A Graph-based Model for Multimodal Information Retrieval

Serwah Sabetghadam

Vienna University of Technology

Supervisors

Andreas Rauber, Mihai Lupu

@serwah

sabetghadam@ifs.tuwien.ac.at

Rapid growth of the multimodal content

On average **350 million photos** are uploaded daily to Facebook

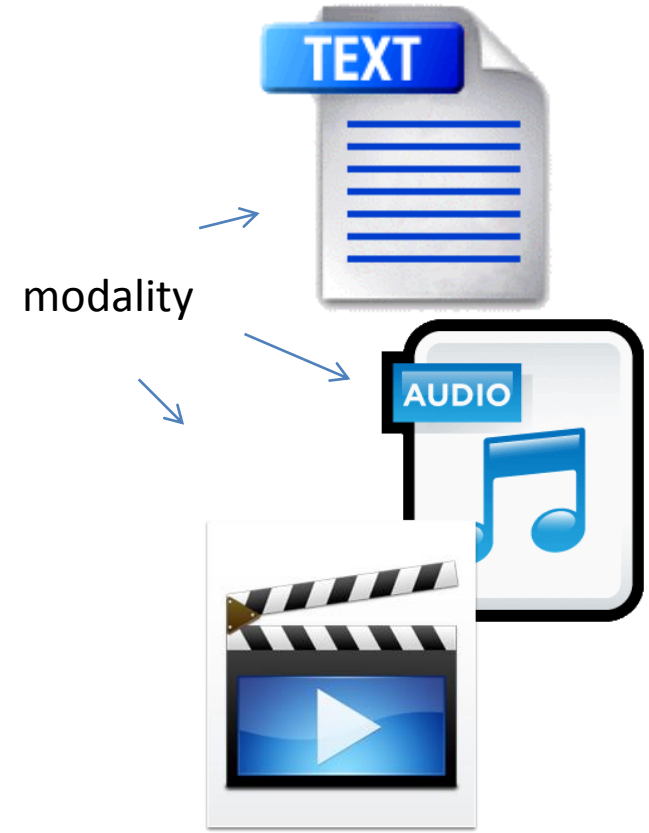
Multimodal Information Retrieval
has become a challenge

Multimodal Information Retrieval (IR)

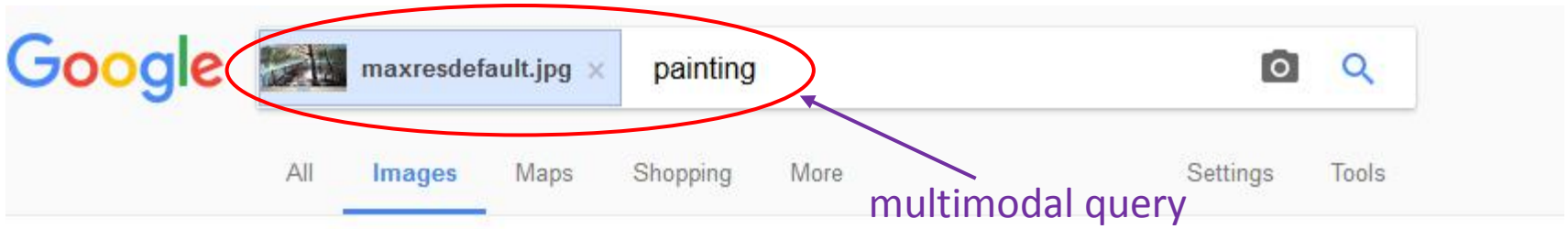
- Search for information of any modality with an information need that can be unimodal or multimodal

- Unimodal query: only keywords or an image example

- Multimodal query: a combination of images, video, or music files



An example of an information need:
find paintings like this image



About 25.270.000.000 results (1,01 seconds)

Image size:
1280 × 720

No other sizes of this image found.

Best guess for this image: ***Rain***

Visually similar images

Motivation (1)

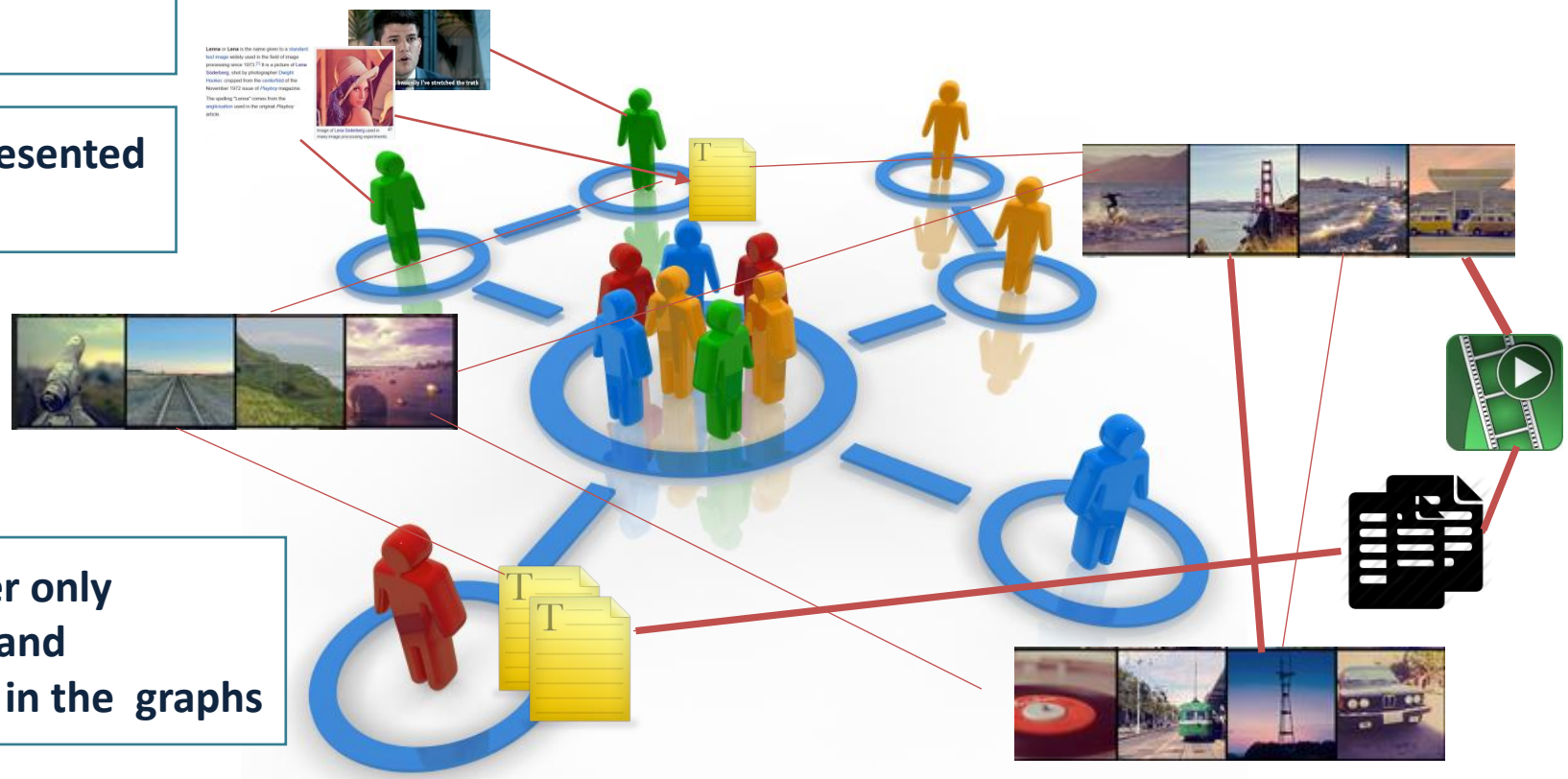
- Conventional text search to find multimodal result
  - E.g. Text based Image retrieval
  - Lack of indexing information of other modalities
    - Content-based image retrieval
  - Motivates to use different modalities

- Traditional IR does not include explicitly relations between docs
  - Documents are not isolated anymore
  - Hyperlinks, Metadata and Semantic connections
  - User-generated multimodal content, multimodal collections
  - Motivates to use structured IR too

Motivation (2)

Multimodal Data is Interlinked

Structured data represented by Graphs



Related work consider only one type of relation and one type of modality in the graphs

Motivation (2)

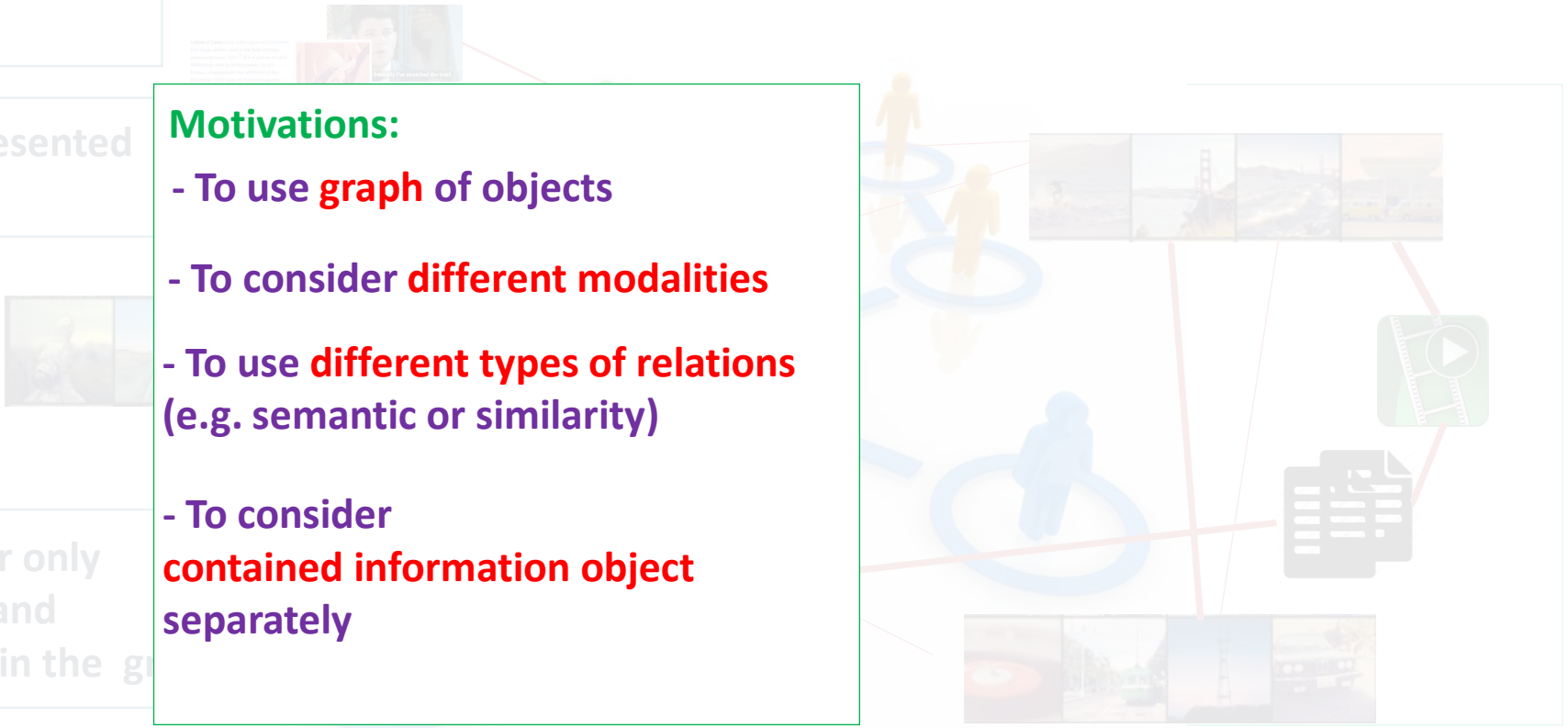
Multimodal Data is Interlinked

Structured data represented by Graphs

Related work consider only one type of relation and one type of modality in the gr

Motivations:

- To use graph of objects
- To consider different modalities
- To use different types of relations (e.g. semantic or similarity)
- To consider contained information object separately



Research Questions

- RQ1: How to define a graph-based model for multimodal information retrieval?
- RQ2: In such a graph model, can the relevant nodes be reached?
- RQ3: In such a model can scores identify the relevant nodes?

Contributions

Model

RQ1: How to define a graph-based model for multi-modal information retrieval (MMIR)?

Contributions

Astera model for MMIR

Defined search based on facets

[ECIR Workshop 2013]
[IRFC 2014]
[ICMR 2014]

Reachability

RQ2: In such a graph model, can the relevant nodes be reached (recall)?

Contributions

Analysis of the effect of different facets and links on reachability of relevant nodes

[ECIR 2015]

Precision

RQ3: In such a model can scores identify the relevant nodes?

Contributions

Analysis of the effect of different facets on precision.

The effect of Query-dependent and –independent routing on precision

[CLEF 2014]
[GBS 2015]
[Keystone 2016]

Contributions

Model

RQ1: How to define a graph-based model for multi-modal information retrieval (MMIR)?

Contributions

Astera model for MMIR

Defined search based on facets

[ECIR Workshop 2013]
[IRFC 2014]
[ICMR 2014]

Reachability

RQ2: In such a graph model, can the **relevant** nodes be **reached**?

Contributions

Analysis of the effect of different **facets** and **links** on **reachability** of relevant nodes

[ECIR 2015]

Score Analysis

RQ3: In such a model can **scores** identify the relevant nodes?

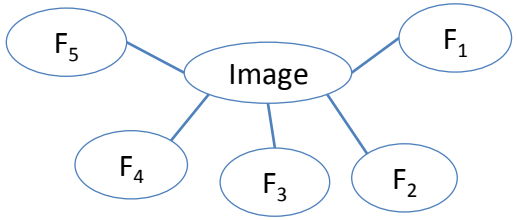
Contributions

Analysis of the effect of different facets on precision.

