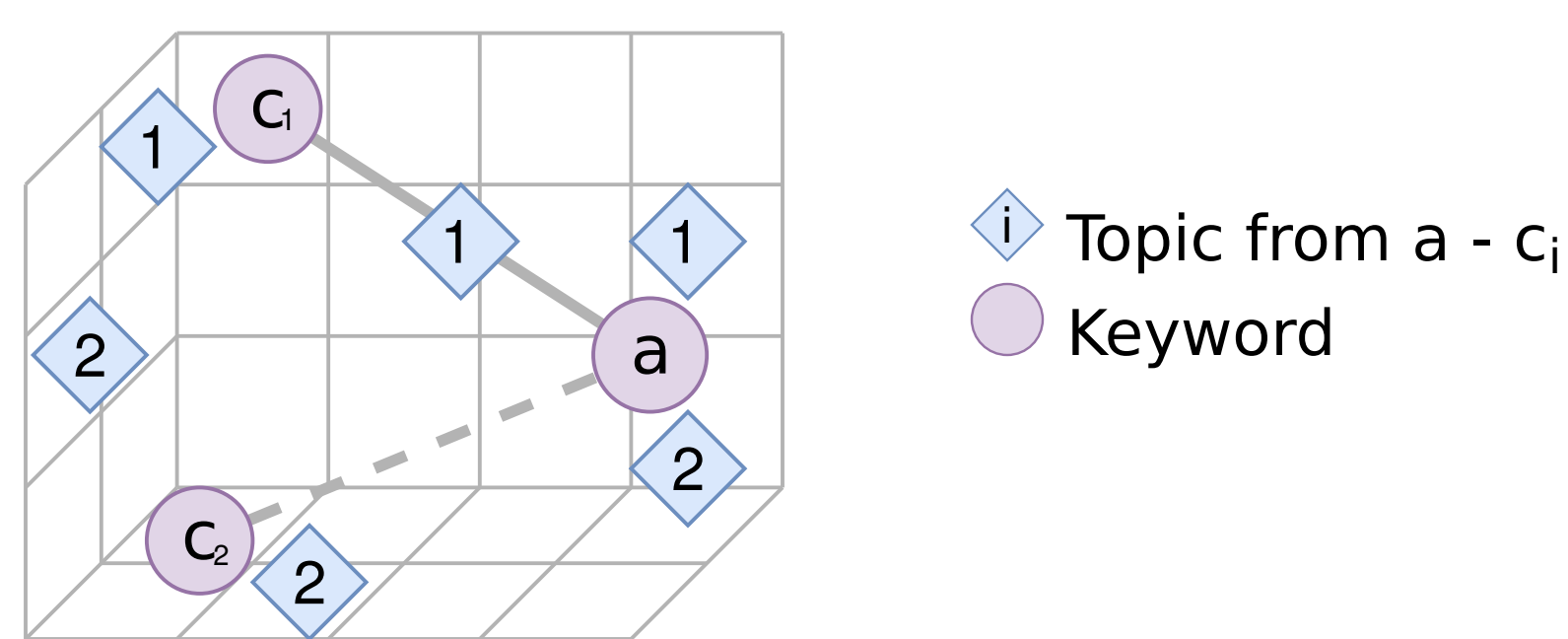
- Construct network with NLM sources



- Describe potential connections