

Align, Disambiguate, and Walk

A Unified Approach for Measuring Semantic Similarity



SAPIENZA
UNIVERSITÀ DI ROMA

Mohammad Taher Pilehvar
David Jurgens
Roberto Navigli



Semantic Similarity;
how similar are a
pair of lexical items?

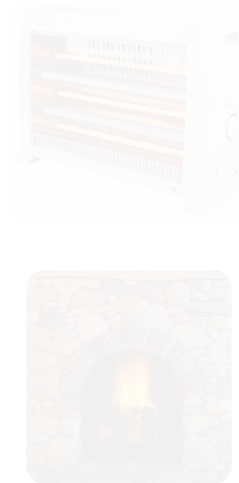


Semantic Similarity

Sentence Level



Word Level



Sense Level

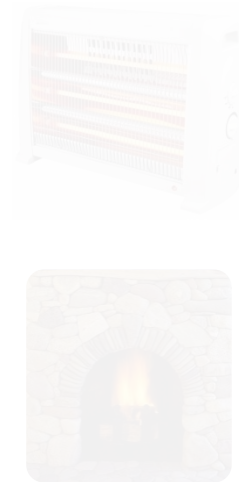


Semantic Similarity

Sentence Level



Word Level



Sense Level



Semantic Similarity

Sentence level

The worker was terminated

The boss fired him



➤ Applications

- **Paraphrase recognition**
(Tsatsaronis et al., 2010)
- **MT evaluation**
(Kauchak and Barzilay, 2006)
- **Question Answering**
(Surdeanu et al., 2011)
- **Textual Entailment**
(Dagan et al., 2006)

Semantic Similarity

Sentence Level



Word Level



Sense Level



Semantic Similarity

Word level

heater



fireplace



➤ Applications

- Lexical simplification
(Biran et al., 2011)

Locuacious → *Talkative*

- Lexical substitution
(McCarthy and Navigli, 2009)

Semantic Similarity

Sentence Level



Word Level



Sense Level



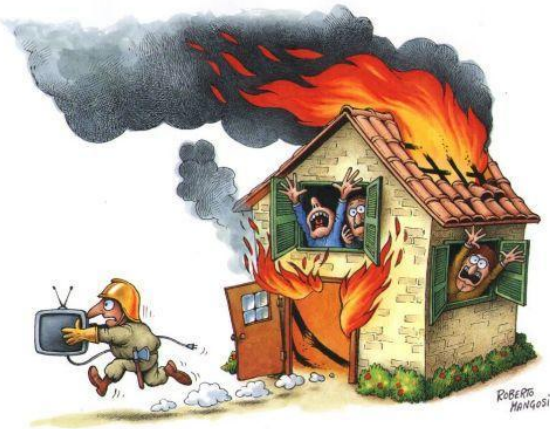
Semantic Similarity

Sense level

fire sense #1



fire sense #8



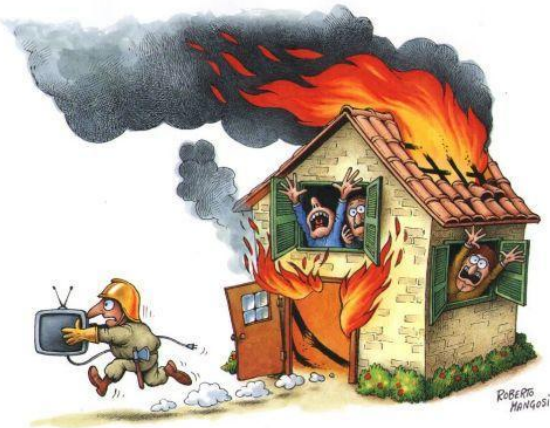
Semantic Similarity

Sense level

fire sense #1



fire sense #8



➤ Applications

- Coarsening sense inventories
(Snow et al., 2007)
- Semantic priming
(Neely et al., 1989)

Existing Similarity Measures

Sentence

Word

Sense

Existing Similarity Measures

Allison and Dix (1986)

Gusfield (1997)

Wise (1996)

Keselj et al. (2003)

Sentence

Word

Sense

Existing Similarity Measures

Allison and Dix (1986)

Gusfield (1997)

Wise (1996)

Keselj et al. (2003)

Salton and McGill (1983)

Gabrilovich and Markovitch (2007)

Radinsky et al. (2011)

Ramage et al. (2009)

Yeh et al., (2009)

Turney (2007)

Landauer et al. (1998)

Sentence

Word

Sense

Existing Similarity Measures

Allison and Dix (1986)

Gusfield (1997)

Wise (1996)

Keselj et al. (2003)

Salton and McGill (1983)

Gabrilovich and Markovitch (2007)

Radinsky et al. (2011)

Ramage et al. (2009)

Yeh et al., (2009)

Turney (2007)

Landauer et al. (1998)

Patwardan (2003)

Banerjee and Pederson (2003)

Hirst and St-Onge (1998)

Lin (1998)

Jiang and Conrath (1997)

Resnik (1995)

Sussna (1993, 1997)

Wu and Palmer (1994)

Leacock and Chodorow (1998)

Sentence

Word

Sense

Existing Similarity Measures

But

Allison and Dix (1986)

Gusfield (1997)

Wise (1996)

Keselj et al. (2003)

Salton and McGill (1983)

Gabrilovich and Markovitch (2007)

Radinsky et al. (2011)

Ramage et al. (2009)

Yeh et al., (2009)

Turney (2007)

Landauer et al. (1998)

Patwardan (2003)

Banerjee and Pederson (2003)

Hirst and St-Onge (1998)

Lin (1998)

Jiang and Conrath (1997)

Resnik (1995)

Sussna (1993, 1997)

Wu and Palmer (1994)

Leacock and Chodorow (1998)

Sense

Word

Sentence



Existing Similarity Measures

But

None directly covers all levels
at the same time

Allison and Dix (1986)

Gusfield (1997)

Wise (1997)

Keselj et al. (2003)

(2003)

First and St-Onge (1998)

Gabrilovich and Markovitch (2007)

Lin (1998)

Radinsky et al. (2011)

Jiang and Conrath (1997)

Ramage et al. (2009)

Resnik (1995)

Yeh et al., (2009)

Sussna (1993, 1997)

Turney (2007)

Wu and Palmer (1994)

Landauer et al. (1998)

Leacock and Chodorow (1998)

Sense

Word

Sentence

Existing Similarity Measures

But

None directly covers all levels
at the same time

Different output scales

Sense

Word

Sentence

Allison and Dix (1986)

Gusfield (1997)

Wise (1997)

Keselj et al. (2003)

Gabrilovich and Markovitch (2007)

Radlinski et al. (1997)

Ramage et al. (2009)

Yeh et al., (2009)

Turney (2007)

Landauer et al. (1998)

Lin (1998)

Lin (1997)

Resnik (1995)

Sussna (1993, 1997)

Wu and Palmer (1994)

Leacock and Chodorow (1998)