The effect of Query-dependent and –independent routing on precision

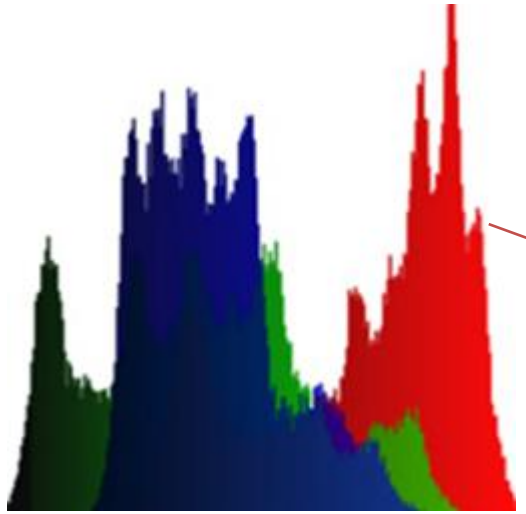
[CLEF 2014]
[GBS 2016]
[Keystone 2016]

An Information Object and Different Representations/Facets



An information object may have several representations or facets.

Facet: an inherent feature or property of an information object



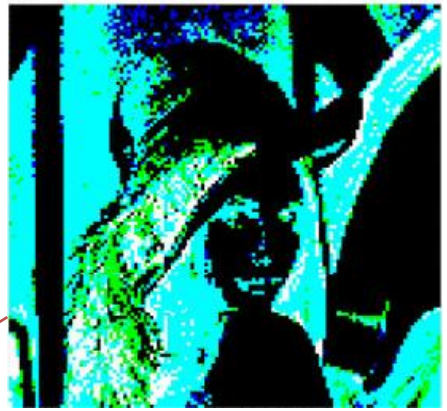


Image of Lena Söderberg used in many image processing experiments.

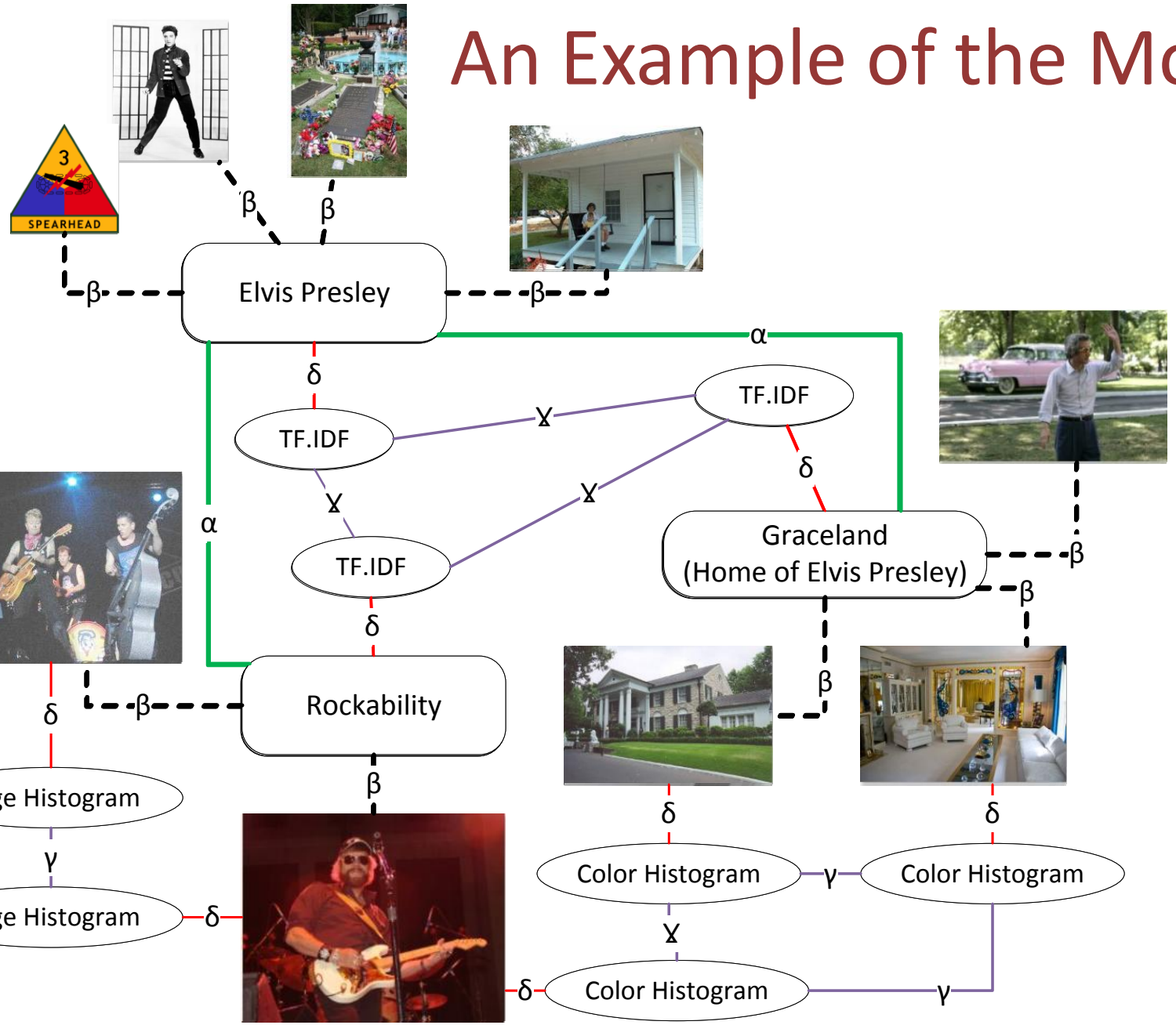
RQ1: How to define a graph-based model for multi-modal information retrieval?

Model

- We propose a graph of information objects $G = (V,E)$ named Astera
  - V is the set of vertices (nodes/facets)
  - E is the set of edges (defined by different types of relations)
- Different types of relations
  - Semantic (α): any semantic relation between two objects
  - Part-of (β): an object as part of another object, e.g. an image in a document
  - Similarity (⧖): relation between the facets of the same type of two information object
  - Facet (δ): linking an object to its facets

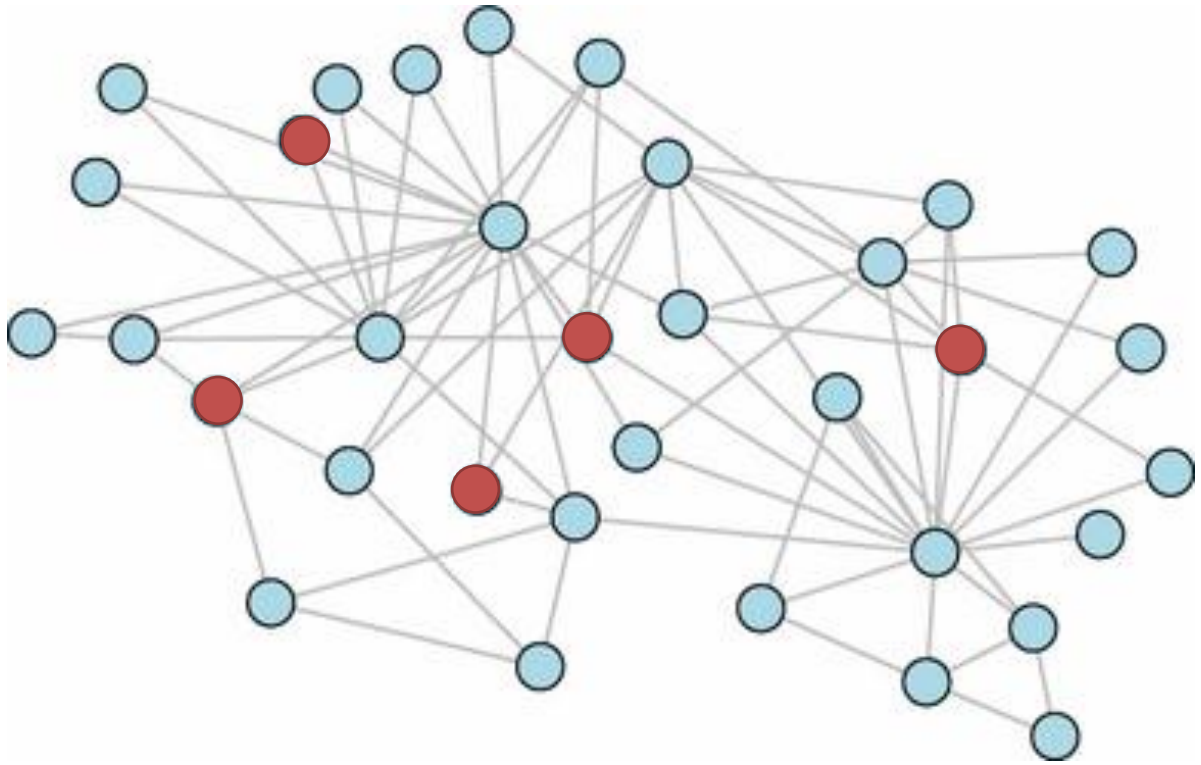
S. Sabetghadam, M. Lupu, and A. Rauber, "Astera - A Generic Model for Multimodal Information Retrieval", in Proceedings of Integrating IR Technologies for Professional Search Workshop, held in ECIR 2013, pp. 551-554.

An Example of the Model



Hybrid Search

- Standard search
  - Text: Lucene
  - Image: LIRE
  - Any similarity computation framework



- Graph search
  - Start from top results of standard search and traverse the graph
  - We take top 20 results of each facet

Relevance Score Value Function (RSV)

Model

$$RSV(q,u) = \sum_{i=1}^{l} norm(sim(q_{f_i}, u_{f_i})) \cdot w_{f_i}$$

Normalization function

Similarity function

Weight of facet f_i

- Reaching a node we calculate the similarity of different facets with the query facets
- The result is a score given to this node

S. Sabetghadam, M. Lupu, and A. Rauber, "A Combined Approach of Structured and Non-structured IR in multimodal domain," in Proceeedings of International Conference on Multimodal Retrieval, ICMR 2014, pp. 491-494.

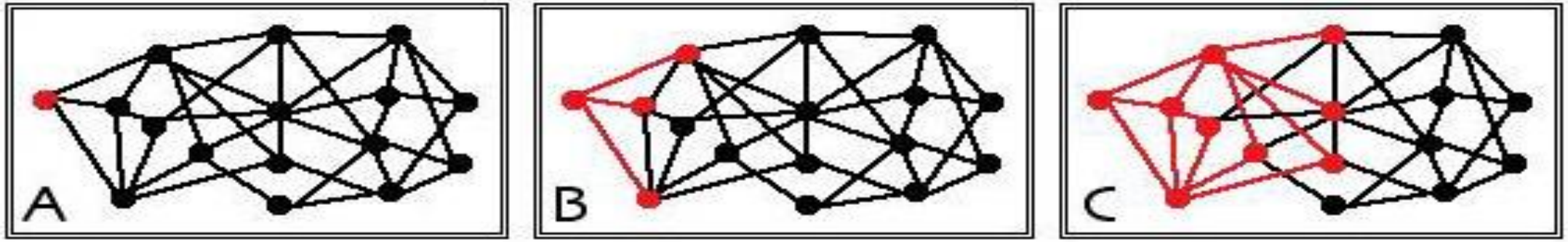
Graph Traversal Methods (1)

Model

- Two of the well-known methods

  Spreading Activation , Random Walks

- Spreading activation
  - Based on associative retrieval idea [Crestani97]
    - Nodes and associations
  - Some nodes get activated
  - Energy propagates to the neighbours



Graph Traversal Methods (2)

Model

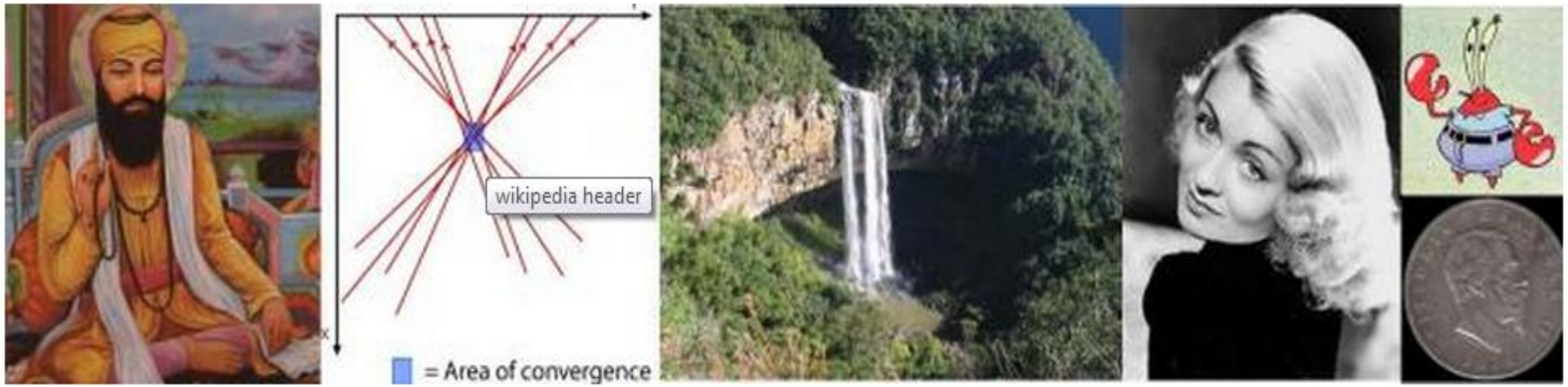
- Random Walks
  - a chain of states created by some stochastic process
  - Stationary distribution of the graph

- Two methods are Principally the same
  - Under certain conditions
  - Used both in different experiments

S. Sabetghadam, M. Lupu, A. Rauber, "Which one to choose? Spreading Activation or Random Walks?", Information Retrieval Facility Conference, IRFC 2014, pp. 112-119.