through shortest-paths

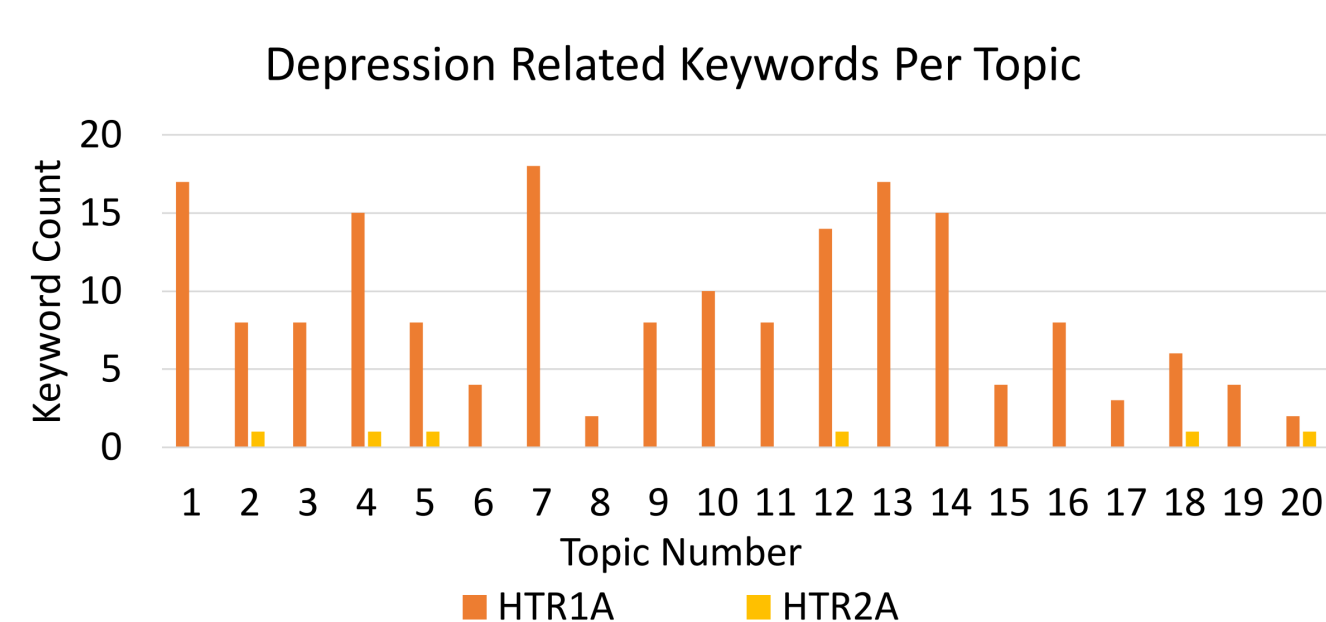
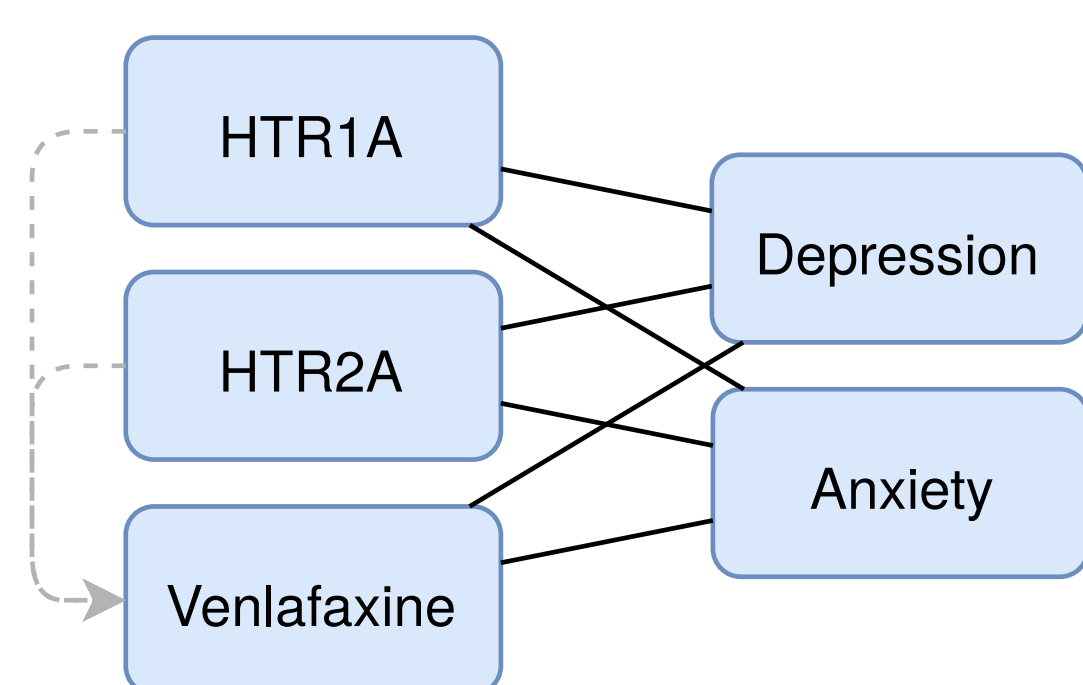