Existing Similarity Measures

But

None directly covers all levels
at the same time

Different output scales

Different internal representations
which are not comparable to each other

Sense

Word

Sentence

Contribution

Sense

Word

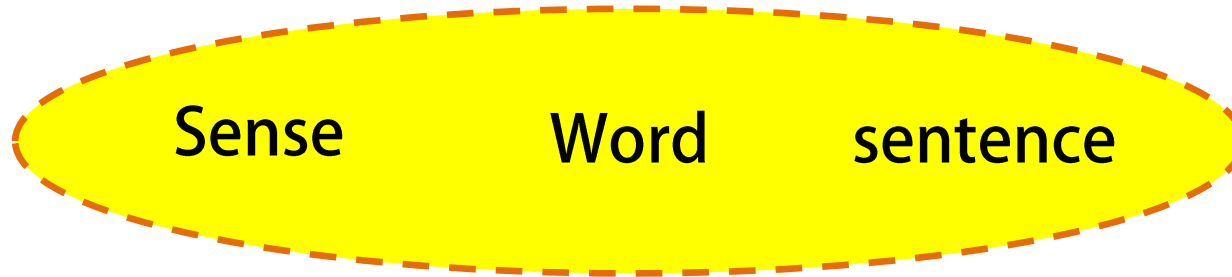
sentence

Contribution



A unified representation
for any lexical item

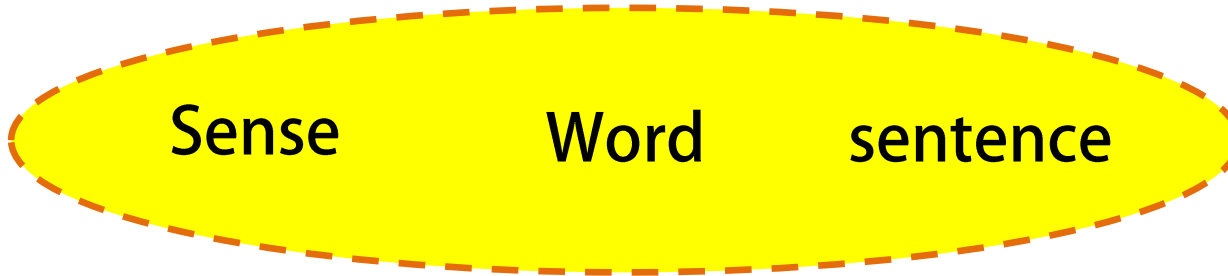
Contribution



A unified representation
for any lexical item

State-of-the-art performance in each level

Contribution



A unified representation
for any lexical item

State-of-the-art performance in each level

Using only WordNet

Advantage I

Unified representation

sense

sentence

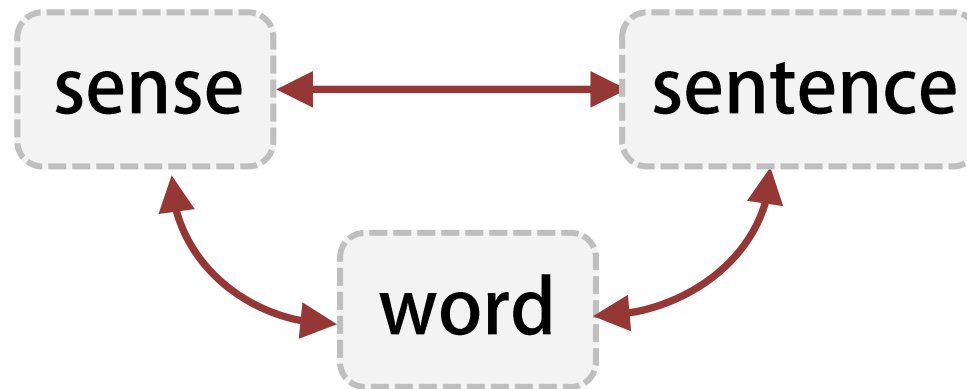
word

All lexical items
this way



Advantage 2

Cross-level semantic similarity



A large and imposing house

vs.

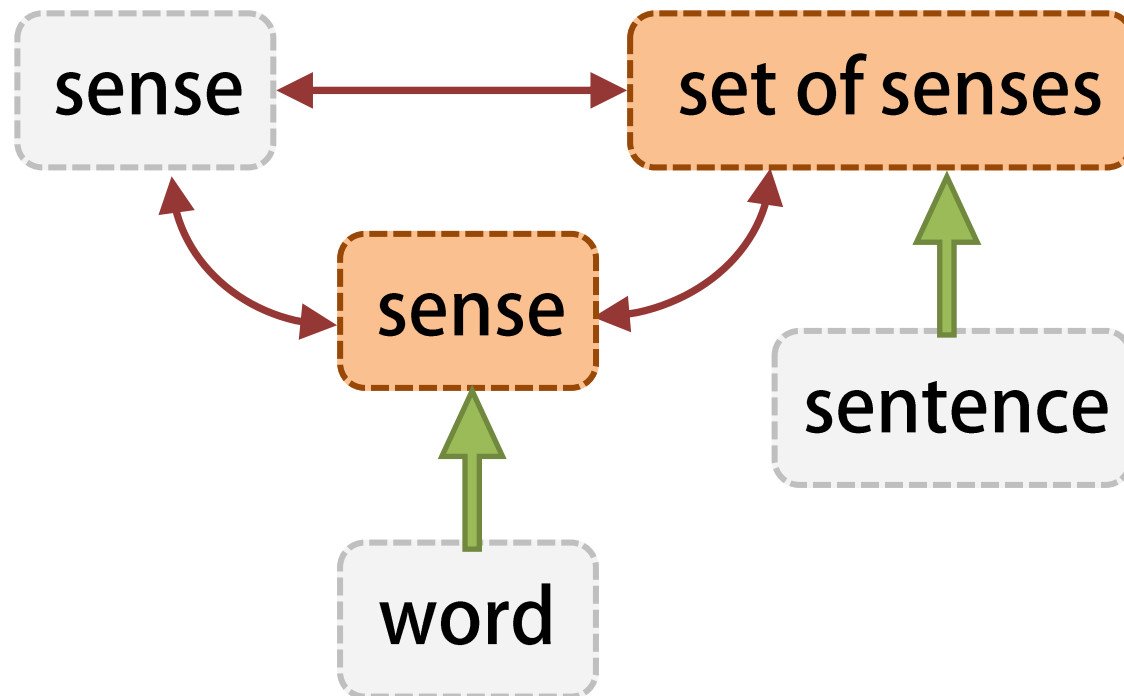
Mansion

vs.

Residence#3

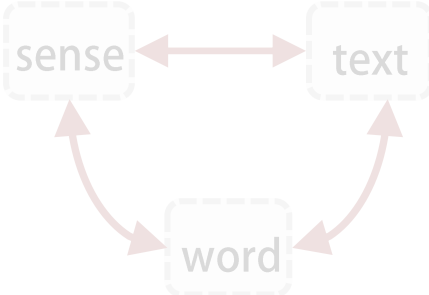
Advantage 3

Sense-level operation



Outline

Introduction



Methodology



Experiments

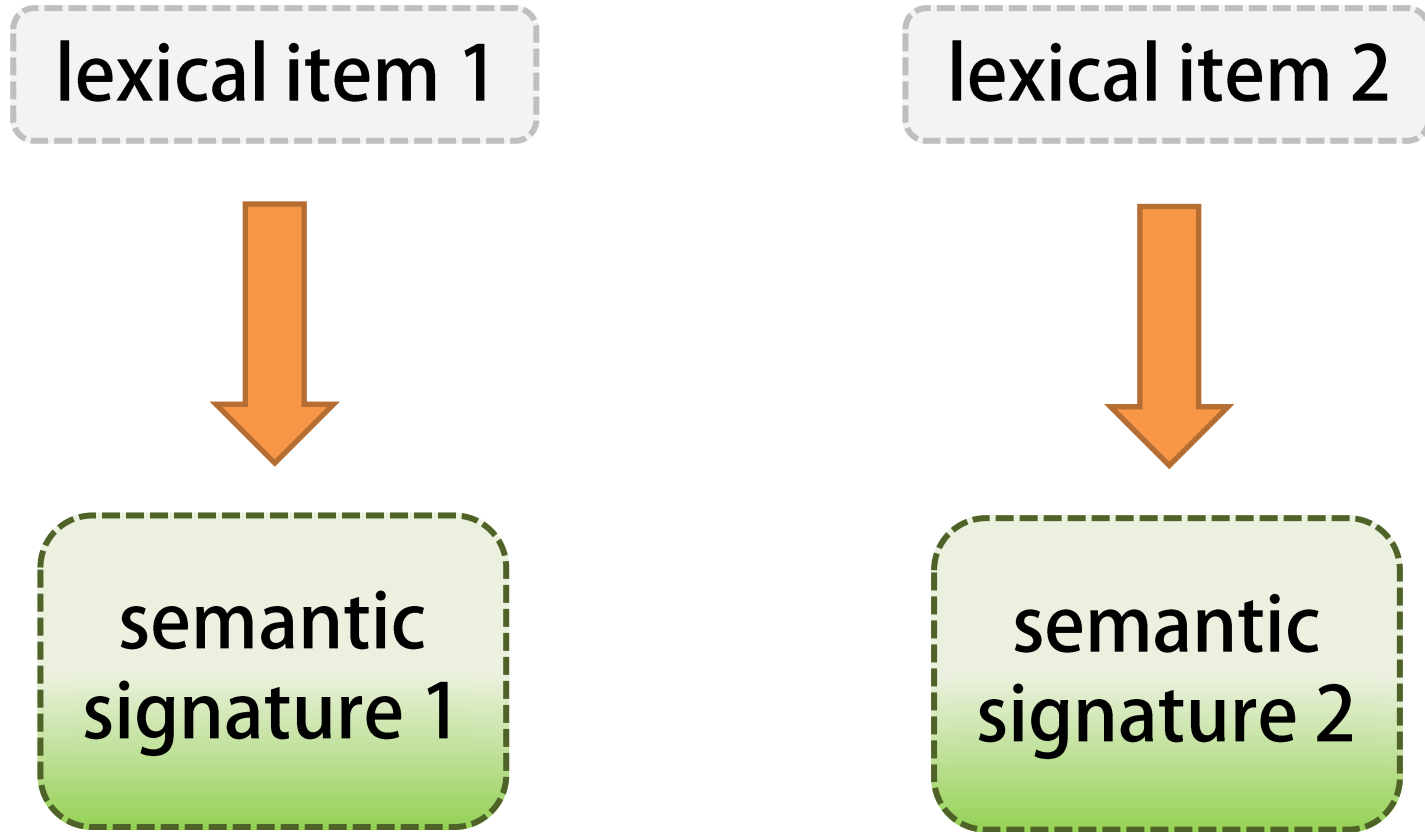


How Does it work?