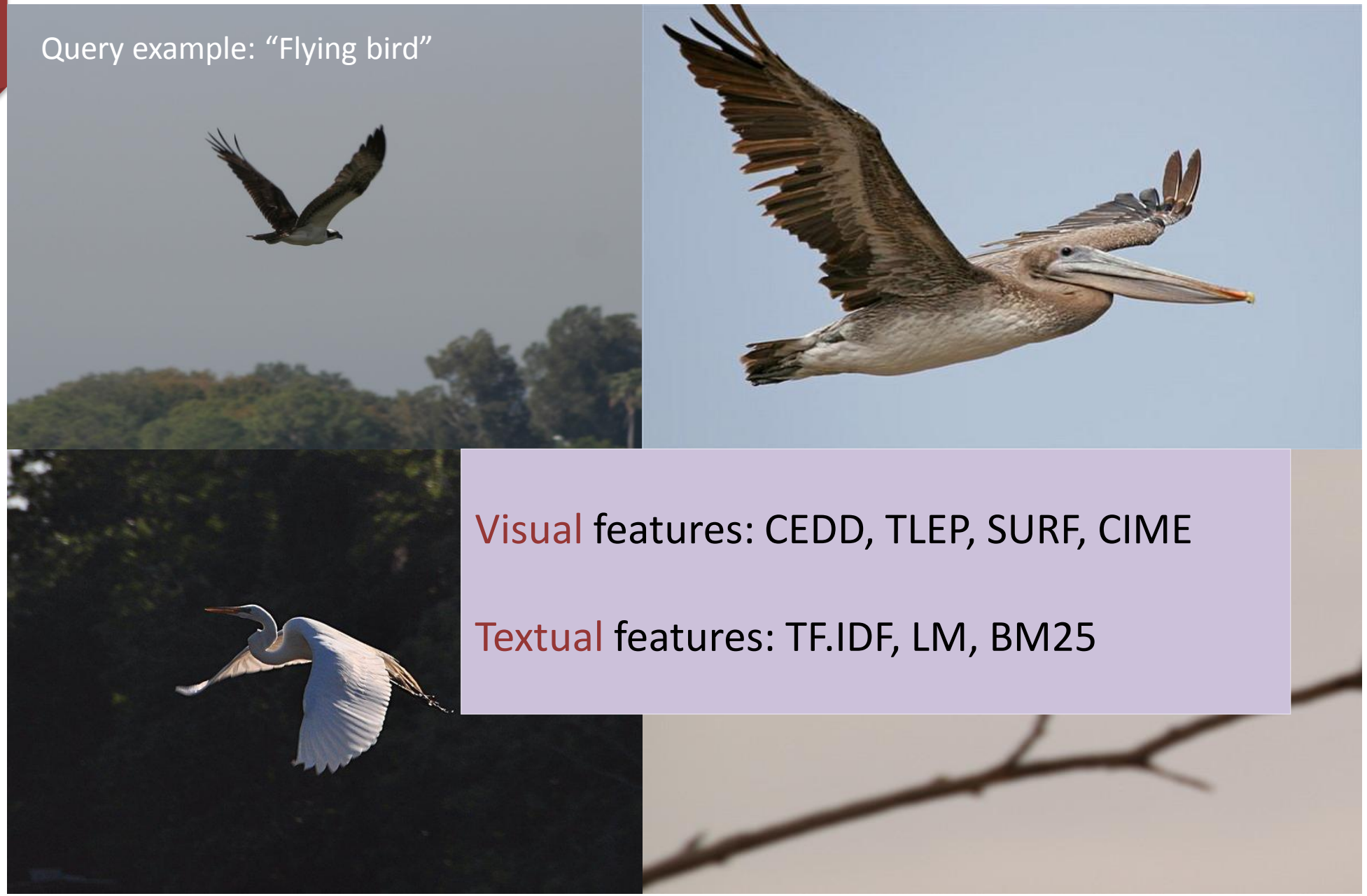
Data collection

- ImageCLEF 2011 Wikipedia collection
- About 400,000 Documents and Images
  - 125,828 documents, 273,434 images



Query data

- 50 topics
  - Easy, medium, hard, very hard [Tsikrika 2011]
- Query
  - keywords
  - four/five image examples



Visual features: CEDD, TLEP, SURF, CIME

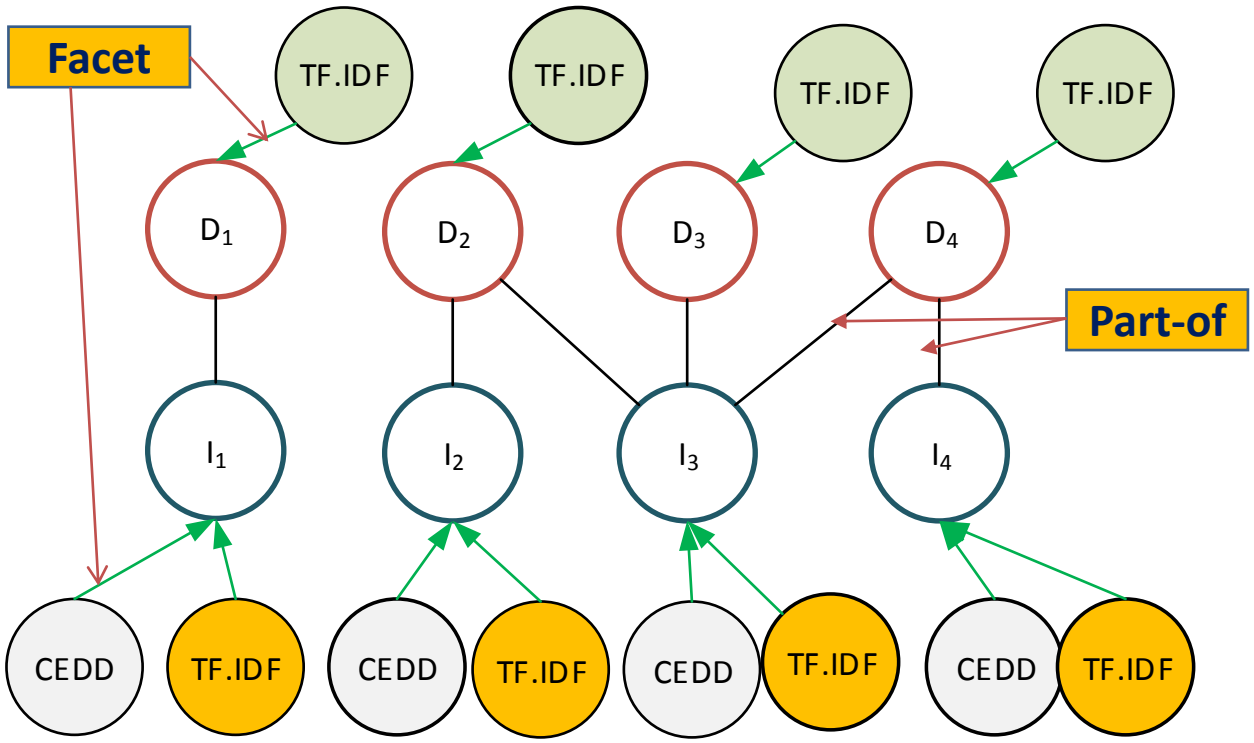
Textual features: TF.IDF, LM, BM25

Image Metadata Provided by the Collection

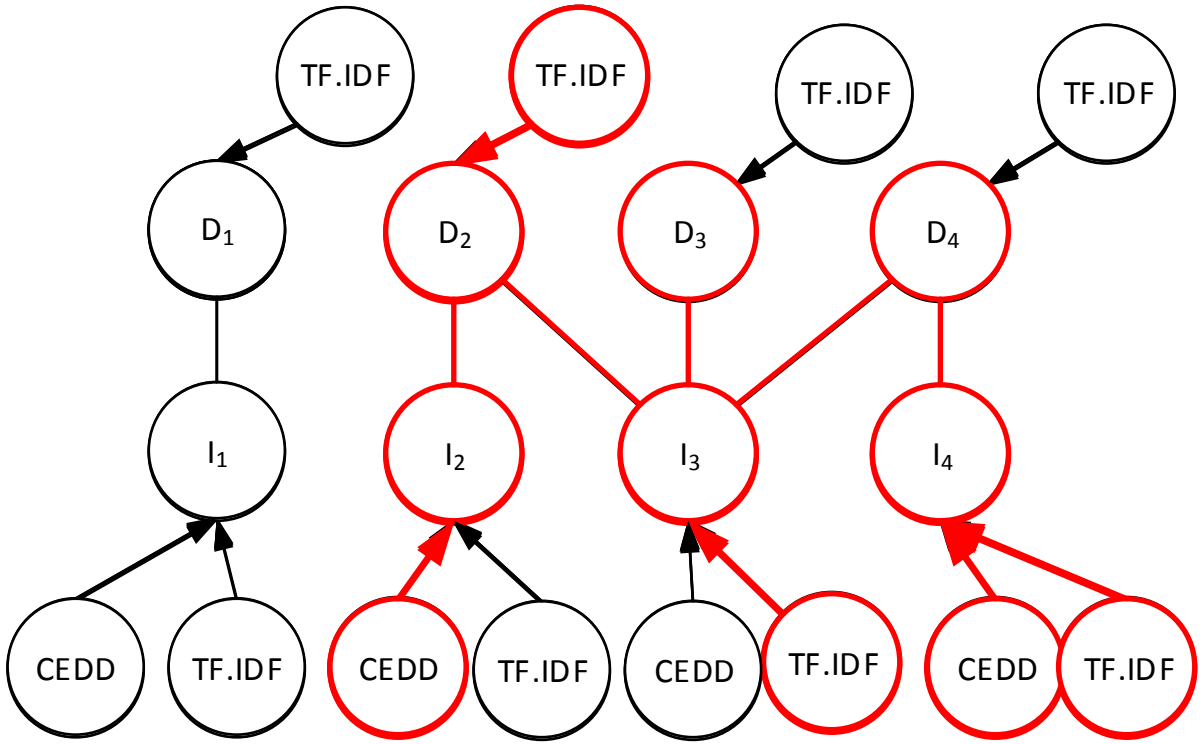
Query: Flying Bird

```
<?xml version="1.0" encoding="UTF-8" ?>
<image id="12031" file="images/2/12031.jpg">
 <name>Colibri-thalassinus-001-edit.jpg</name>
 <text xml:lang="en">
  <description>Green Violet-ear -- Finca Lerida, Boquete, Panama.</description>
  <comment />
  <caption article="text/en/1/307941">Green Violetear at a flower.</caption>
 </text>
 <text xml:lang="de">
  <description>Ein Grün-Veilchenohr kolibri ( Colibri thalassinus ), Finca Lerida, Boquete , Panama .</description>
  <comment />
  <caption />
 </text>
 <text xml:lang="fr">
  <description>Un Colibri thalassinus , Finca Lerida, District de Boquete, Panama .</description>
  <comment />
  <caption />
 </text>
 <comment>({{Information |Description= |Source= |Date= |Author= |Permission= |other_versions= }}
 <license>GFDL</license>
</image>
```

The Collection Mapped to our Model



Hybrid Search Example



Summary

- We proposed a model which supports
  - Different modalities
  - Different relation types
  - Decomposed an information object into facets
- Hybrid Search
  - Standard and Graph search
- Mapped the collection to our model

Contributions

Model

RQ1:Can we define a graph-based model for multi-modal multi-faceted information retrieval (MMIR)?

Contributions

Astera model for MMIR

Modelled faceted search and relevancy computation function

[ECIR Workshop 2013]
[IRFC 2014]
[ICMR 2014]

Reachability

RQ2: In such a graph model, can the relevant nodes be reached?

Contributions

Analysis of the effect of different facets and **links** on reachability of relevant nodes

[ECIR 2015]

Score Analysis

RQ3: In such a model can scores identify the relevant nodes?

Contributions

Analysis of the effect of different facets on precision.

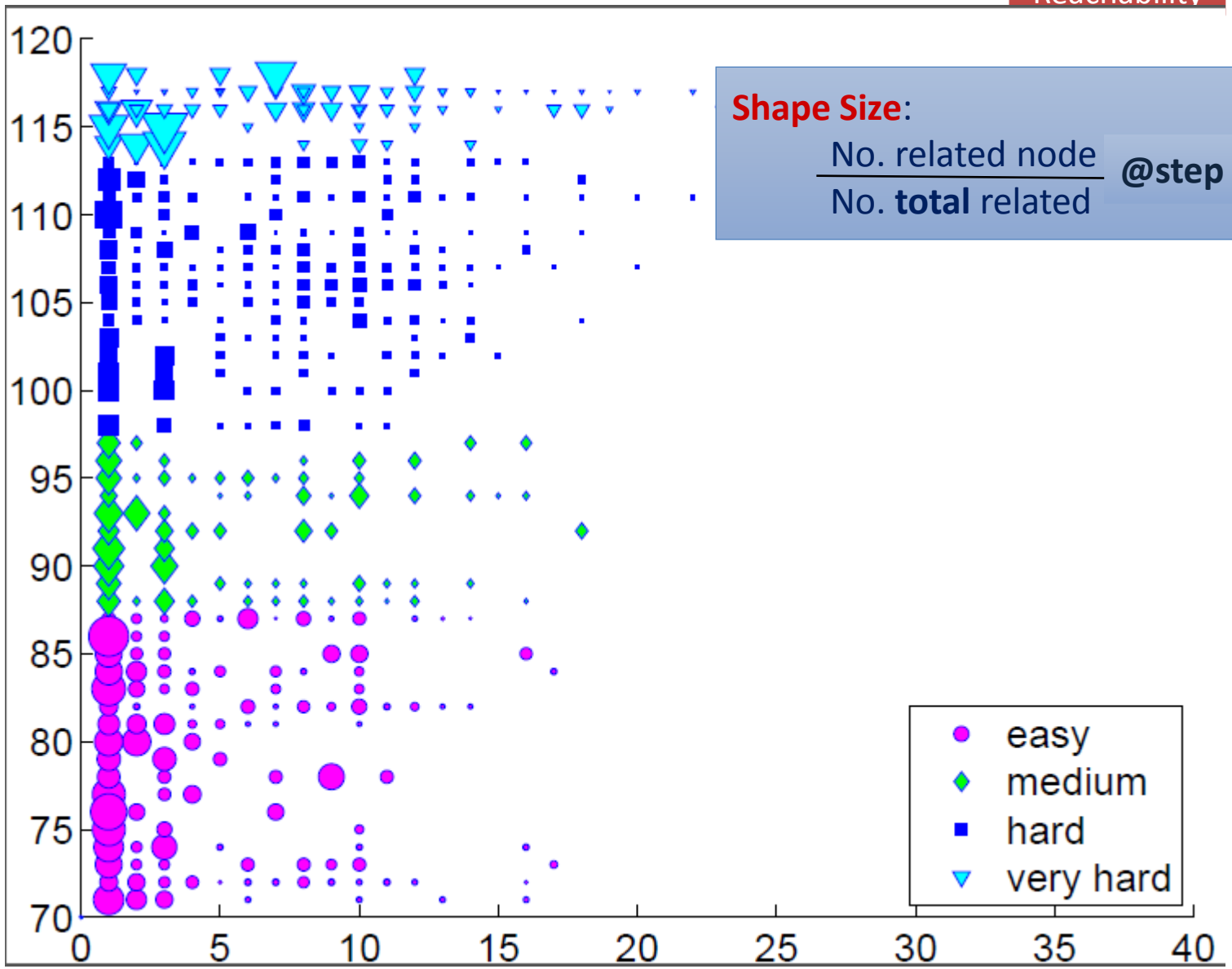
The effect of Query-dependent and –independent routing on precision

[CLEF 2014]
[GBS 2016]
[Keystone 2016]

Relevant Data Distribution

- 40 steps
- In each step
  - Check if we visit

    new relevant image



RQ2: In such a graph model, can the **relevant** nodes be **reached**?