- Generate topic model from cloud and apply metrics



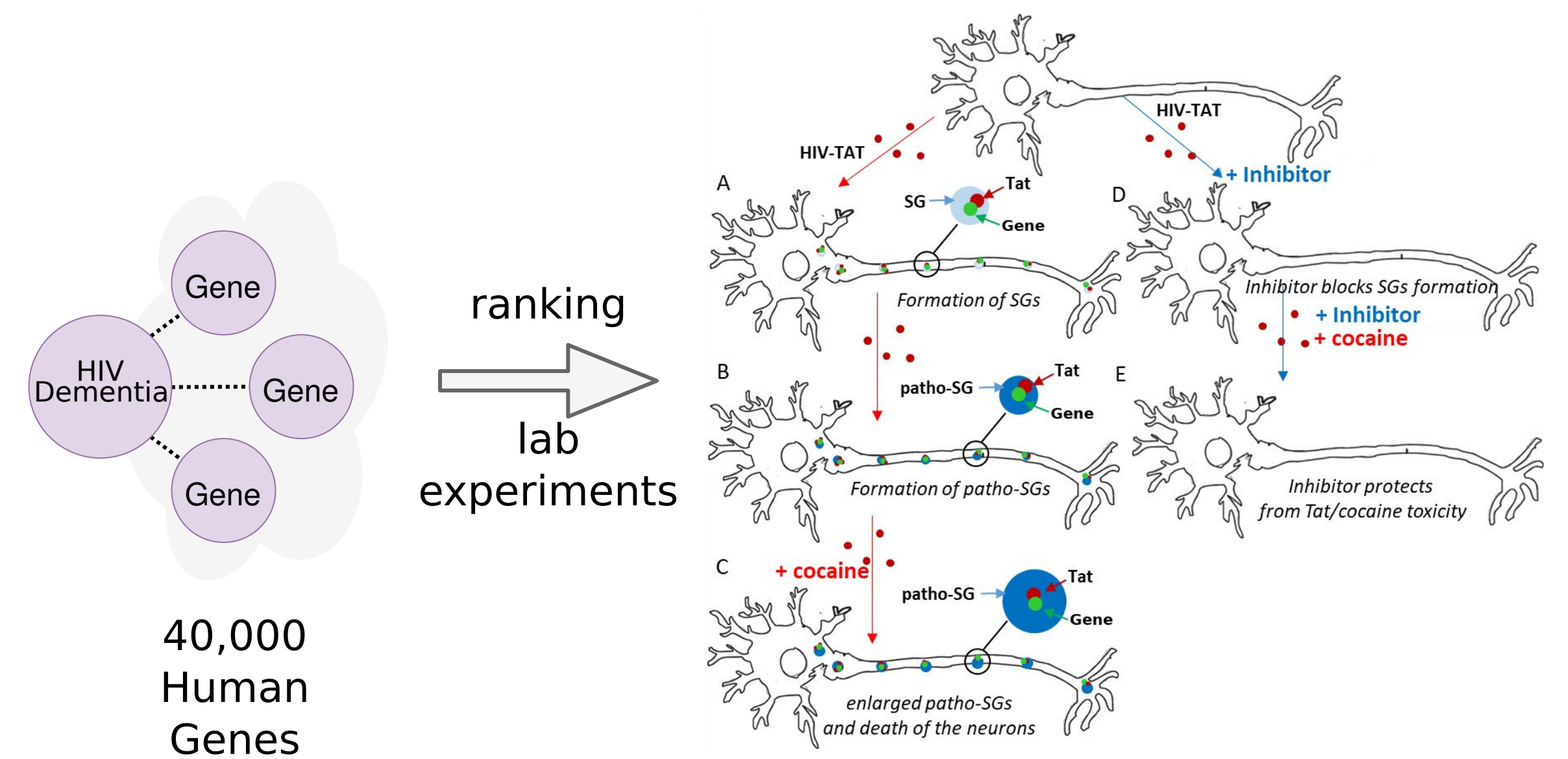
HTR[12]A & Venlafaxine

- 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

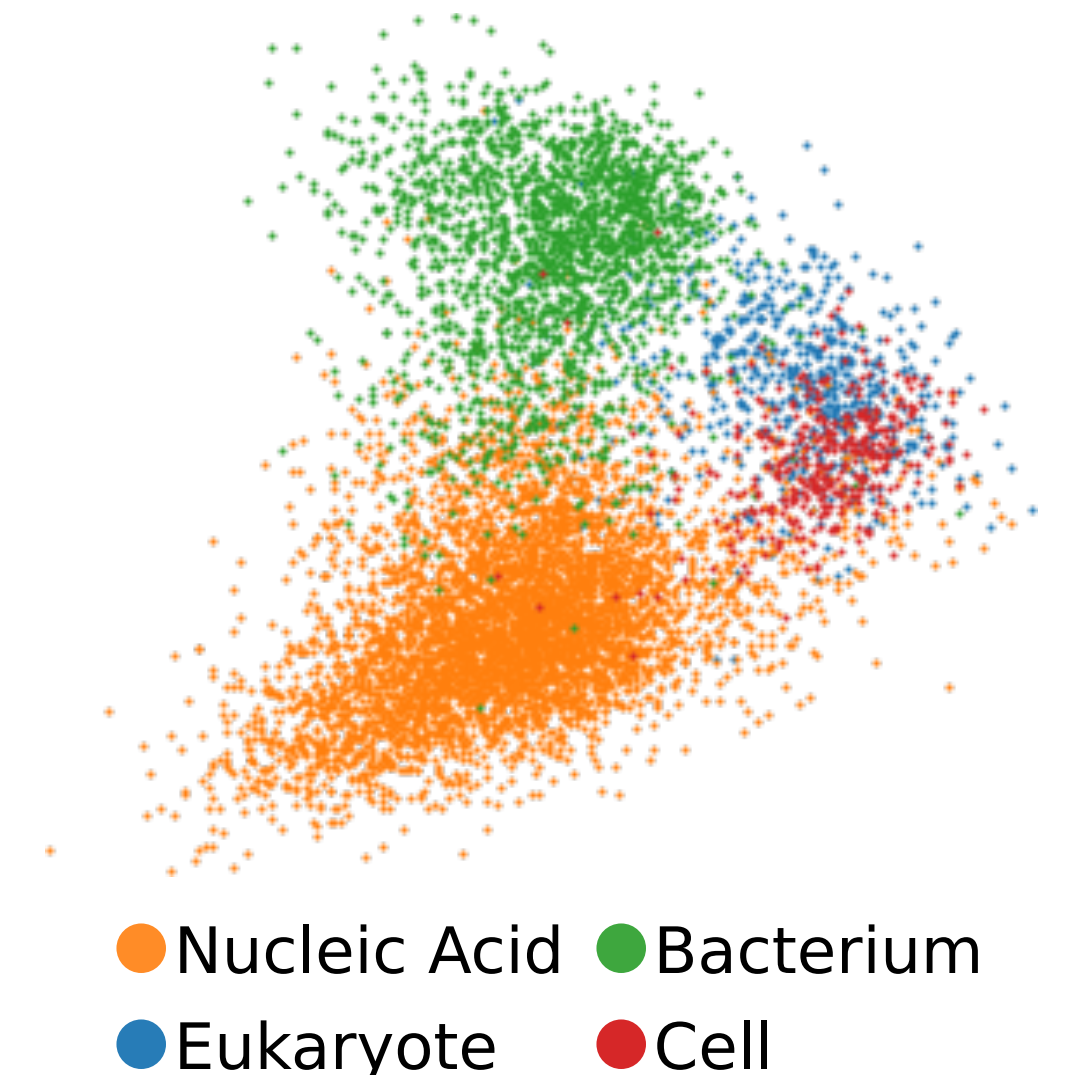