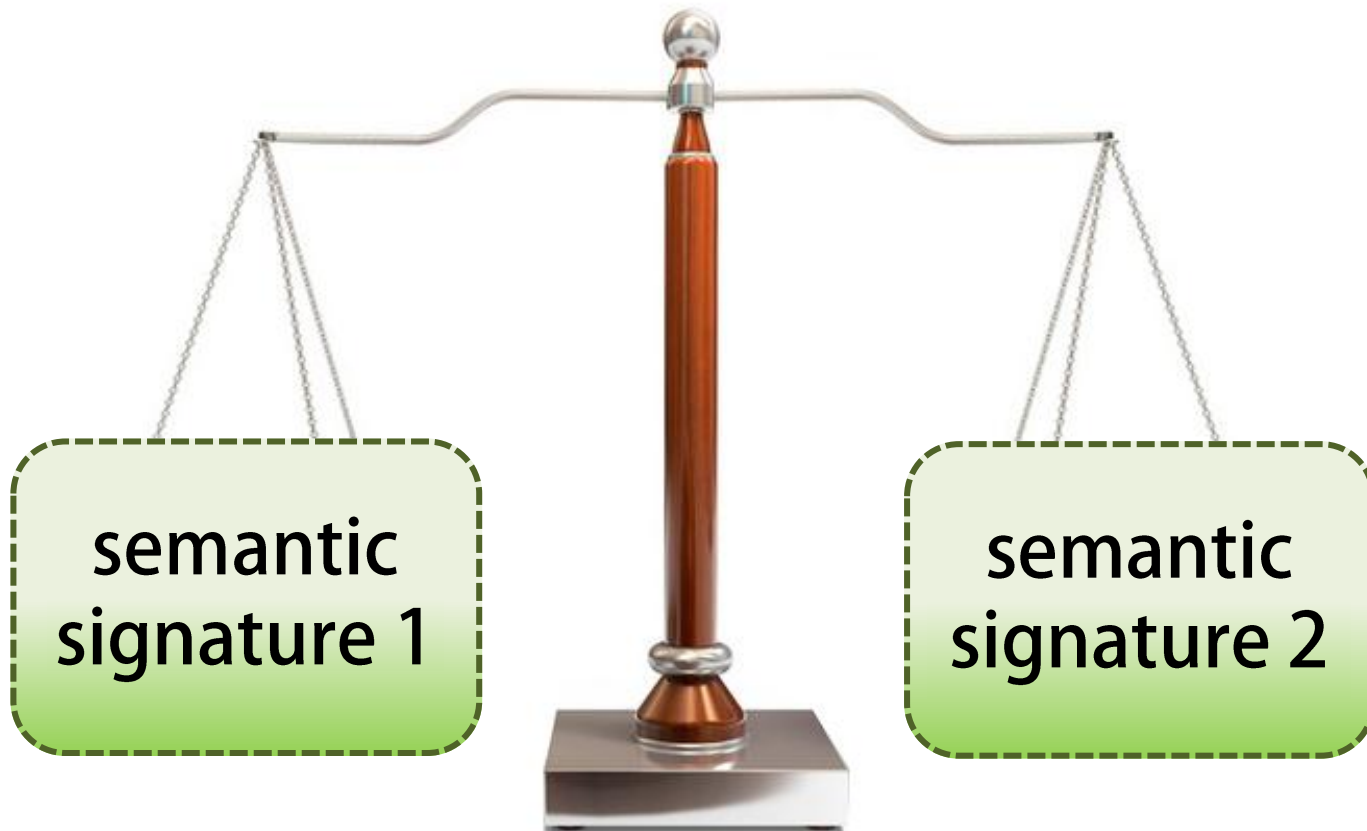
lexical item 1

lexical item 2

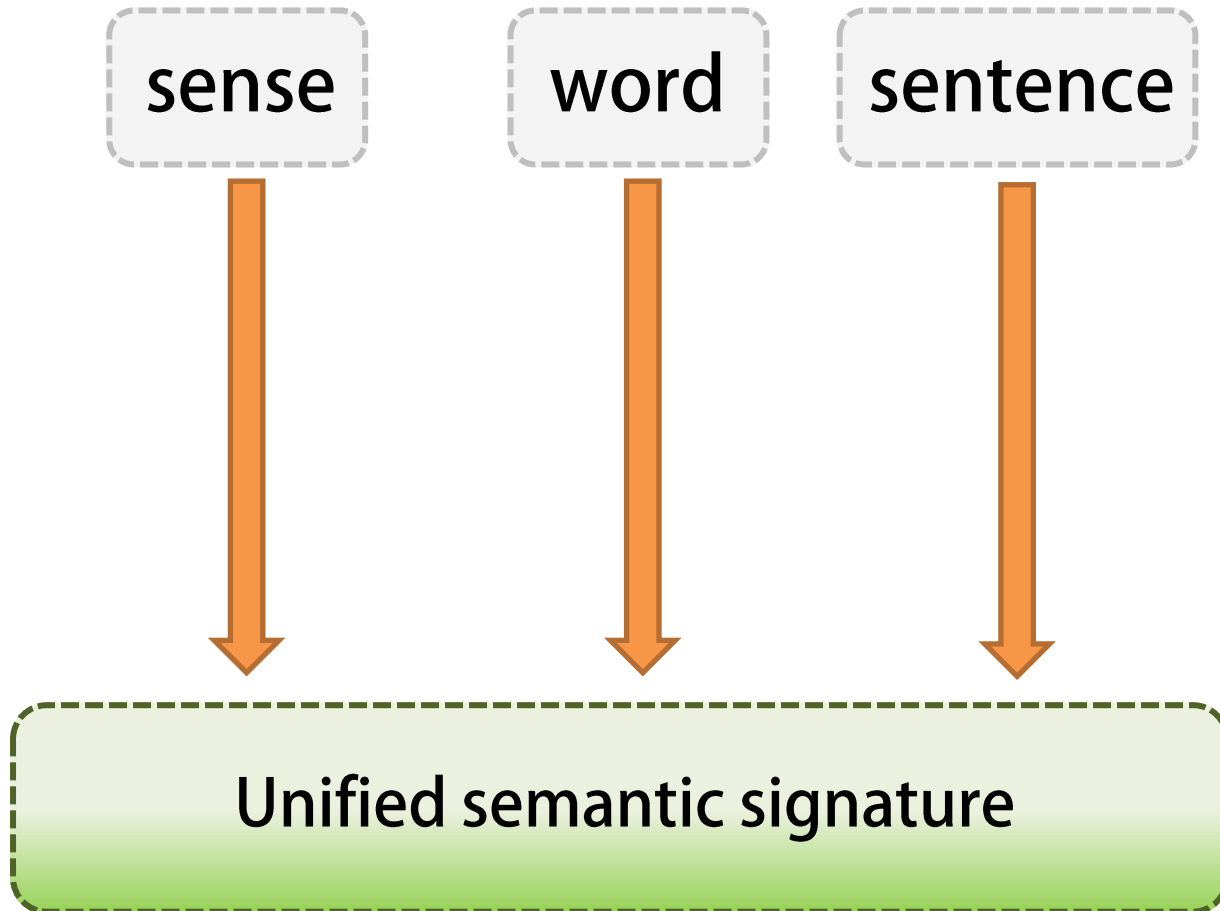
How Does it work?