C1: Reachability Analysis from Different Facets

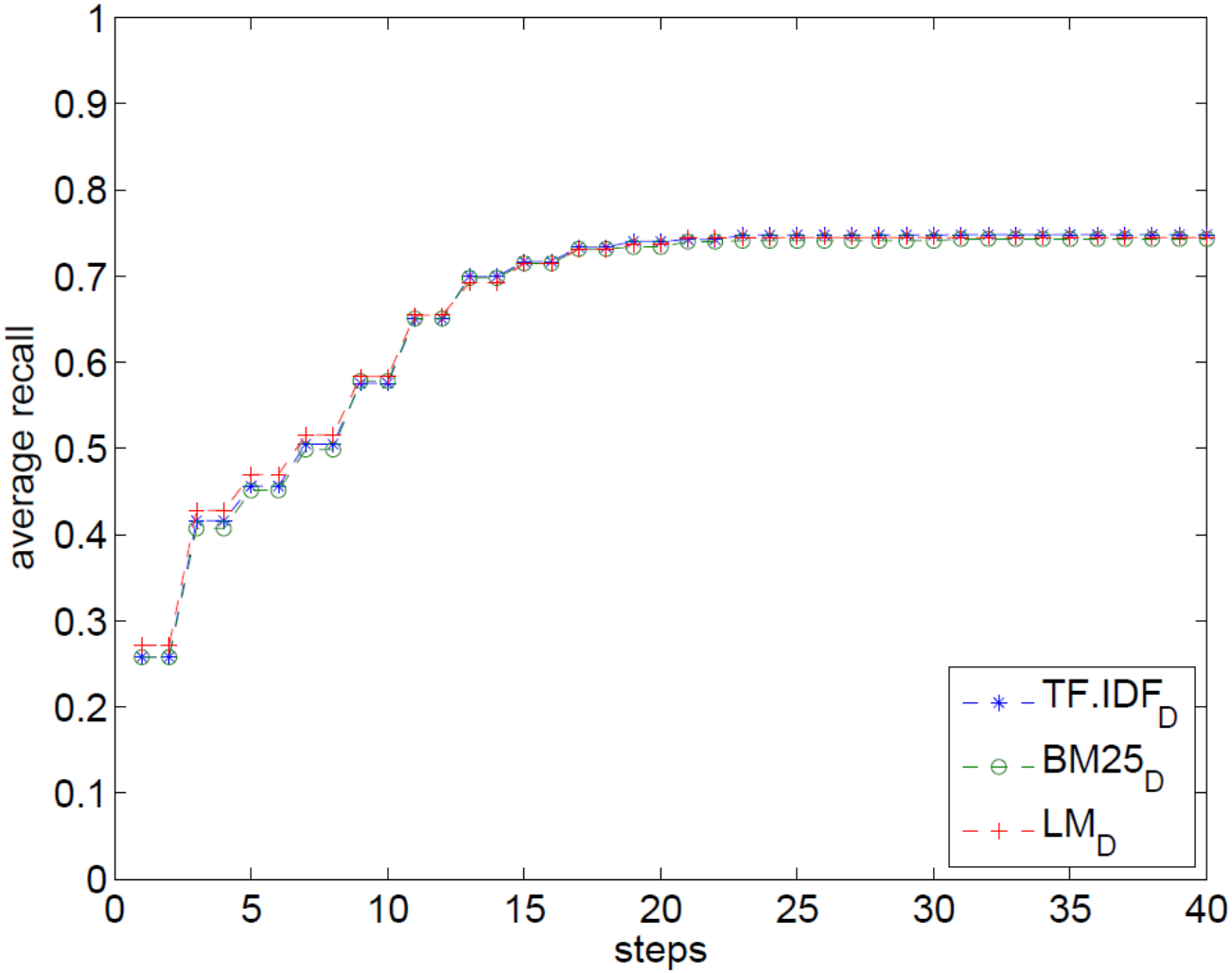
C2: Reachability Analysis from Different Links

C3: Reachability Analysis of Different Topic Categories

C4: Graph Visit from Different Facets

Recall from Document Textual Facets

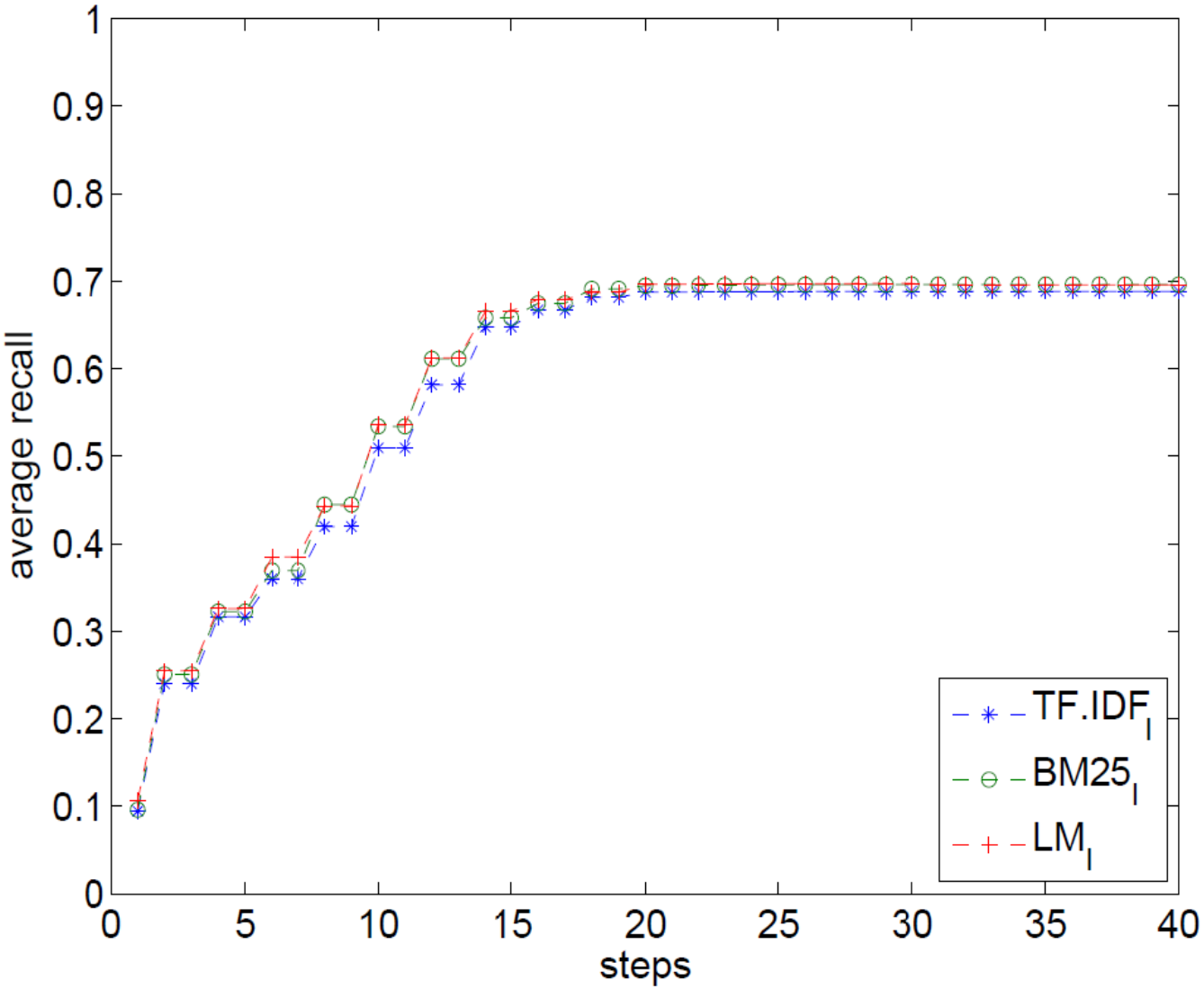
- Links used
  - Facet and Part-of
- Top 20 results of each facet
- Traverse the graph from these results
- Calculate recall in each step



