

# SemanTex: Semantic Text Exploration Using document Links Implied by Conceptual Networks

Suad Aldarra(FI<sup>\*1</sup>), Emir Muñoz(FI<sup>\*1</sup>), Pierre-Yves Vandenbussche (FI<sup>\*1</sup>), Vít Nováček (Insight<sup>\*2</sup>)

## Summary

- Automatic method for computing semantic links between documents
- Provide the conceptual paths (explanation) between two documents
- Allow topical navigation within and through documents
- Application to biomedical literature (PubMed) Parkinson Disease article

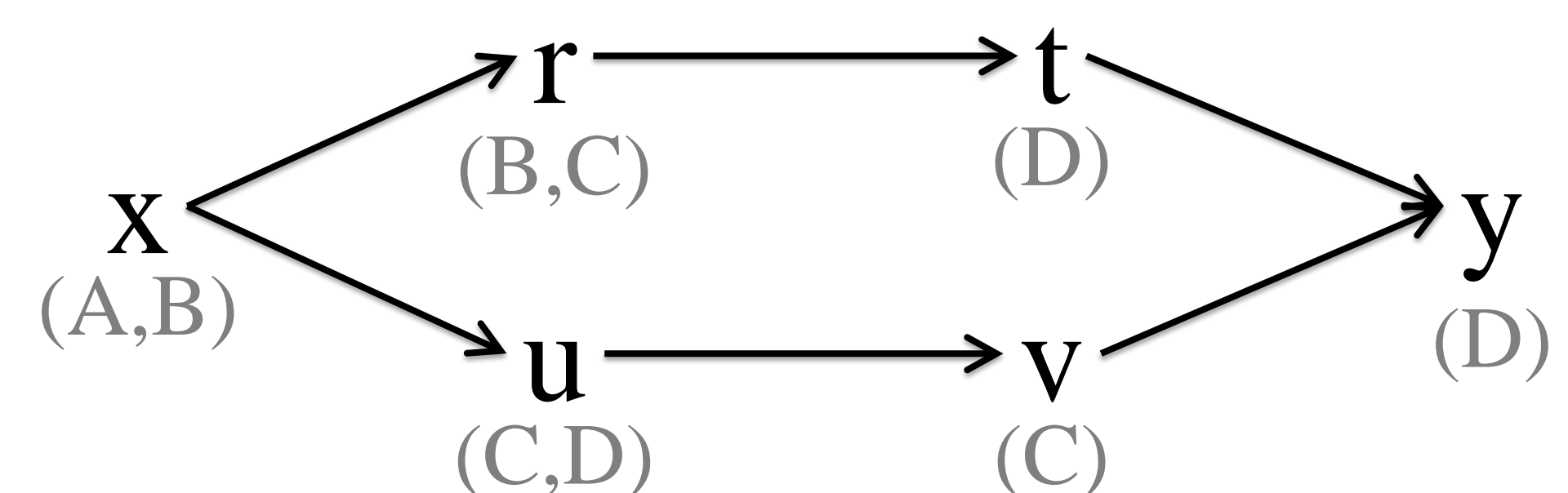
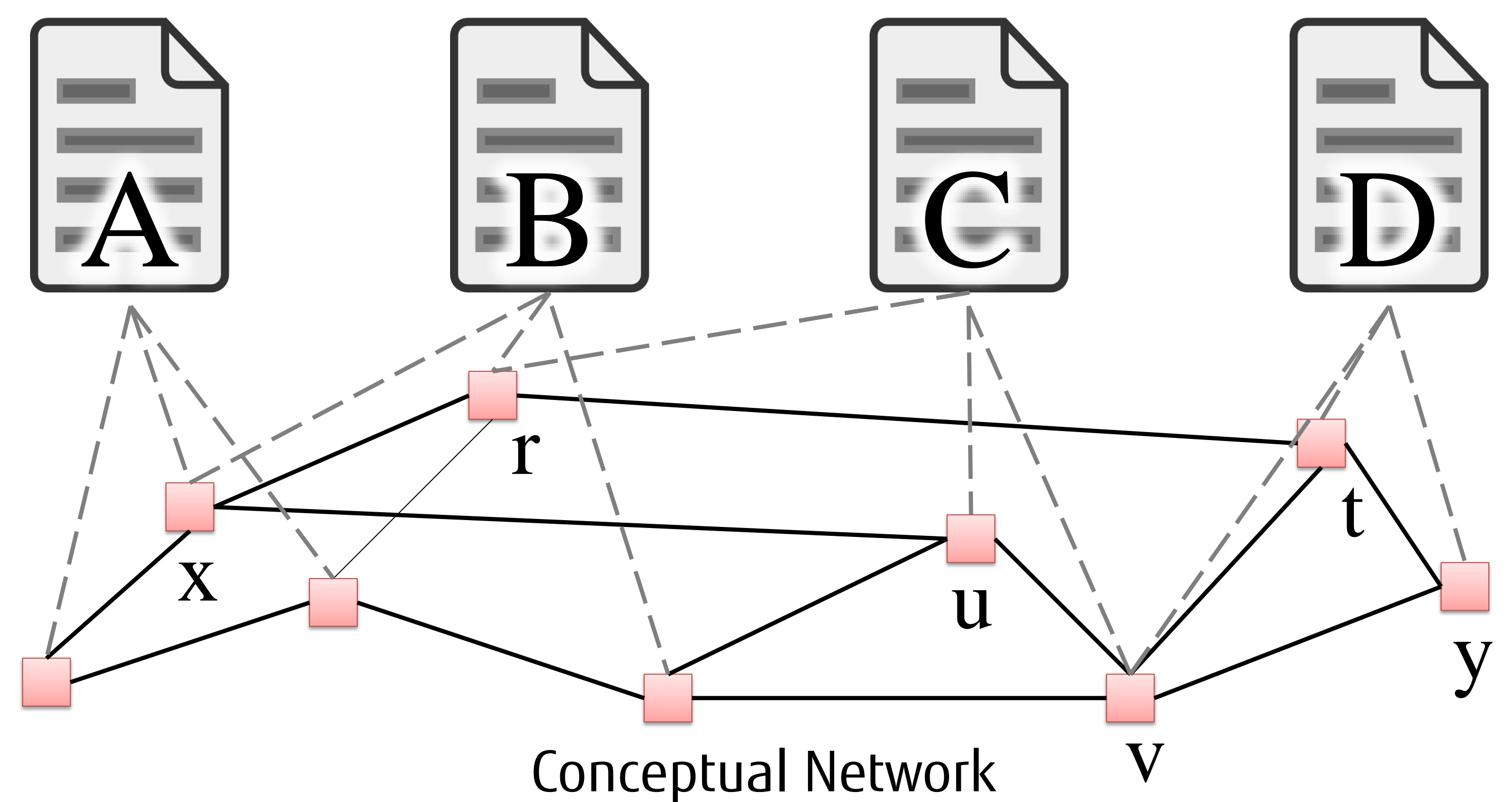
## Method