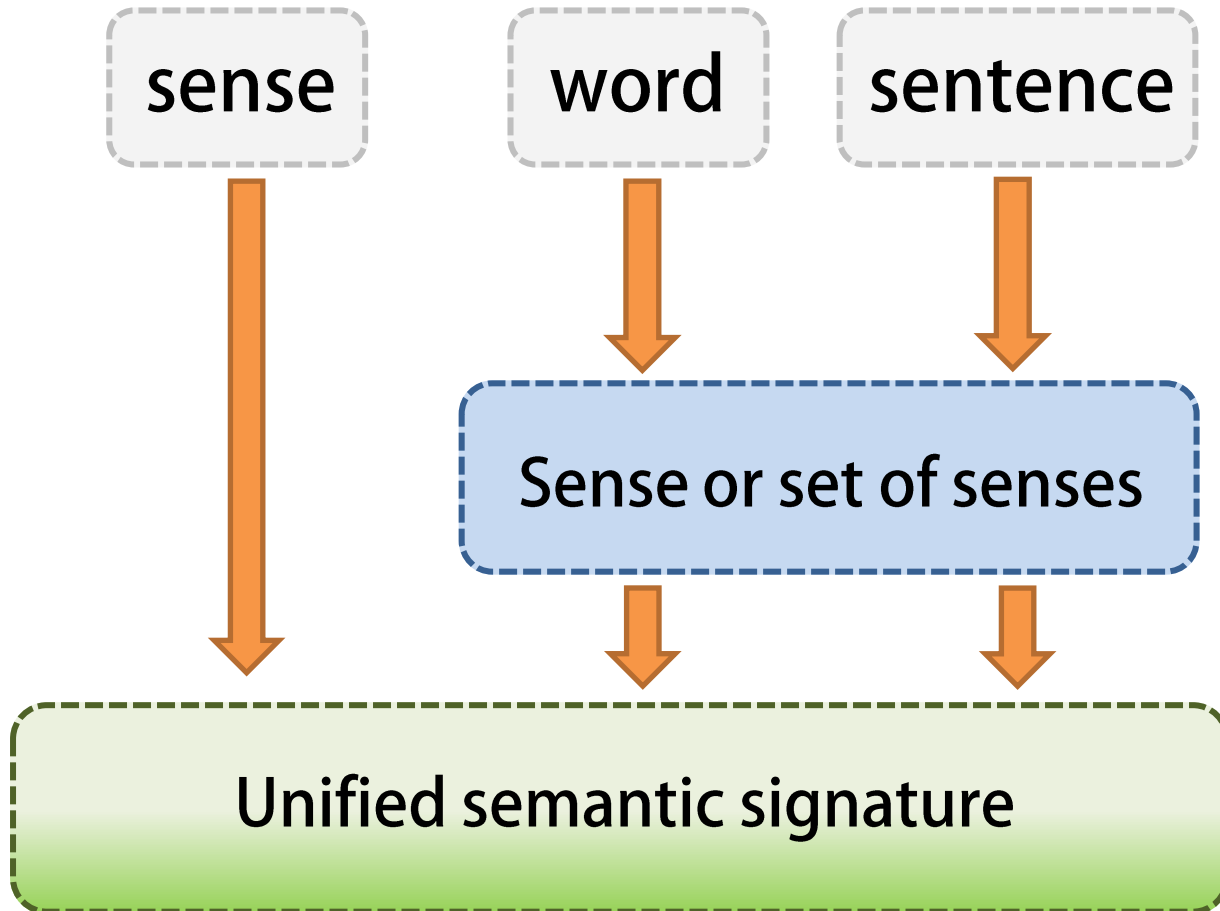
How Does it work?