Similar recall behaviour

Recall from Image Textual Facets

- Same experiment starting from results of Image metadata textual facet

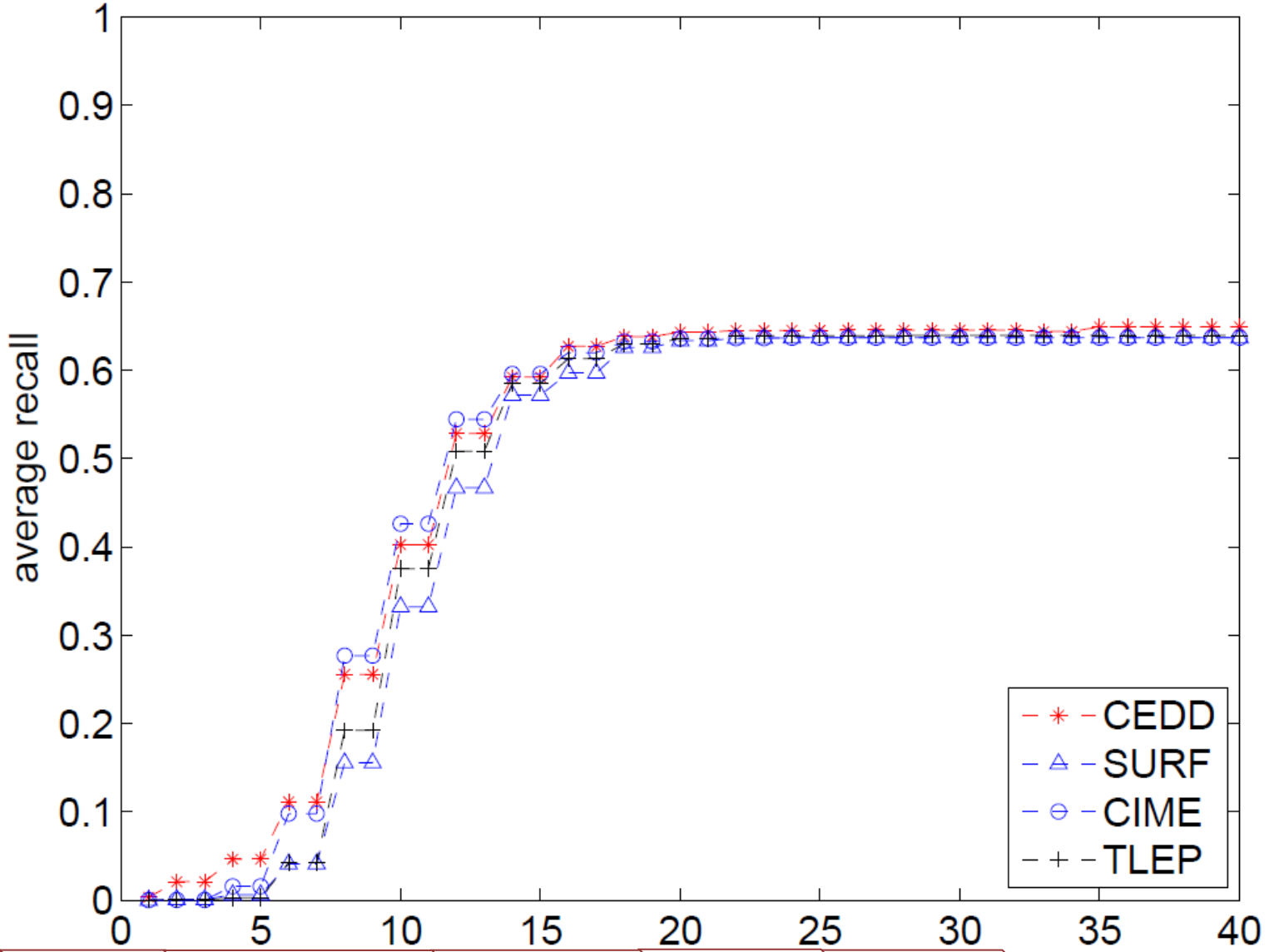


Similar recall behaviour

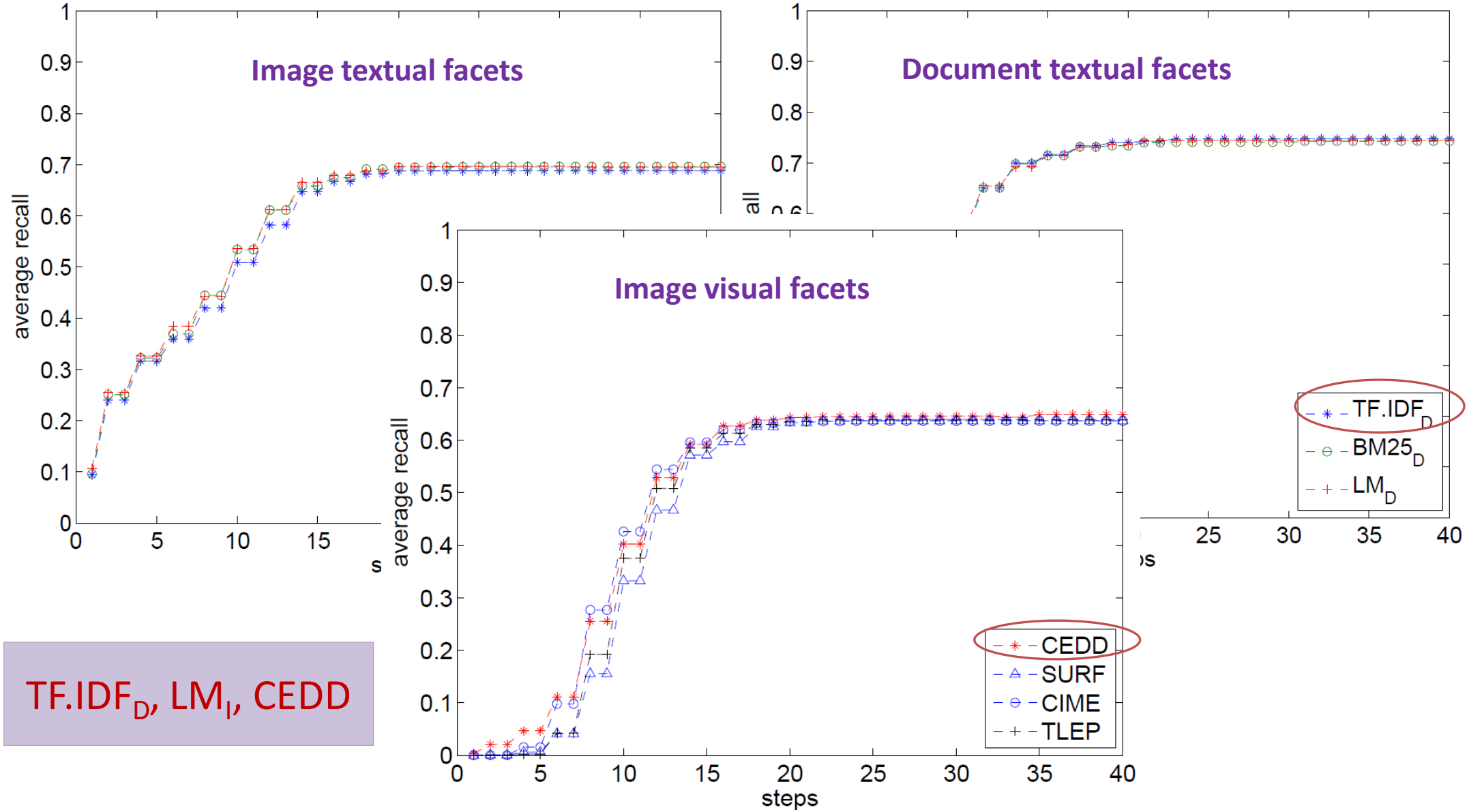
Recall from Image Visual Facets

- Same experiment starting from results of Image visual facets

Similar recall behaviour



C1 Representative Facet from each Category of



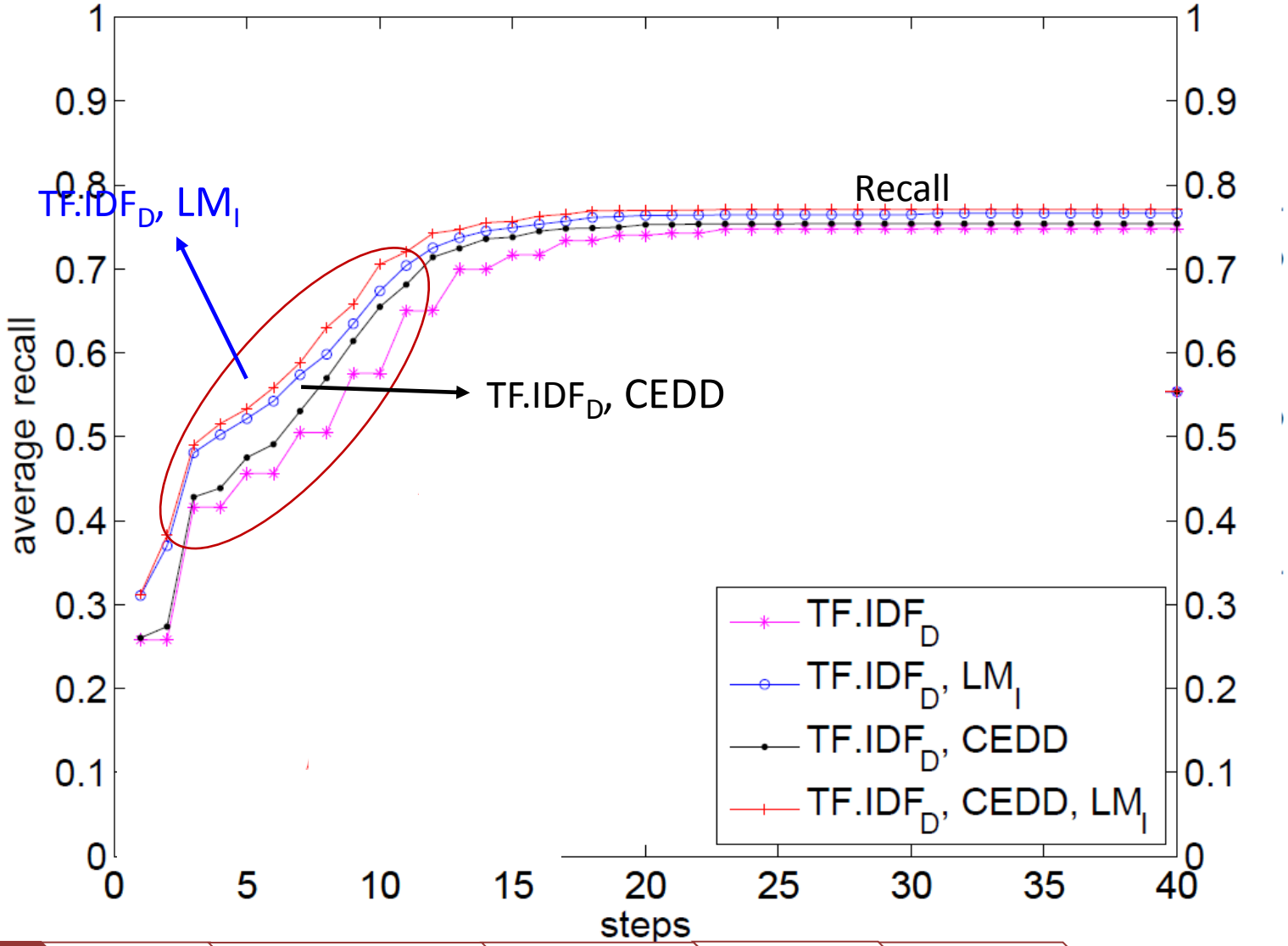
$TF.IDF_D$, LM_I, CEDD

Facet Combinations (1)

- Links: part-of, facet
  - No semantic/ similarity links
- Fewer visited nodes, higher recall

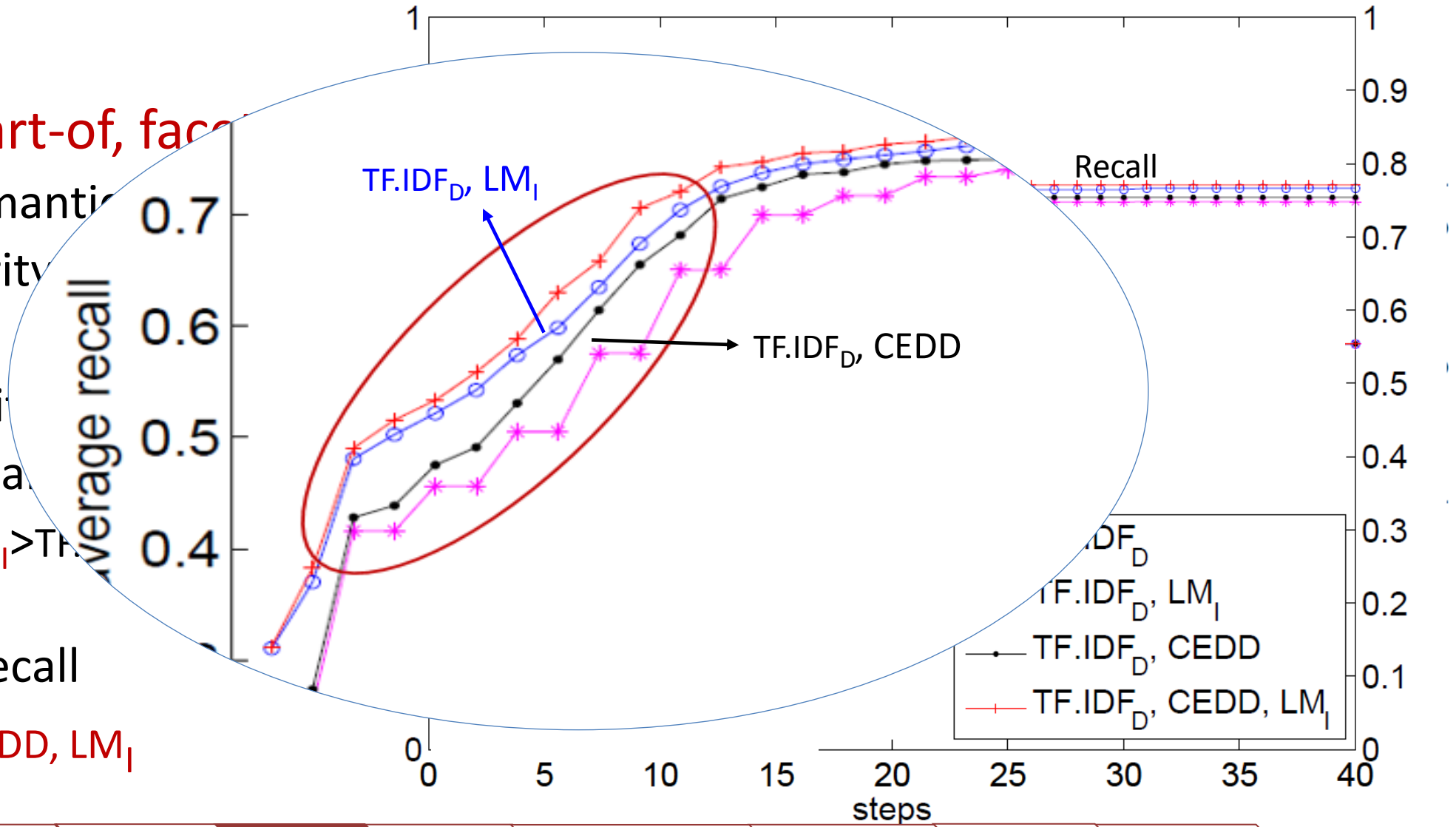
  $TF.IDF_D$, LM_l > $TF.IDF_D$, CEDD
- Highest recall

  $TF.IDF_D$, CEDD, LM_l



Facet Combinations (1)

- Links: part-of, face
  - No semanti
    similarity
- Fewer visi
  higher reca
  TF.IDF$_D$, LM$_I$>TF
- Highest recall
  TF.IDF$_D$, CEDD, LM$_I$

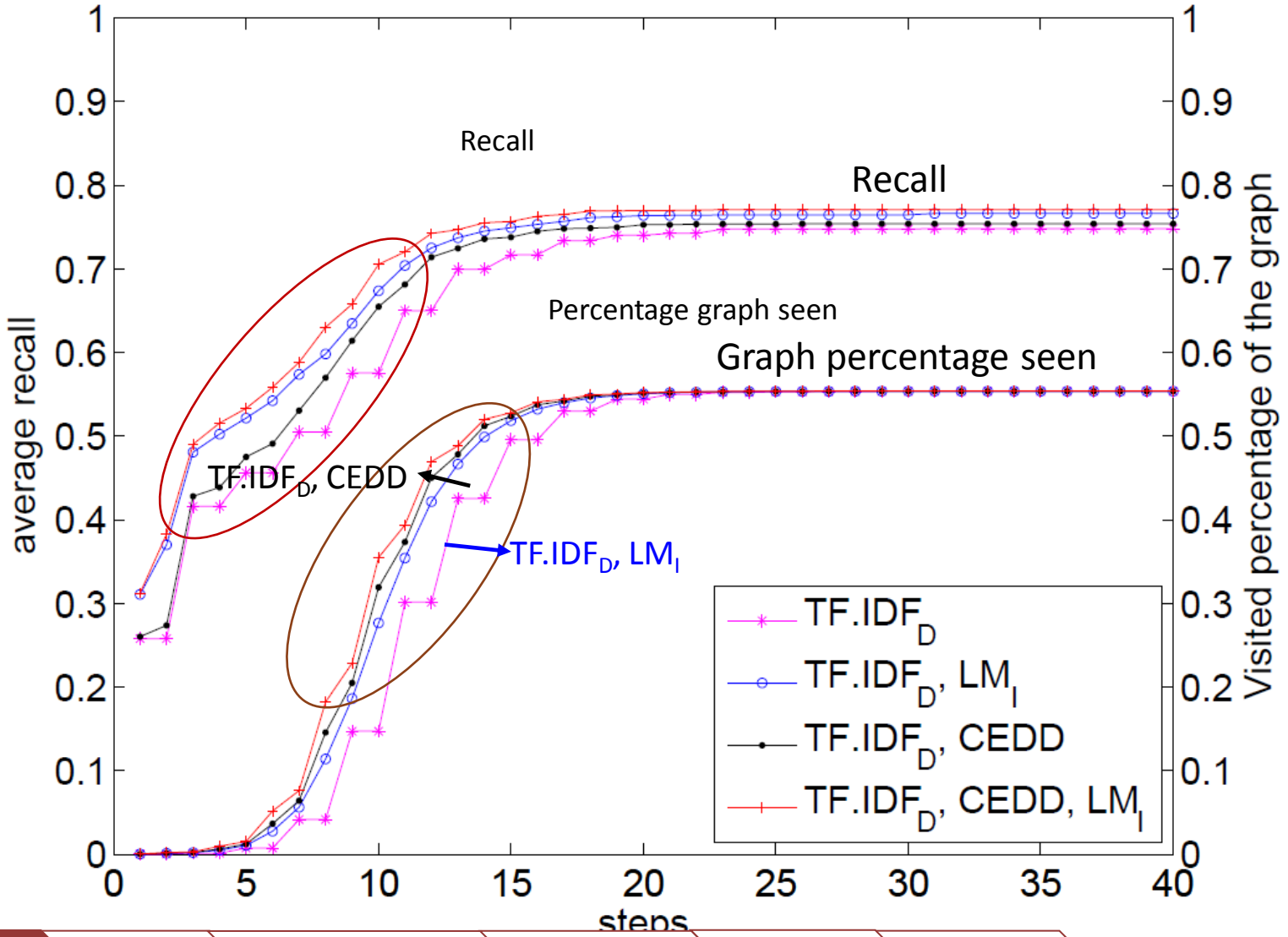


Facet Combinations (2)