Ranking Hypotheses

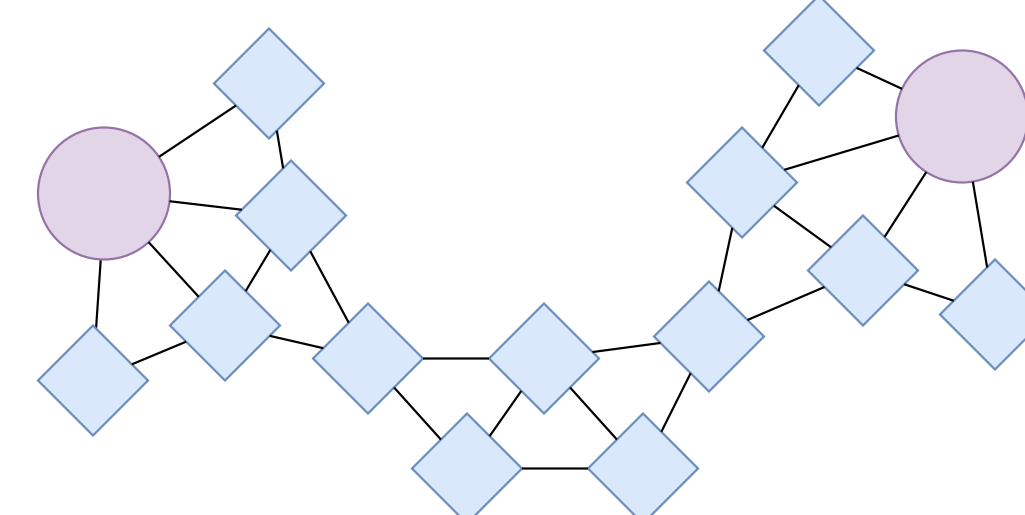
- Quantify connections through heuristics
- Derived via embeddings & topic network

- Example Metrics:

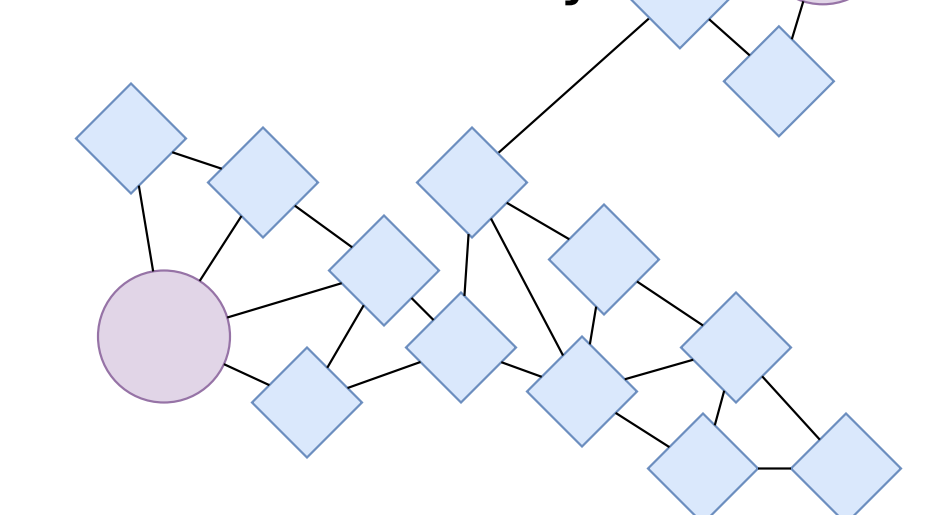
- Distance between terms
- Distance to topic centroids
- Correlation w.r.t. topic model
- Betweenness of topic network
- Topic network clustering coef.
- Polynomial combination