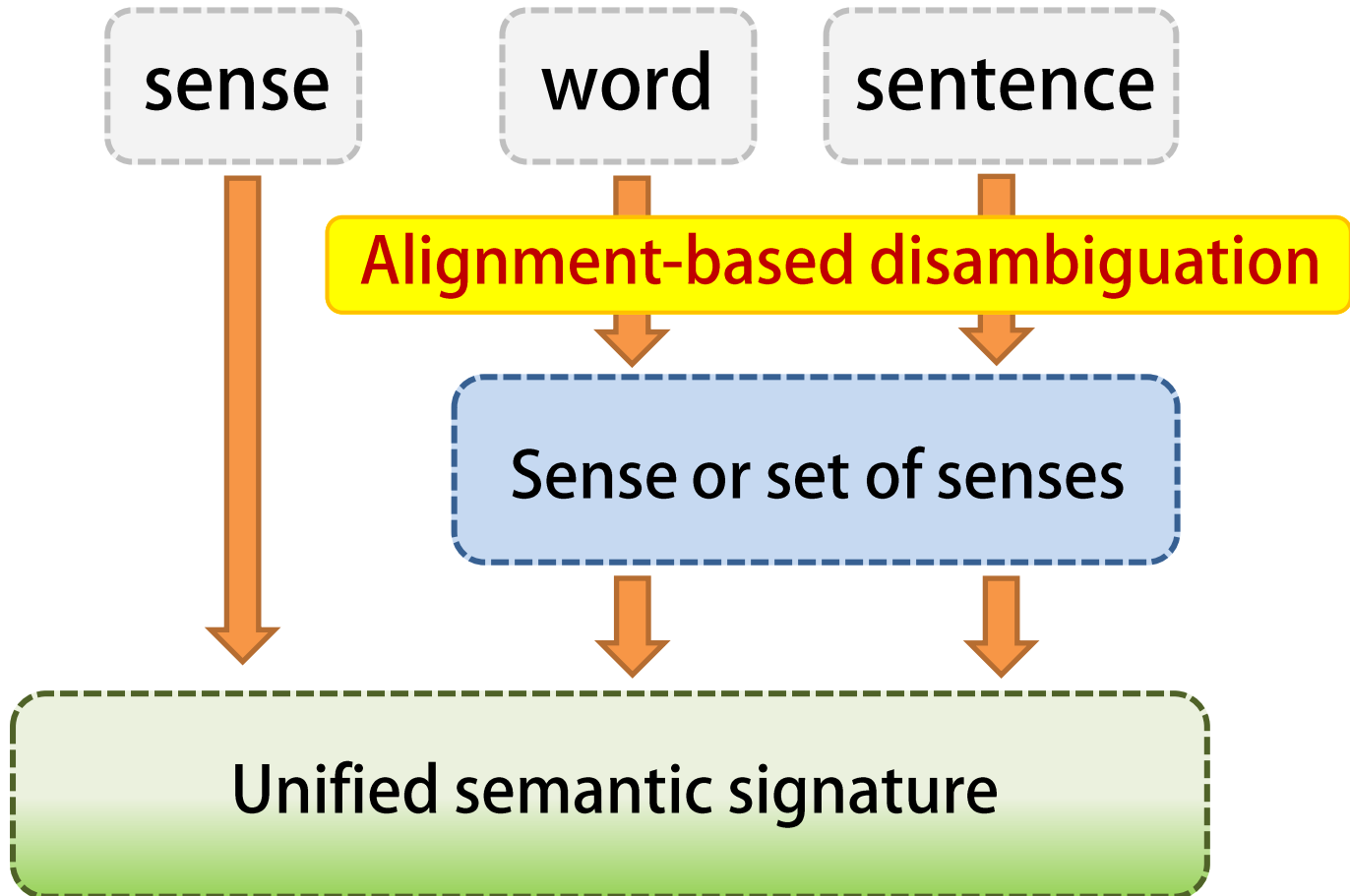
Semantic Signature



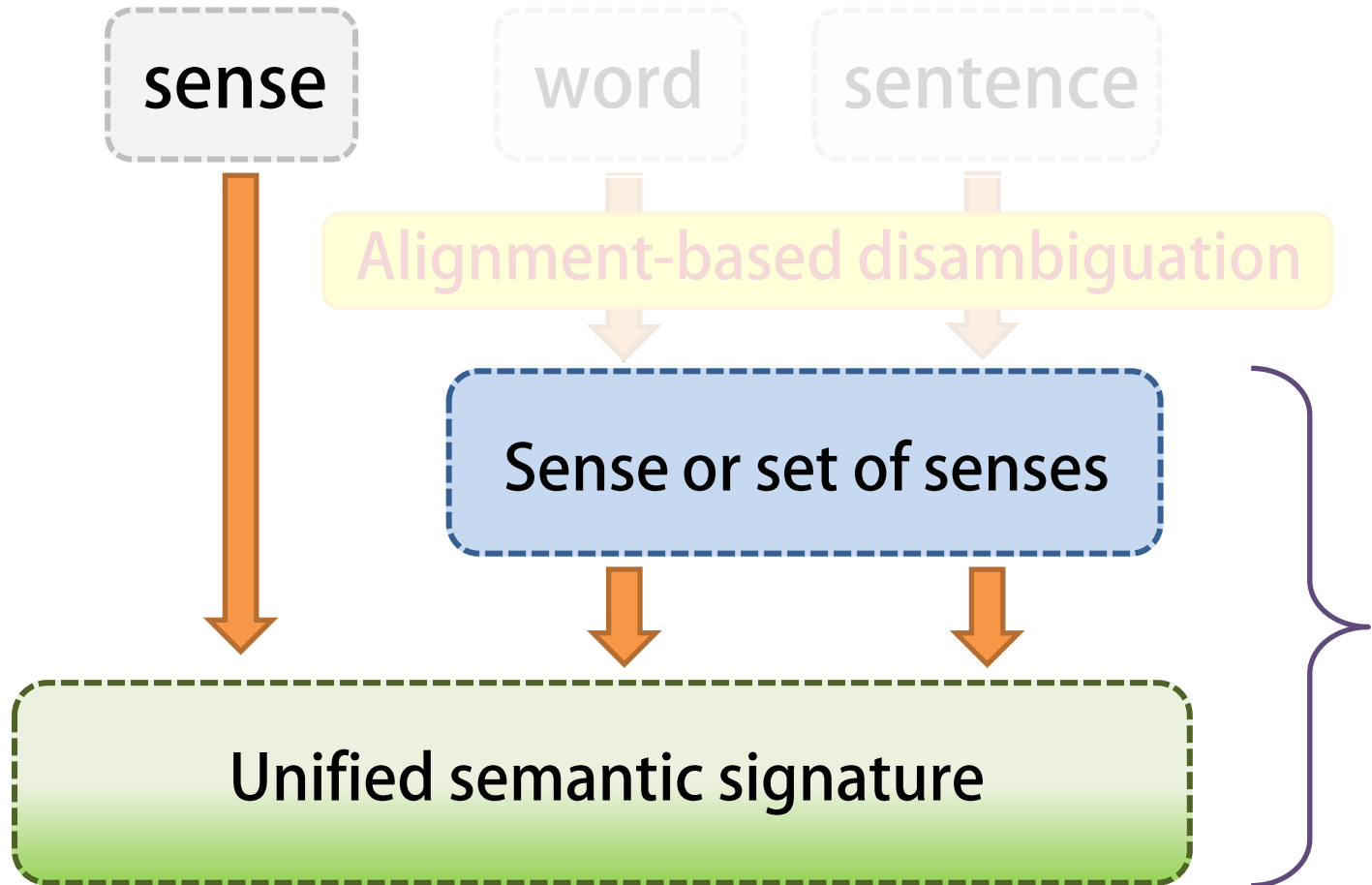
Semantic Signature



Semantic Signature



Semantic Signature



Semantic Signature

A woman is frying food

Semantic Signature



A woman is frying food

Semantic Signature



A woman is frying food



Semantic Signature



A woman is frying food



Semantic Signature



A woman is frying food



Semantic Signature



A woman is frying food