- Links: part-of, facet
  - No semantic/ similarity links
- Fewer visited nodes, higher recall

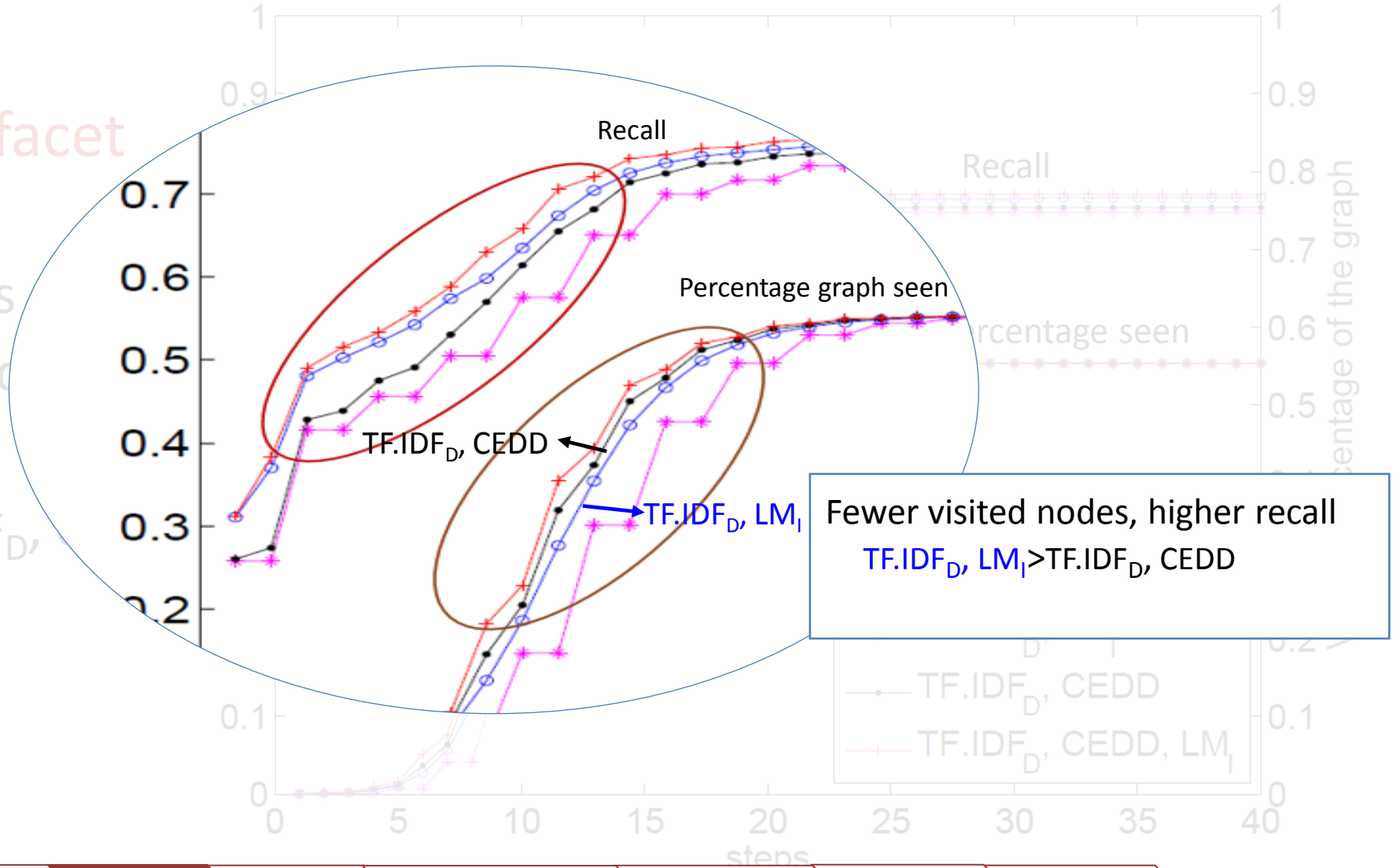
  $TF.IDF_D$, LM_l>$TF.IDF_D$, CEDD
- Highest recall

  $TF.IDF_D$, CEDD, LM_l



Facet Combinations (2)

- Links: part-of, facet
  - No semantic/ similarity links
- Fewer visited nodes, higher recall
  TF.IDF$_D$, LM$_I$>TF.IDF$_D$, CEDD
- Highest recall
  TF.IDF$_D$, CEDD, LM$_I$



C1

Observations

- Different facets leading to visiting different parts of the collection

- This reinforces the importance of the poly-representation idea to identify the relevant objects.

Sabetghadam S., Lupu M., Bierig R., and Rauber A, ."Reachability Analysis of Graph Modelled Collections". 37th European Conference on Information Retrieval, ECIR 2015, pp. 370-381

C1

Observations

- Different facets leading to visiting different parts of the collection

- This reinforces the importance of the poly-representation idea to identify the relevant objects.

Still limited access to the graph,
only half of the graph is reachable

Sabetghadam S., Lupu M., Bierig R., and Rauber A, ."Reachability Analysis of Graph Modelled Collections". 37th European Conference on Information Retrieval, ECIR 2015, pp. 370-381

RQ2: In such a graph model, can the **relevant** nodes be **reached**?

C1: Reachability Analysis from Different Facets

C2: Reachability Analysis from Different Links (α,β,δ,Ɣ)

C3: Reachability Analysis of Different Topic Categories

C4: Graph Visit from Different Facets

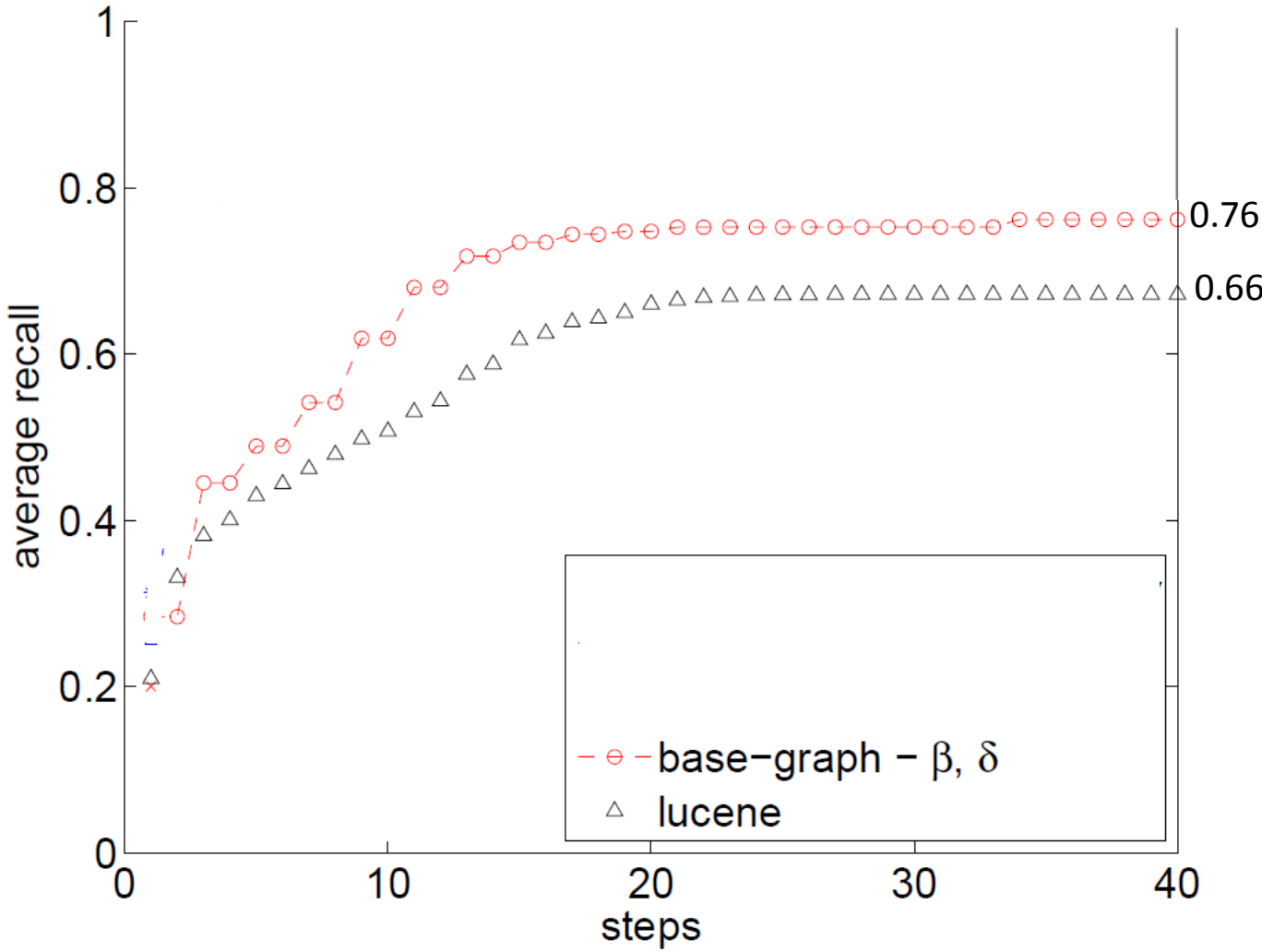
C2

Recall Baseline Graph vs Lucene Results

β: Part-of

δ: Facet

- Base graph recall: 0.76
- Better than Lucene: 0.66



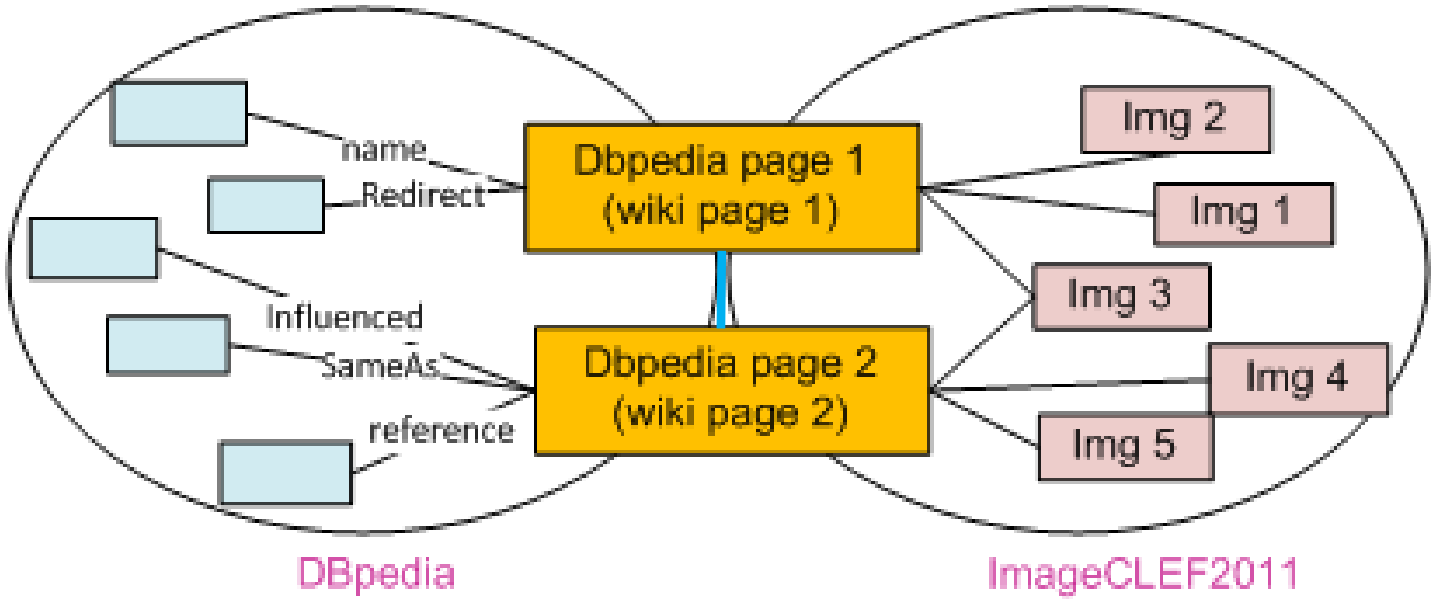
C2

Adding Semantic Links

Reachability

- Using DBpedia dump
- Adding semantic links between equivalent pages
- 55,544 Intra-lingual links added
- 100,653 Inter-lingual links added

link type	*frequency*
FR-DE sameAs	36,295
EN-DE sameAs	34,530
EN-FR sameAs	29,828

