# Generating Text Summaries of Graph Snippets

Shruti Chhabra, Srikanta Bedathur

Indraprastha Institute of Information Technology, Delhi
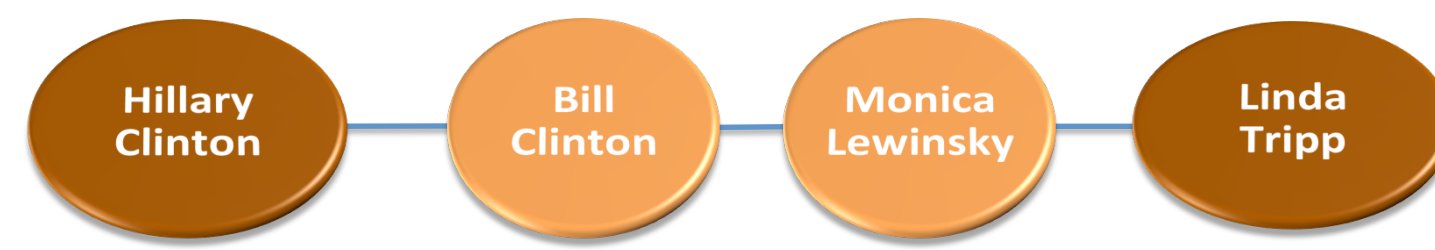
{shrutic, bedathur}@iiitd.ac.in

## ✧ Introduction

- o Large entity-relationship graphs available, rich source, knowledge graphs
- o "Given a set of entities, how to best characterize the relationships between entities?"
- o Recent studies – extract graphical sub-structures connecting entities from knowledge graph   **hard to interpret!**

**Entity-Centric Summarization:** Framework to generate summaries for graph structures (star graph or entity chain)
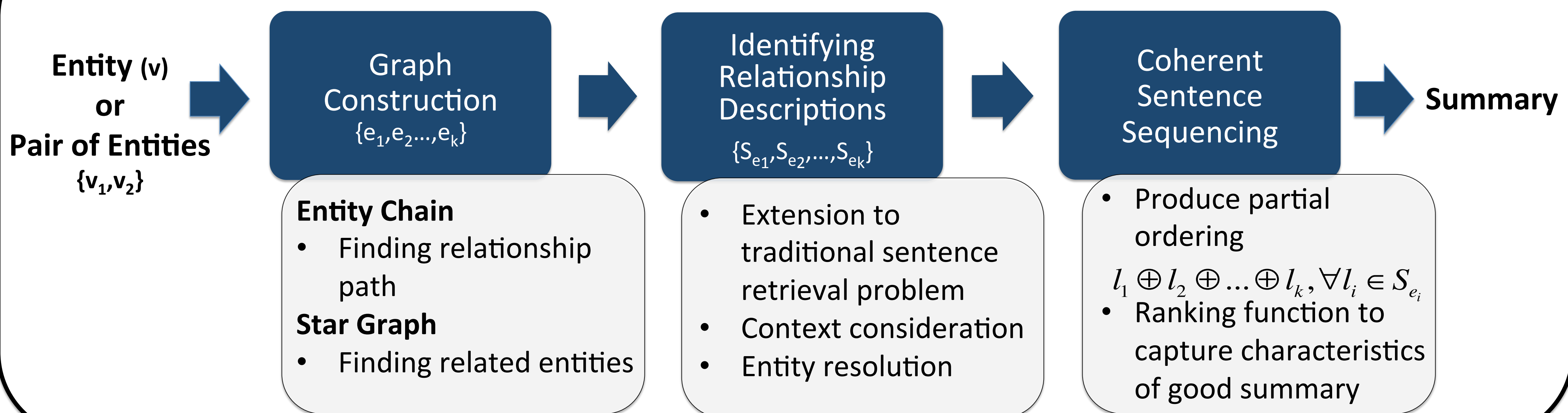
**Star graph for an entity**                     **Entity chain for pair of entities**

## ✧ Framework

**Entity (v)
or
Pair of Entities
{v₁,v₂}** → **Graph Construction {e₁,e₂...,eₖ}** → **Identifying Relationship Descriptions {S_{e1},S_{e2},...,S_{ek}}** → **Coherent Sentence Sequencing** → **Summary**

**Entity Chain**
- • Finding relationship path

**Star Graph**
- • Finding related entities

- • Extension to traditional sentence retrieval problem
- • Context consideration
- • Entity resolution

- • Produce partial ordering
  $$l_1 \oplus l_2 \oplus ... \oplus l_k, \forall l_i \in S_{e_i}$$
- • Ranking function to capture characteristics of good summary

## ✧ Preliminary Results

- o Assumption: given with the graph (star graph or entity chain)
- o Open IE tool is used to identify set of relationship descriptions for edges between entity pairs
- o Query set comprises of varied types of entities

### Summary of "Leonardo da Vinci"

Leonardo da Vinci was born in 1452, a hundred years after the exhibition.
Let 's trace the history of this rumor ... Following the death of his patron Giuliano deMedici , Leonardo left Italy and came to France in 1516 , at the age of 64.
Leonardo Da Vinci painted The Mona Lisa, Many have flocked to The Louvre to marvel at its beauty.
Leonardo died in France and is buried in the grounds of the chateau .
The record sale price of a nonliterary manuscript was \$ 30.8 million that Microsoft chairman Bill Gates paid for Leonardo da Vinci 's 16th-century scientific tract , the Codex Leicester , in 1994.

### Summary of Entity Chains

**Bill Clinton – Hillary Clinton – Monica Lewinsky – Linda Tripp**
Hillary is married to Bill , and Bill will come to the WH with her . News in the US is dominated by the allegation that Bill Clinton had a sexual relationship with Monica Lewinsky and persuaded her to lie under oath . Unknown to either Lewinsky or President Clinton , a former White House staffer Linda Tripp had recorded Lewinsky talking about her relations with President Clinton .

**Monsanta – Bt Cotton – Bacillus Thuringiensis**
Then in 1996 , Monsanto introduced BT cotton - a GMO that employs a gene from the bacterium Bacillus thuringiensis to make a powerful pesticide in the plant . Bt cotton , widely grown around the world , contains a gene from Bacillus thuringiensis , a bacterium species .

## ✧ Insights

- o Single sentence sufficient to describe entity pair
- o Relevance essential for coherence based ranking
- o Evidence of Redundancy
- o Sentence selection also depends on other edges
- o Sentence ordering crucial
- o Temporal aspect must be considered

## ✧ Research Challenges

- o How to cluster related entities and summarize separately? Can clustering be pushed to last stage?
- o How to quantize strength of relationships and thus rank related entities?
- o Which ranking model to use to find sentences describing relationship between entities?

## ✧ References

[1] X. Liu, Z. Nie, N. Yu, and J.-R. Wen. Biosnowball: automated population of wikis. In KDD, 2010.

[2] Rohini K. Srihari, Li Xu, and Tushar Saxena. Use of ranked cross document evidence trails for hypothesis generation. In KDD, 2007.