Semantic Signature

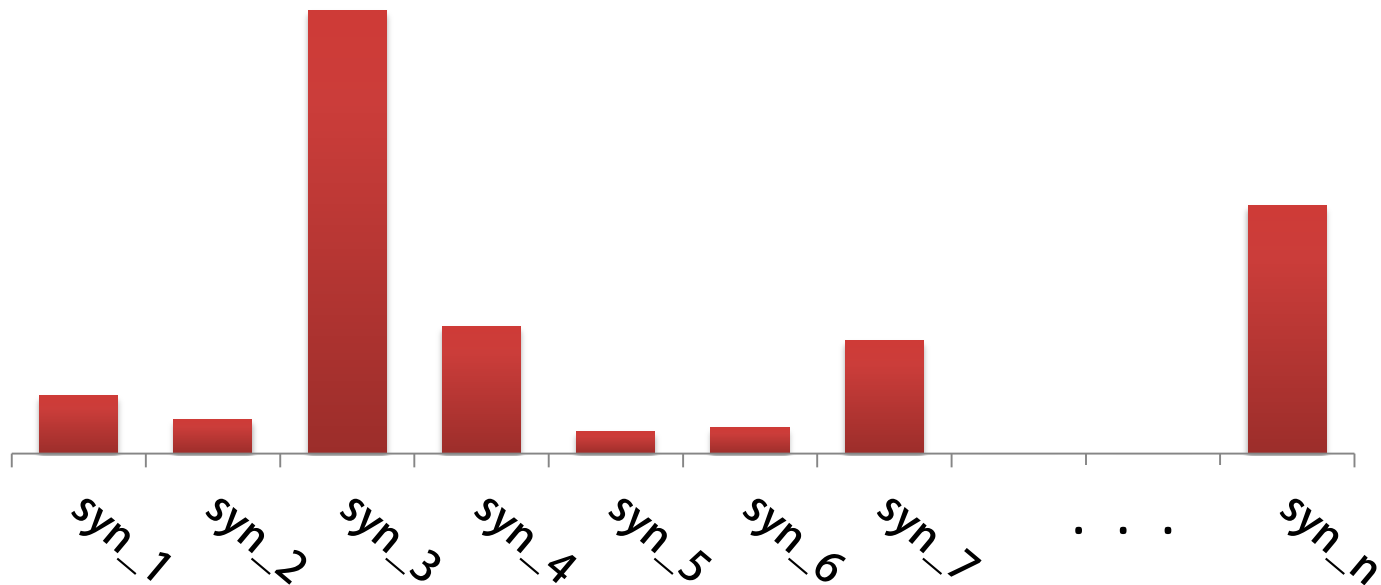


A woman is frying food



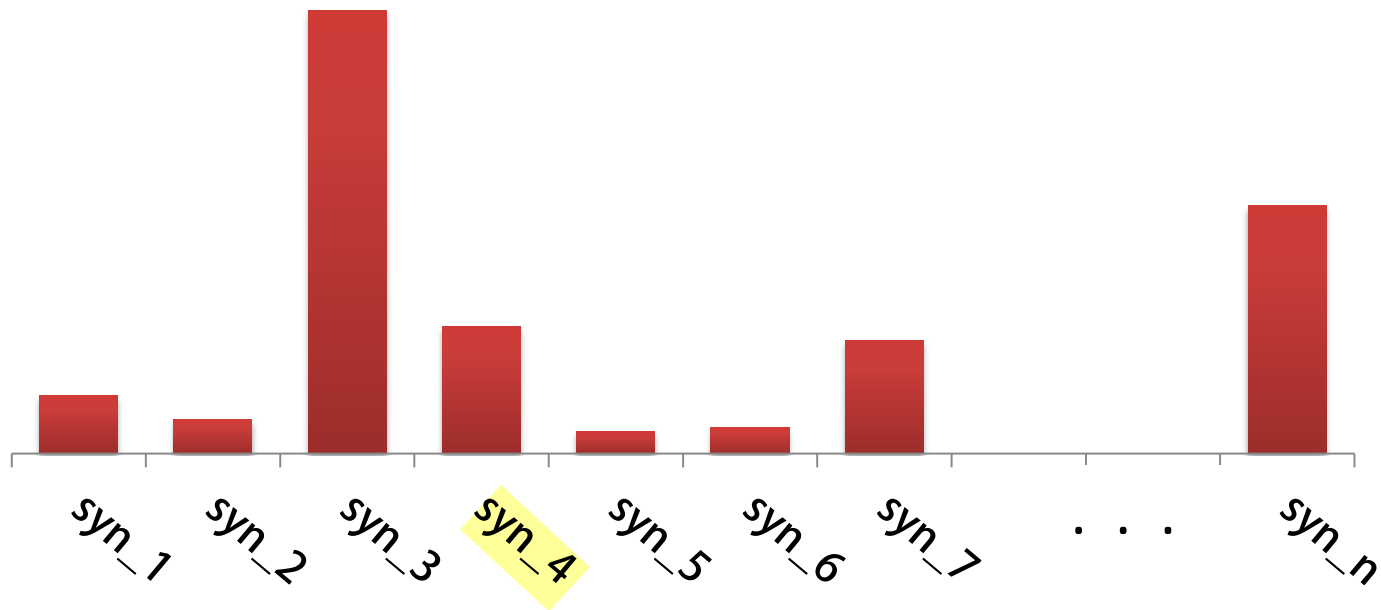
Semantic Signature

Distributional representation
over all synsets in WordNet



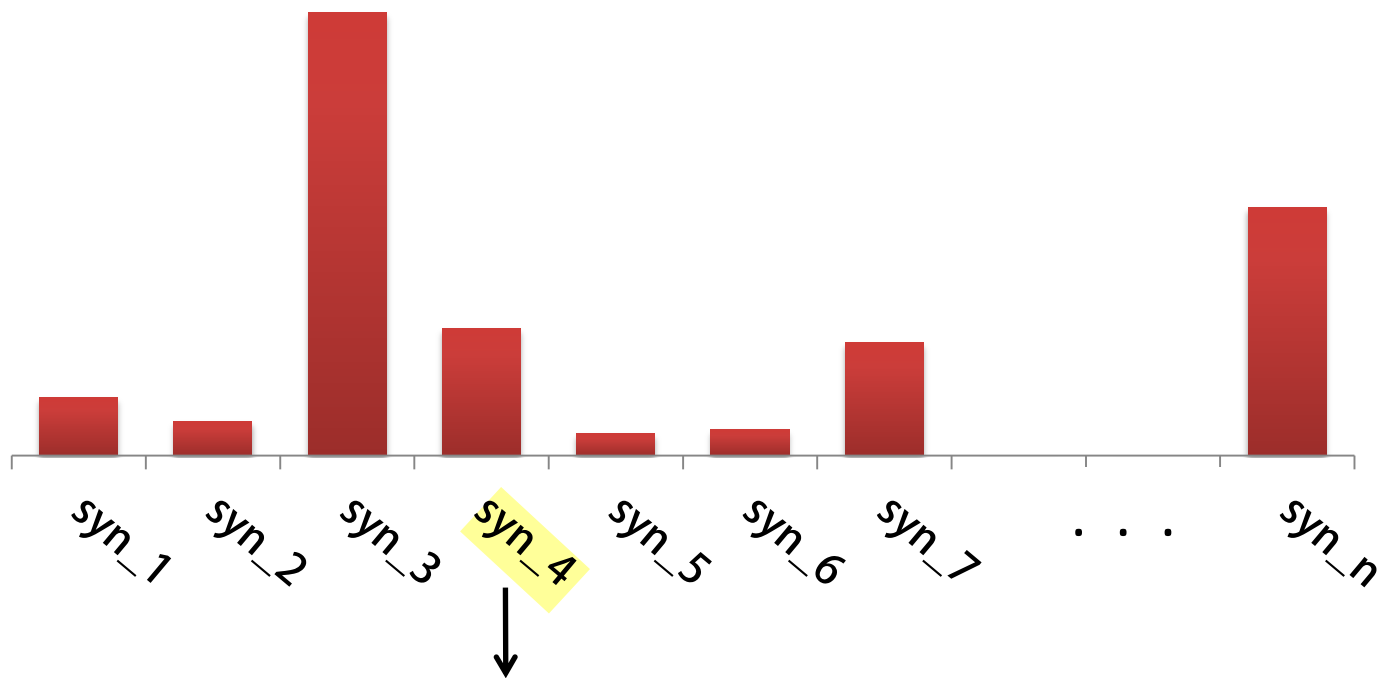
Semantic Signature

Distributional representation
over all synsets in WordNet



Semantic Signature

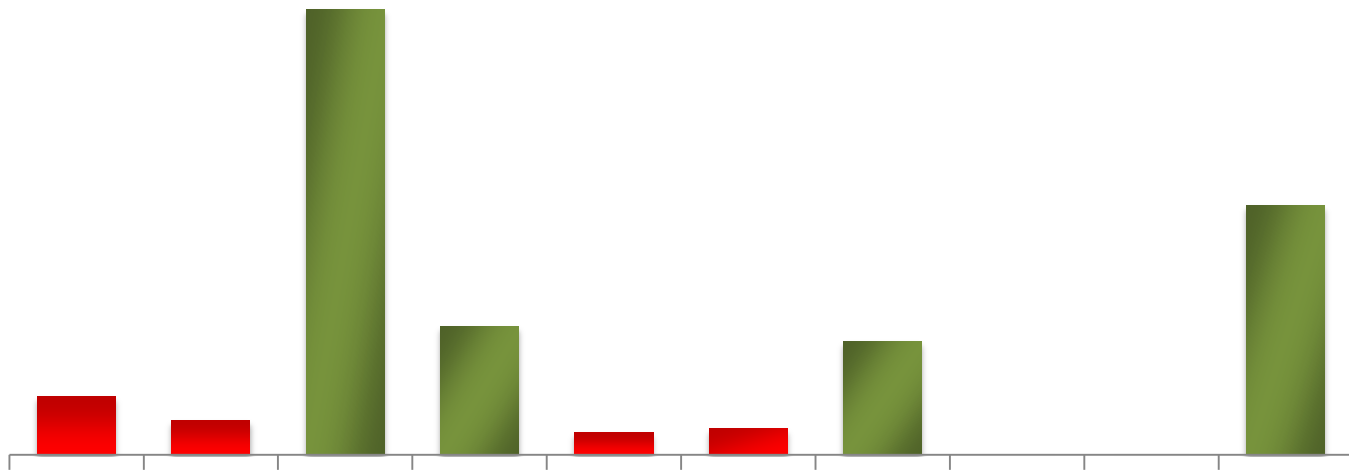
Distributional representation
over all synsets in WordNet