Published Connection High Connectivity

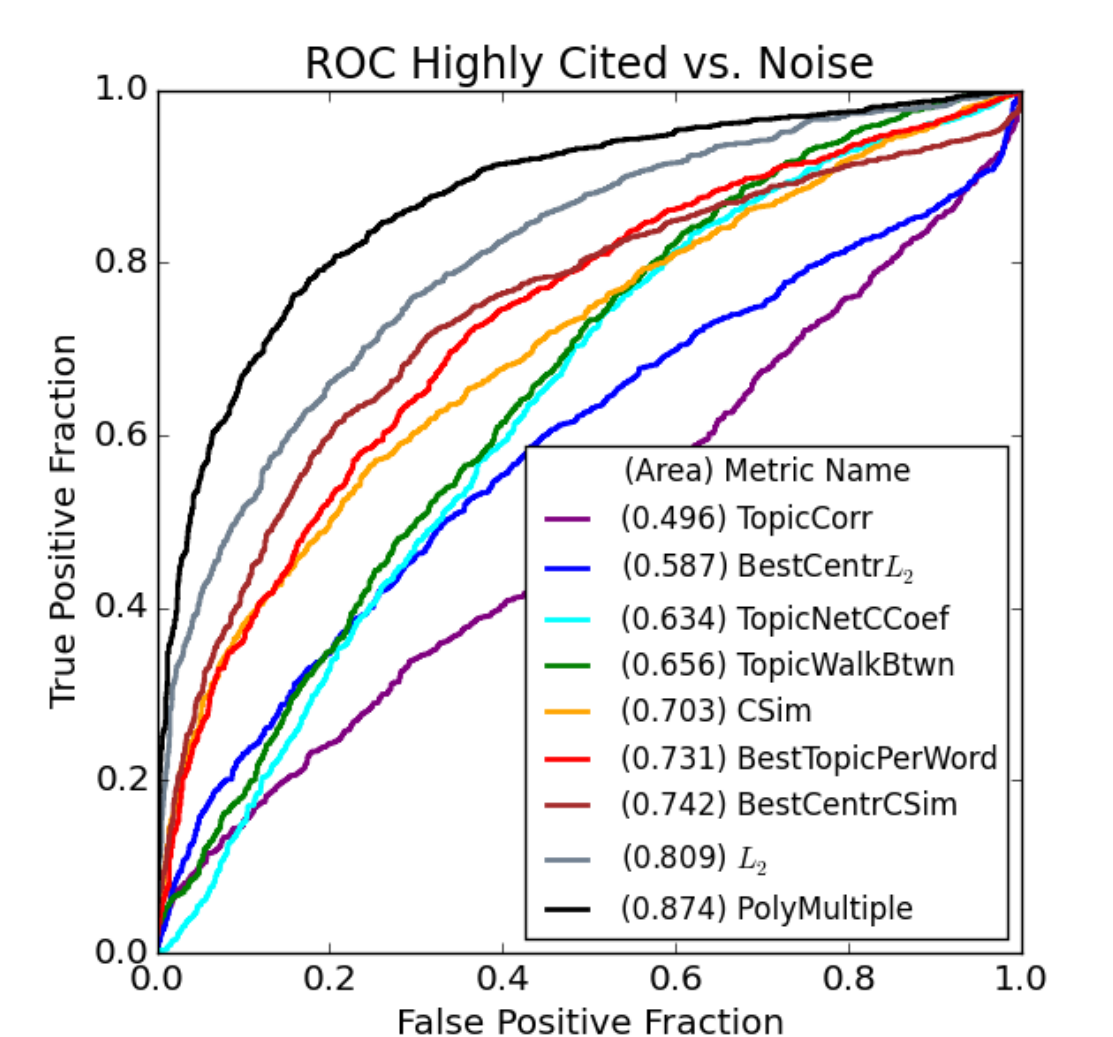
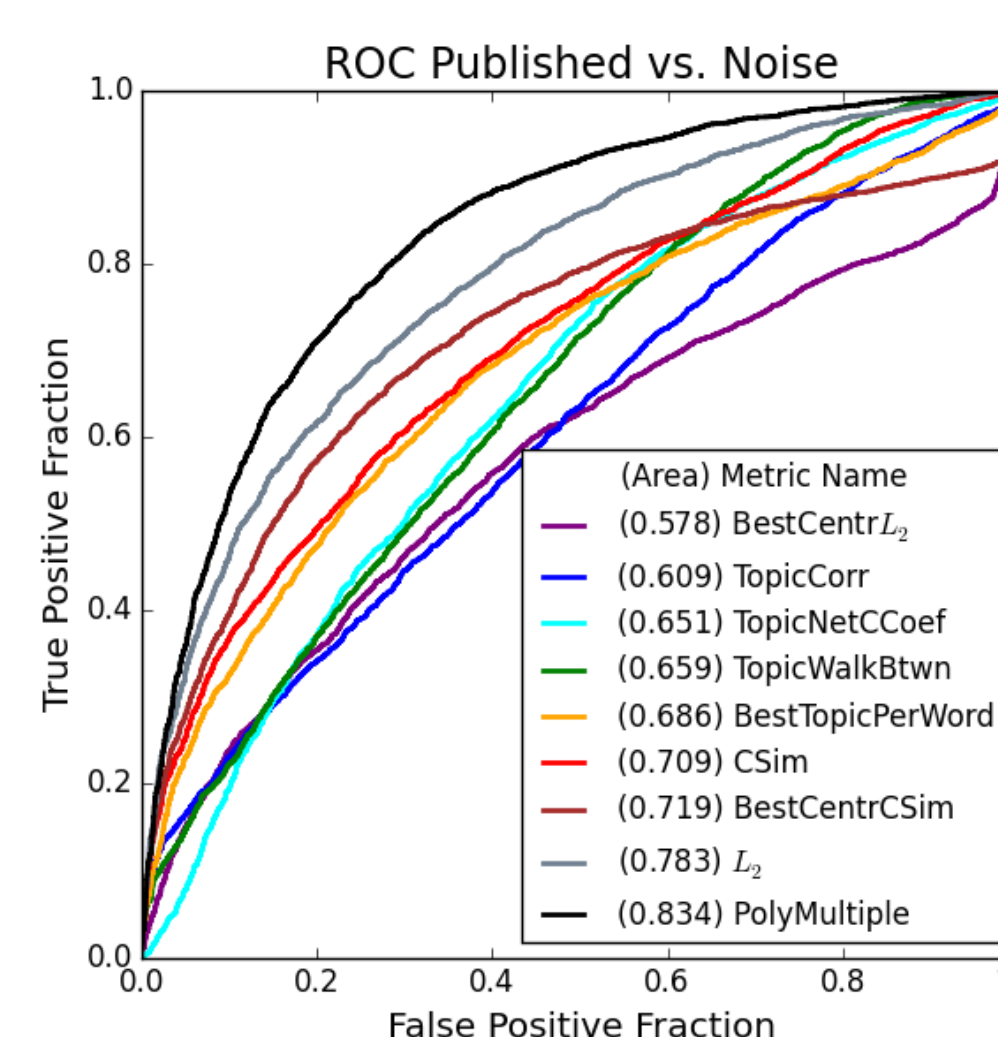


Noise Connection Low Connectivity



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