### Conceptual Network Extraction

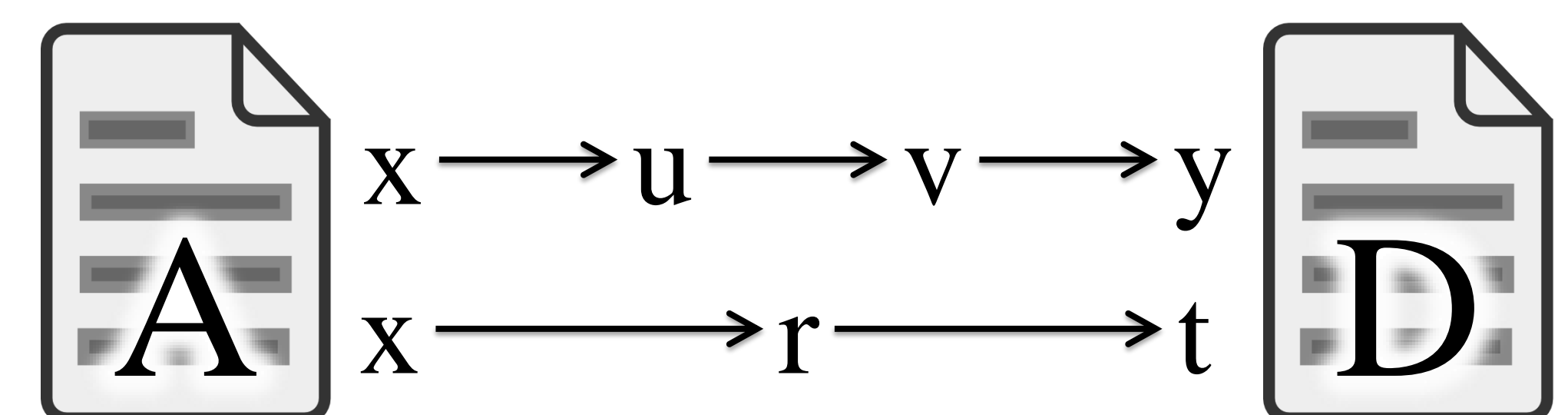
1. Noun-Phrase extraction  
(Using Biomedical Named Entity recognition)
2. Co-occurrence Relationship computation  
(Using Point wise Mutual Information)
3. Cosine similarity computation

### Paths Extraction and Selection

4. Extraction of paths between all nodes and between all documents  
(Using path length and product of the edges weight threshold)
5. Multi-objective optimization
  - Complexity
  - Coherence
  - Entropy



Most relevant Paths from x



Most relevant Paths between documents A and D

## Implementation and Evaluation

- Experimentation on 4722 abstracts from PubMed Parkinson disease related articles that led to the extraction of 43,362 concepts and 38M paths possible before paths selection
- Automated evaluation (TREC dataset: doc-doc relationship; MeSH: topical progression)
- Expert evaluation (Quality of paths selection)

Application to PubMed abstracts

Most relevant paths from "maternal transmission" concept