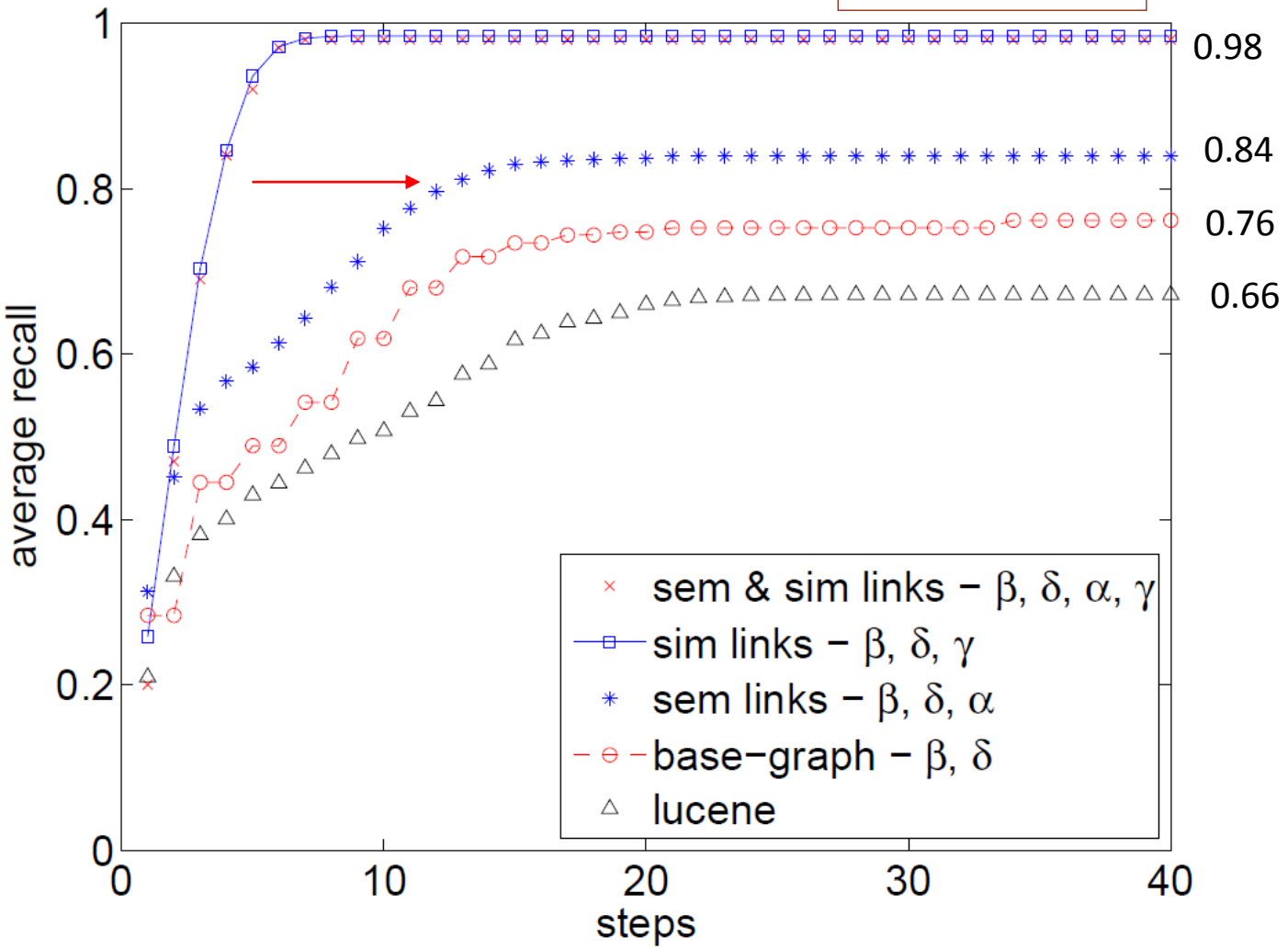


C2

Recall after Adding Semantic Links

α: Semantic
β: Part-of
δ: Facet
χ: Similarity

- Semantic links added from DBpedia dump
- Facets to start from
  - $TF.IDF_D$, CEDD, LM_I
- Recall increase of %10 after adding semantic links

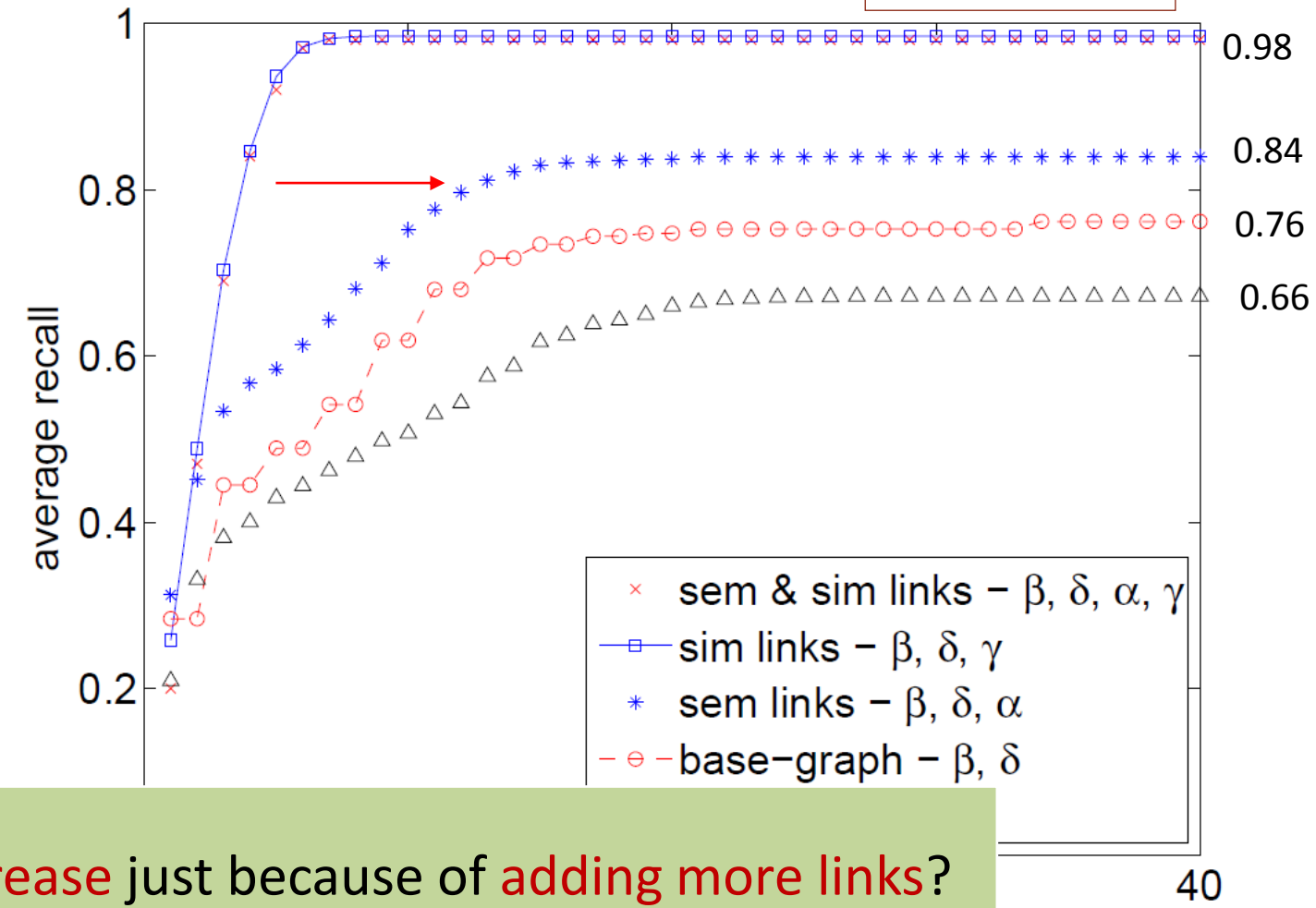


C2

Recall after Adding Semantic Links

α: Semantic
β: Part-of
δ: Facet
χ: Similarity

- Semantic links added from DBpedia dump
- Facets to start from
  - $TF.IDF_D$, CEDD, LM_l
- Recall increase of %10 after adding semantic lin



Question: Is this recall increase just because of adding more links?

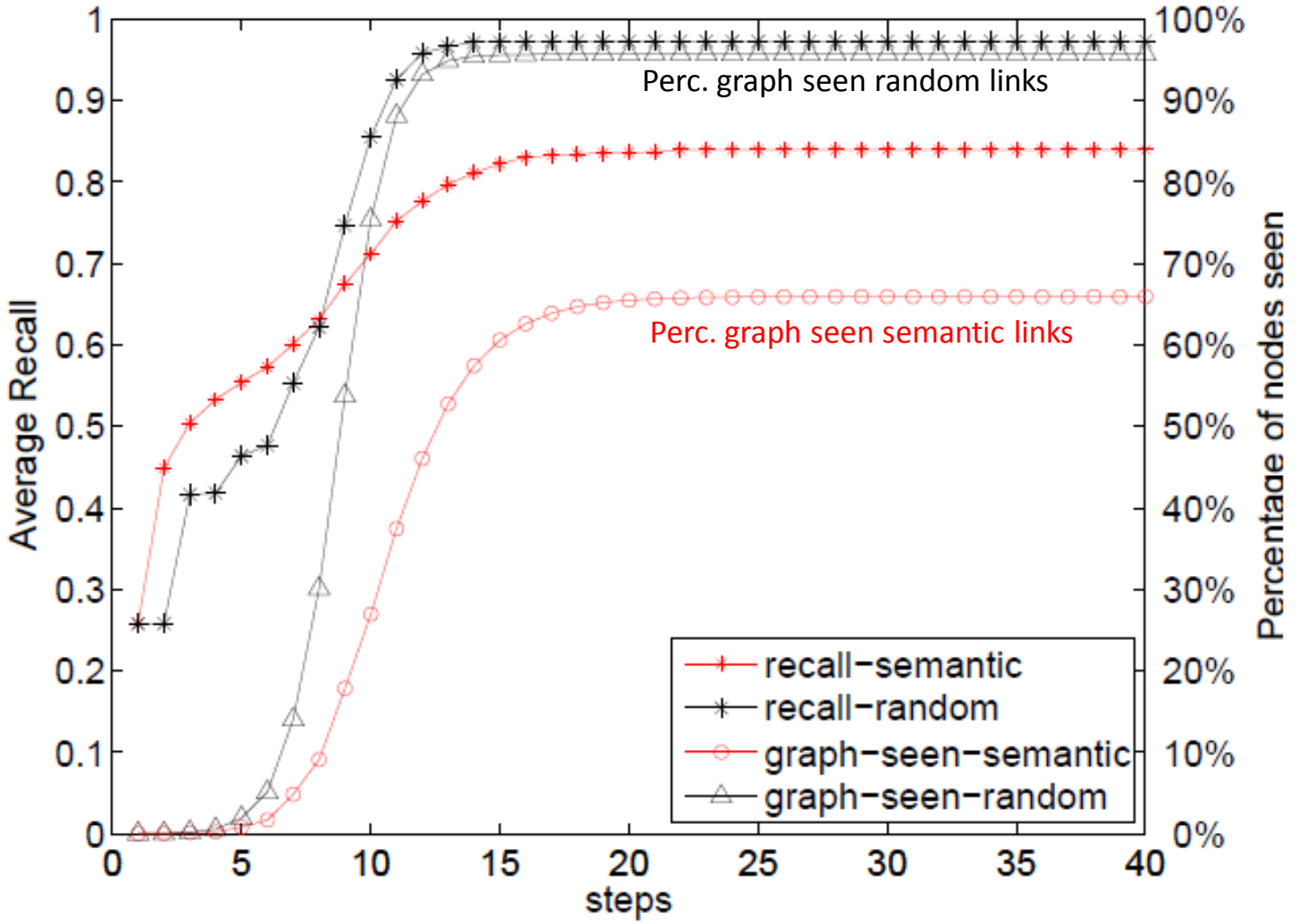
C2

Recall comparison: Semantic vs. Random

Reachability

- We added the same number of random links between docs

Higher recall with random links but with the expense of visiting almost all the graph!