Importance of this synset (syn_4) for our lexical item

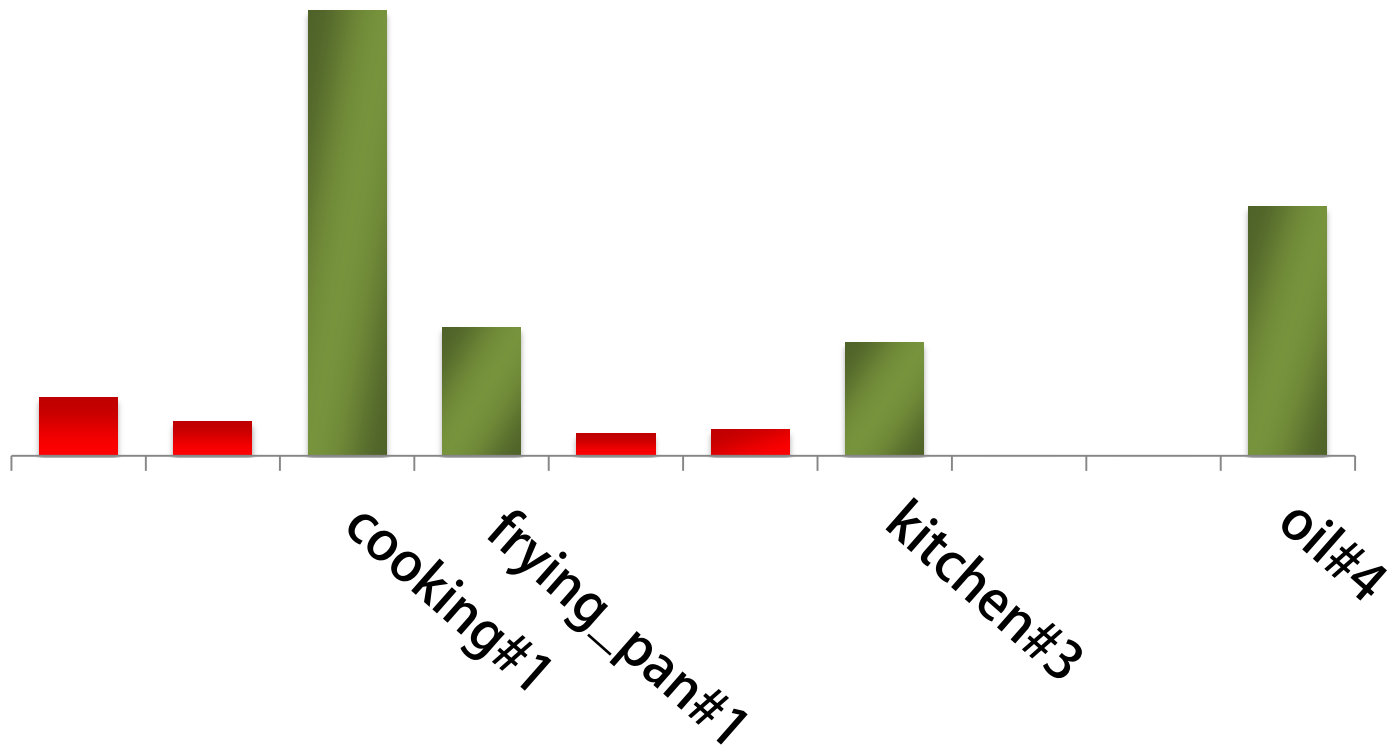
Semantic Signature

a woman is frying food



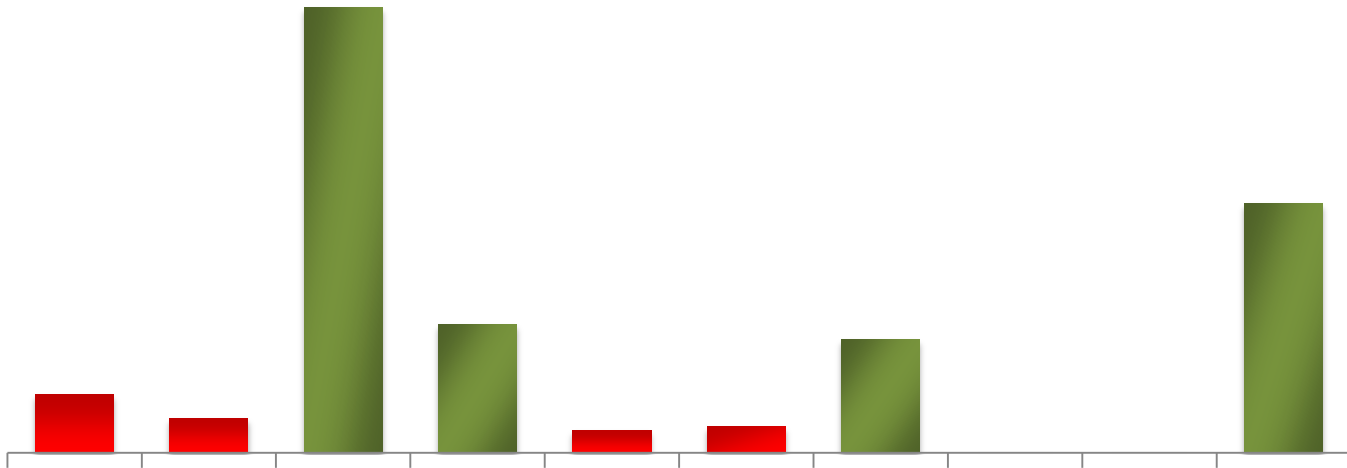
Semantic Signature

a woman is frying food



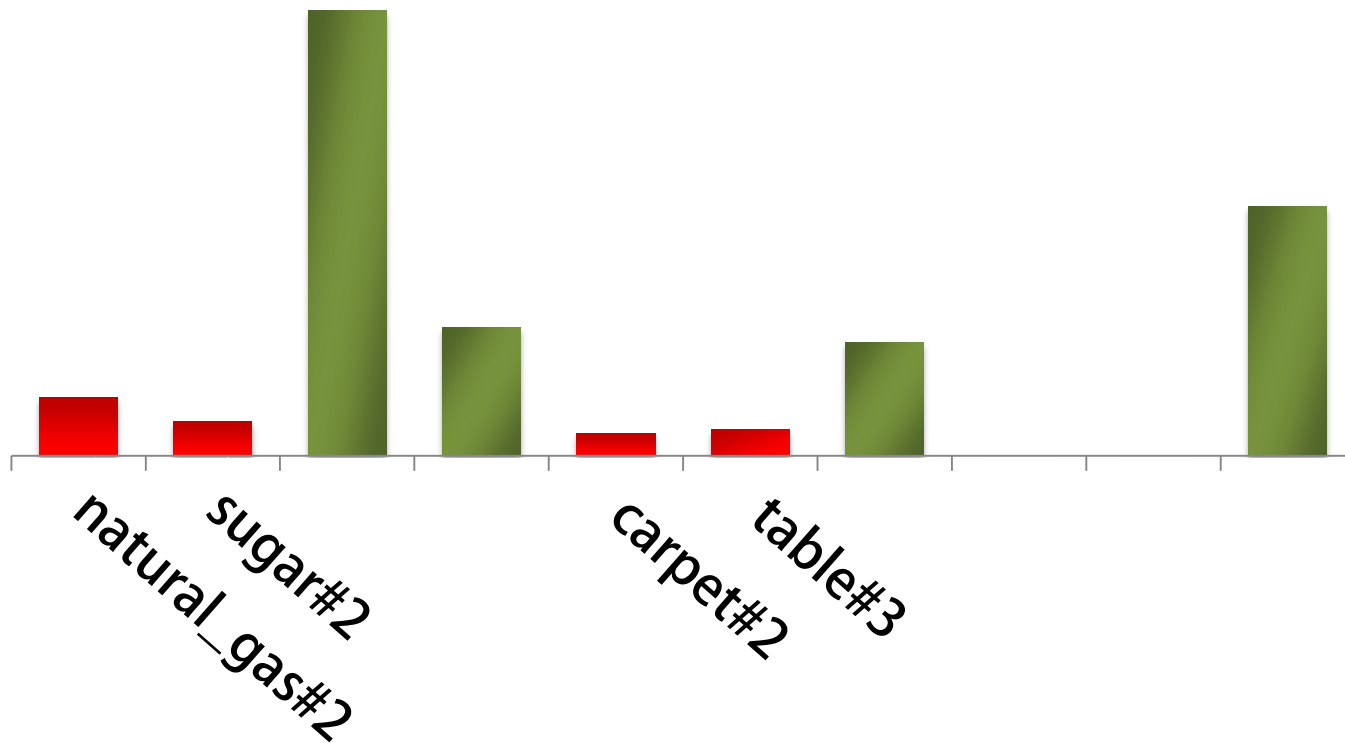
Semantic Signature

a woman is frying food