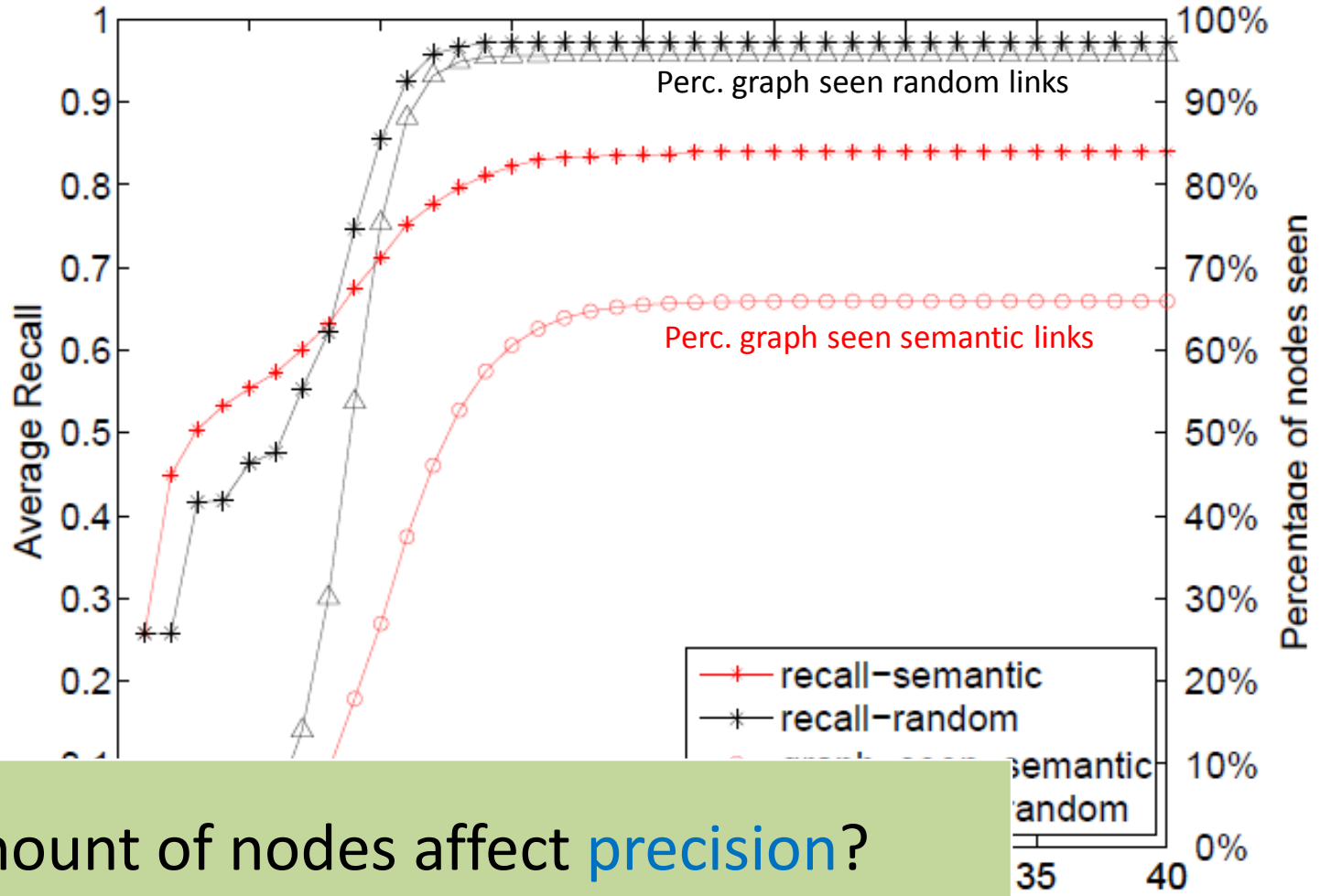


C2

Recall comparison: Semantic vs. Random

Reachability

- We added the same number of random links between docs



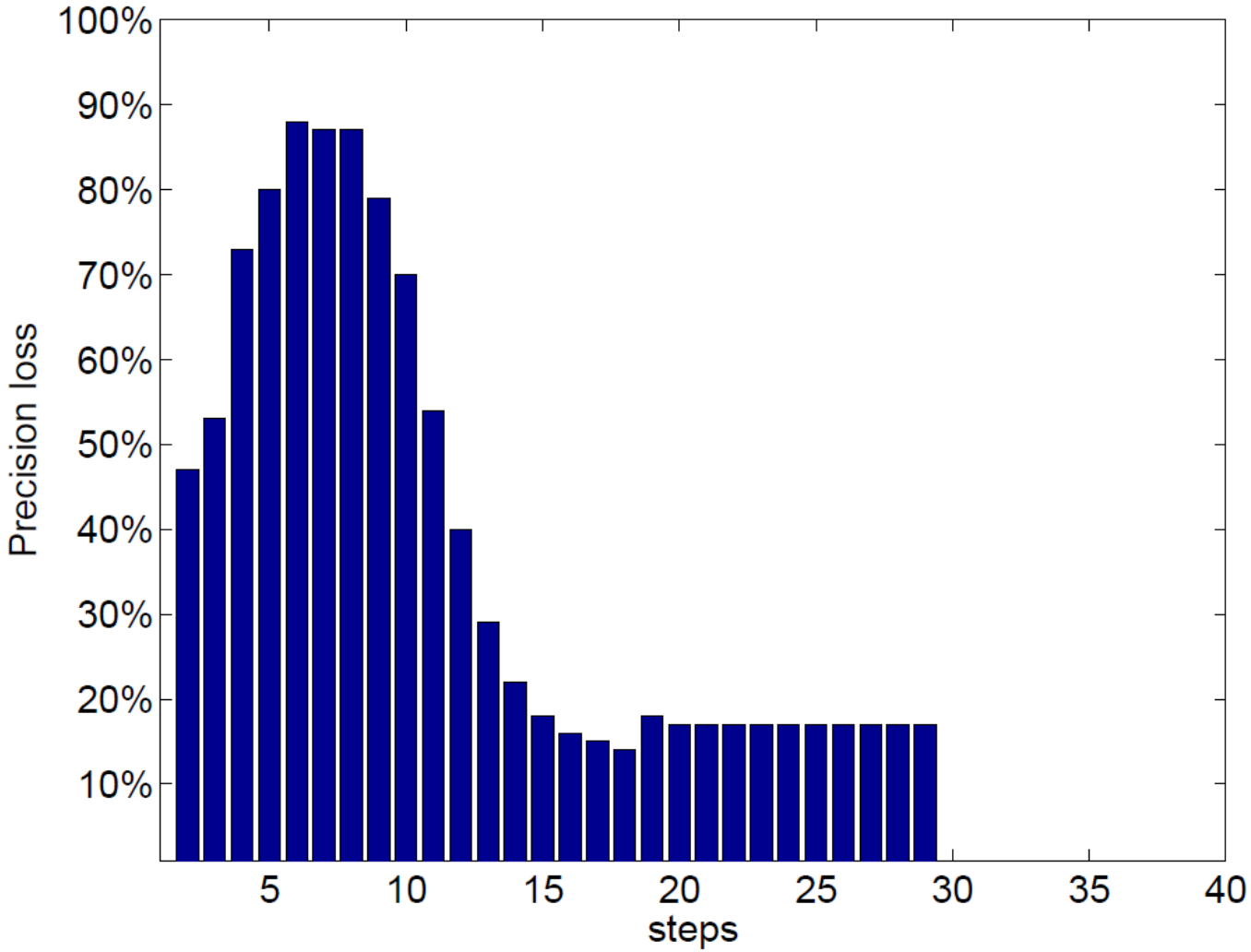
Higher recall with random links but with the expense of visiting almost all the graph!

How visiting this large amount of nodes affect precision?

Adding random links - Precision loss

$$\text{precLoss} = 1 - \left(\frac{\text{precision with random semantic links}}{\text{precision with real semantic links}}\right)$$

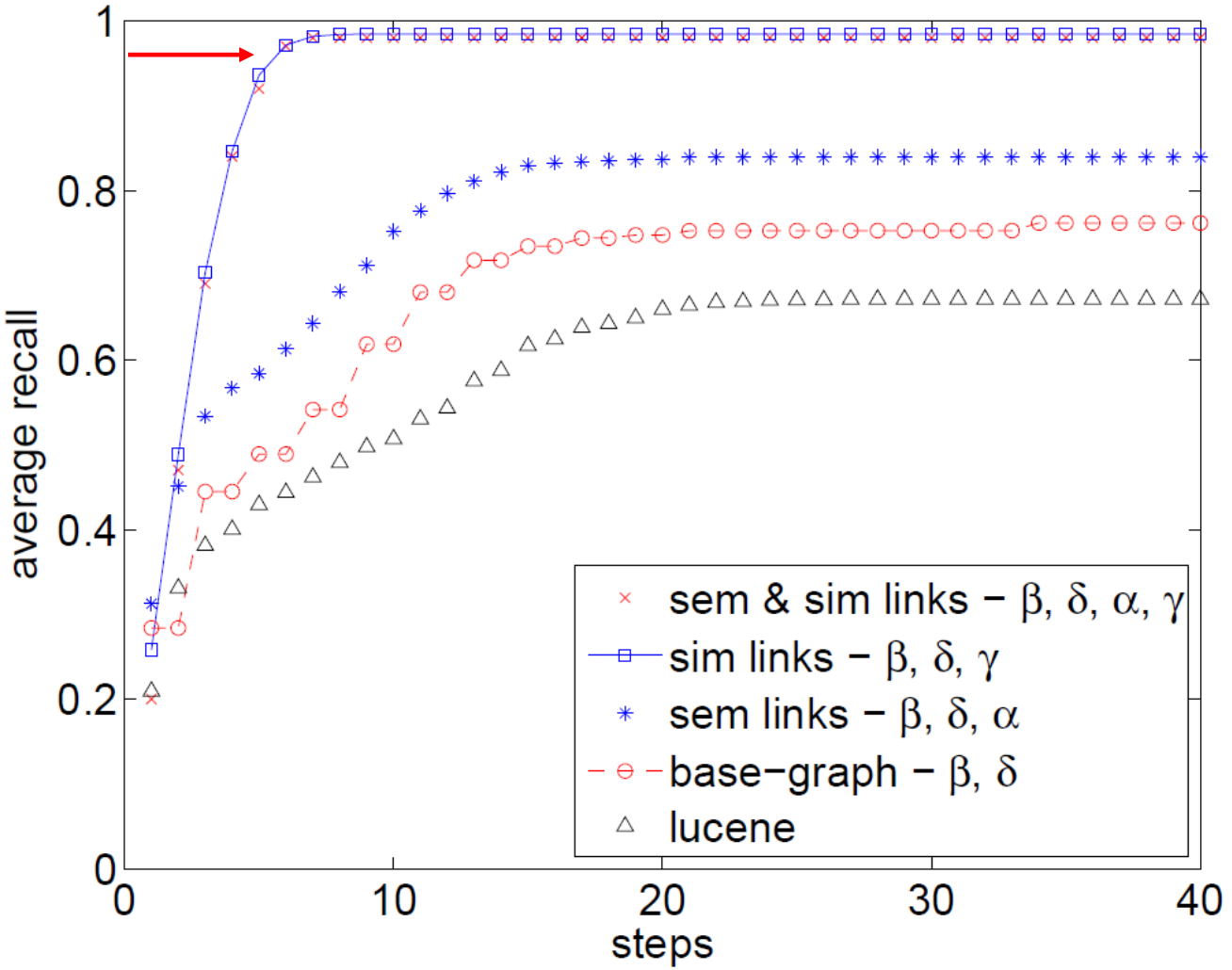
Adding semantic links helped effectively to the recall increase.



Recall after Adding Similarity Links

We did the same analysis with adding similarity links

They are effectively helpful in reaching more relevant nodes



Conclusion (1)

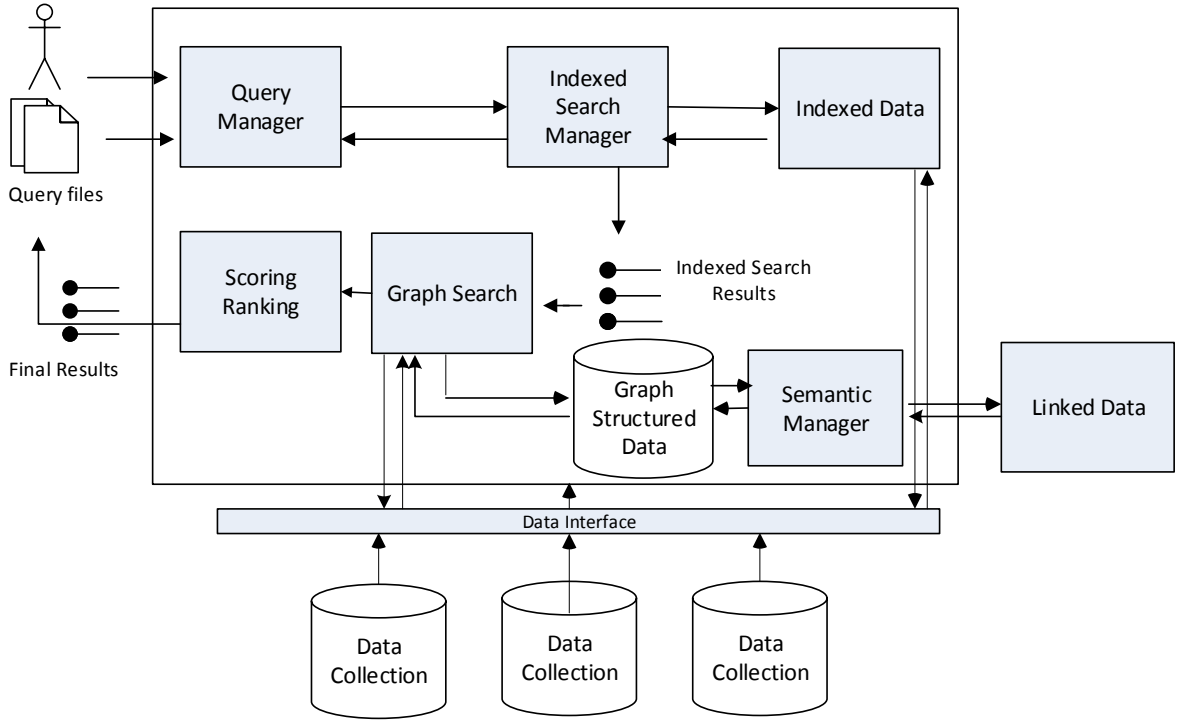
- RQ1: How to define a graph-based model for multi-modal multi-faceted information retrieval?
- We defined a graph-based model which supports
  - Different modalities
  - Different types of relations
  - Our defined search based on facets
- Decomposes an information object to its facets
- Defined relevancy function Based on Poly-representation principle
- Calculate relevancy of an information object to the query as starting points in the graph

Conclusion (2)

- RQ2: In such a graph model, can the relevant nodes be reached?
- We showed the effect of poly-representation on recall
  - Combination of facets increases recall
  - Not every combination is effective
- The results from adding random links showed:
  - The effect of adding meaningful in higher reachability

Astera Status

- Astera: 2012 - now
- open source and available online
  - http://ifs.tuwien.ac.at/~sabetghadam/Astera.html
- Highly configurable to work with other collections



Scientific Publications

- **Sabetghadam S.**, Lupu M., Rauber A., "Astera - A generic model for multi-modal Information Retrieval", Workshop on Integrating IR technologies for Professional Search, held in **ECIR 2013**, pp. 551-554.
- **Sabetghadam S.**, Lupu M., and Rauber A., "Which one do you choose? Spreading Activation or Random Walks?". Information Retrieval Facility Conference, **IRFC 2014**, pp. 112-119.
- **Sabetghadam S.**, Lupu M., Bierig R., and Rauber A., "A combined approach of structured and non-structured IR in multi-modal domain", In Proceedings of International Conference on Multimedia Retrieval, **ICMR 2014**, pp. 491-494.
- **Sabetghadam S.**, Astera - A model for Multimodal IR with a Combined Approach of Structured and Non-structured Retrieval, Doctoral Symposium, **ICMR 2014**, pp. 551.
- **Sabetghadam S.**, Lupu M., Bierig R., and Rauber A., "A Hybrid Approach for Multi-Faceted IR in multi-modal Domain", 5th Conference of Labs and Evaluation Forums, **CLEF 2014**, pp. 86-97.
- **Sabetghadam S.**, Lupu M., Bierig R., and Rauber A,."Reachability Analysis of Graph Modelled Collections". 37th European Conference on Information Retrieval, **ECIR 2015**, pp. 370-381.
- **Sabetghadam S.**, Palotti J., Rekabsaz N., Lupu M., Hanbury A.."TUW at MediaEval 2015". MediaEval, 2015. Obtained first place in the task of „Diverise Social Image Retrieval“, **Media Eval 2015**.
- **Sabetghadam S.**, Lupu M., and Rauber A,."Leveraging Metropolis-Hastings Algorithm on Graph-based Model for multimodal IR". GSB'15: First International Workshop on Graph Search and Beyond, held at **SIGIR 2015**, pp. 14-18.
- Navid Rekabsaz, **Serwah Sabetghadam**, Mihai Lupu, Linda Andersson, Allan Hanbury, Standard Test Collection for English-Persian Cross-Lingual Word Sense Disambiguation, Proceeding on Language Resources and Evaluation Conference, **LREC 2016**.
- **Sabetghadam S.**, Lupu M., Rauber A., "Random Walks Analysis on Graph Modelled Multimodal Collections", Second International KEYSTONE Conference **Keystone 2016**.
- Eskevich M., Larson M., Aly R., **Sabetghadam S.**, Jones G., Ordelman R., Huet B. “Multimodal Video-to-Video Linking: Turning to the Crowd for Insight and Evaluation”, Proceeding on Multimedia Modeling, **MMM 2017**, pp. 280-292.
- S. **Sabetghadam**, M. Lupu, R. Bierig, A. Rauber, "*A faceted approach to reachability analysis of graph modelled collections*"; International Journal of Multimedia Information Retrieval, **IJMIR (2017),** 10.1007/s13735-017-0145-8; S. 1 - 15.