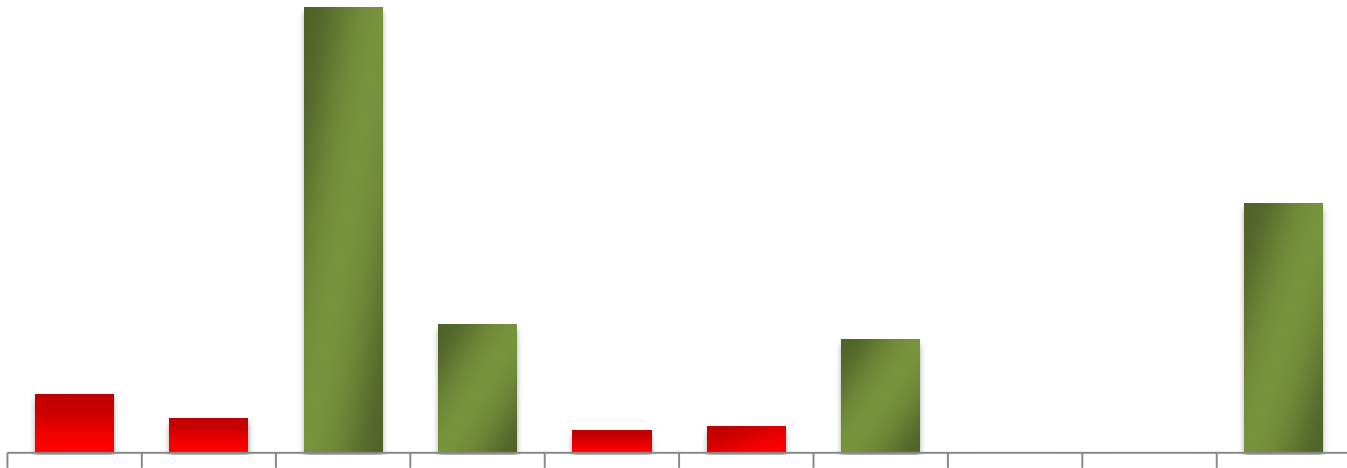
Semantic Signature

a woman is frying food



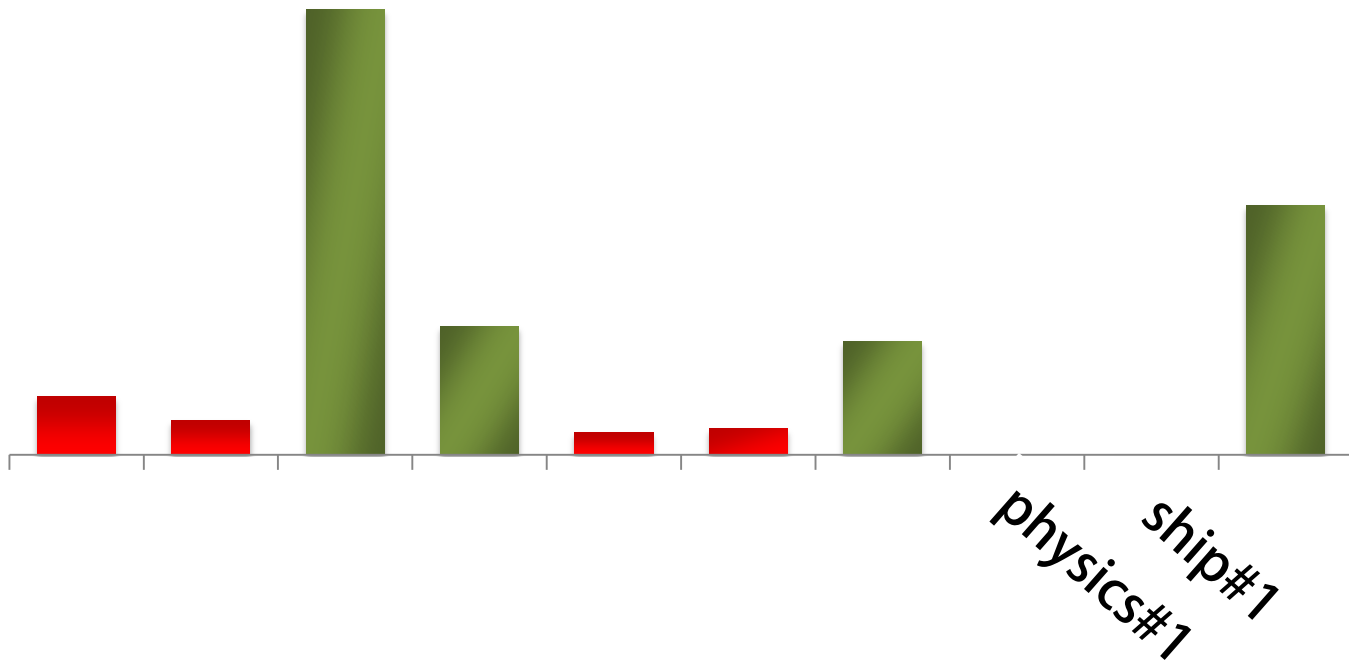
Semantic Signature

a woman is frying food

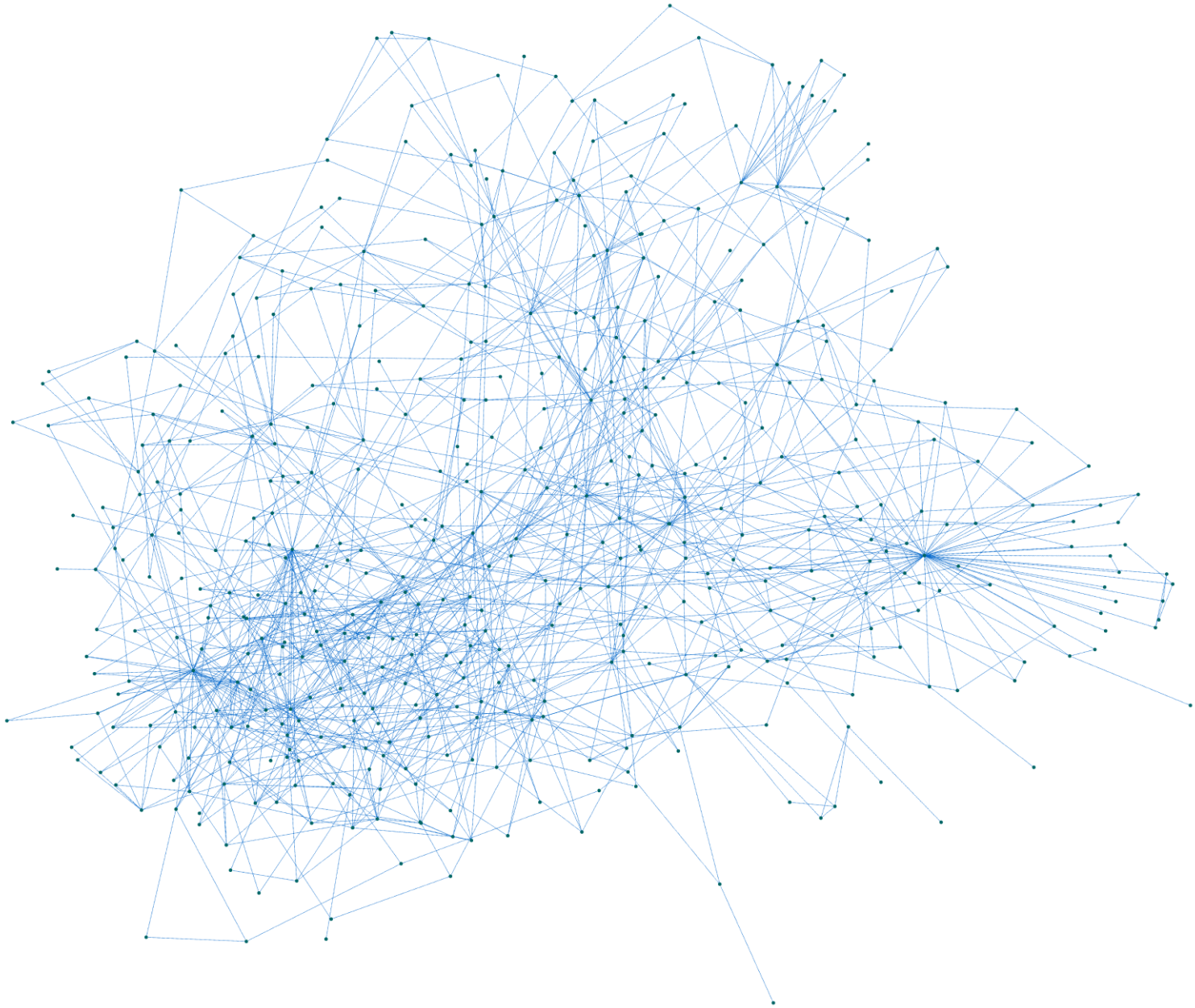


Semantic Signature

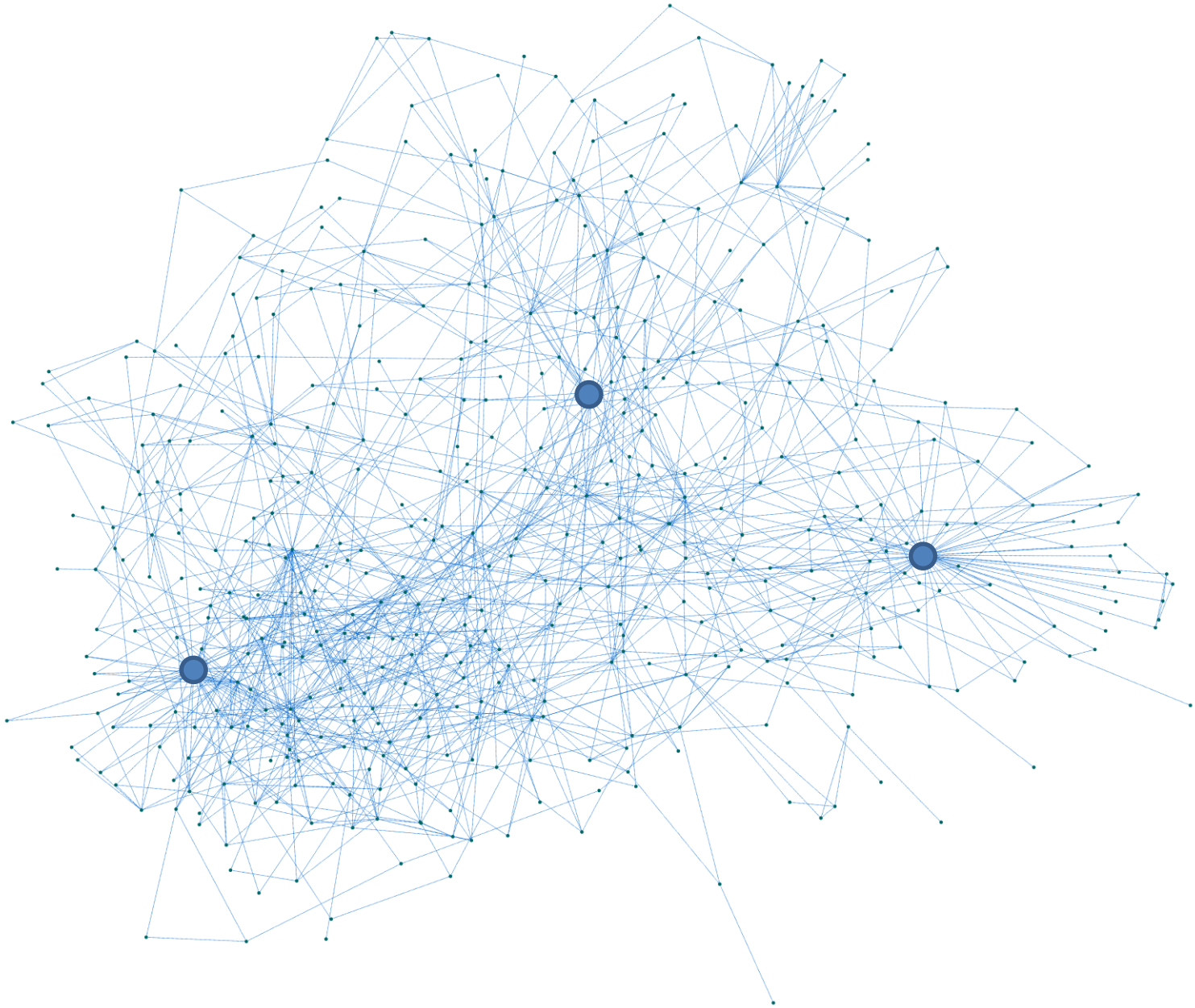
a woman is frying food



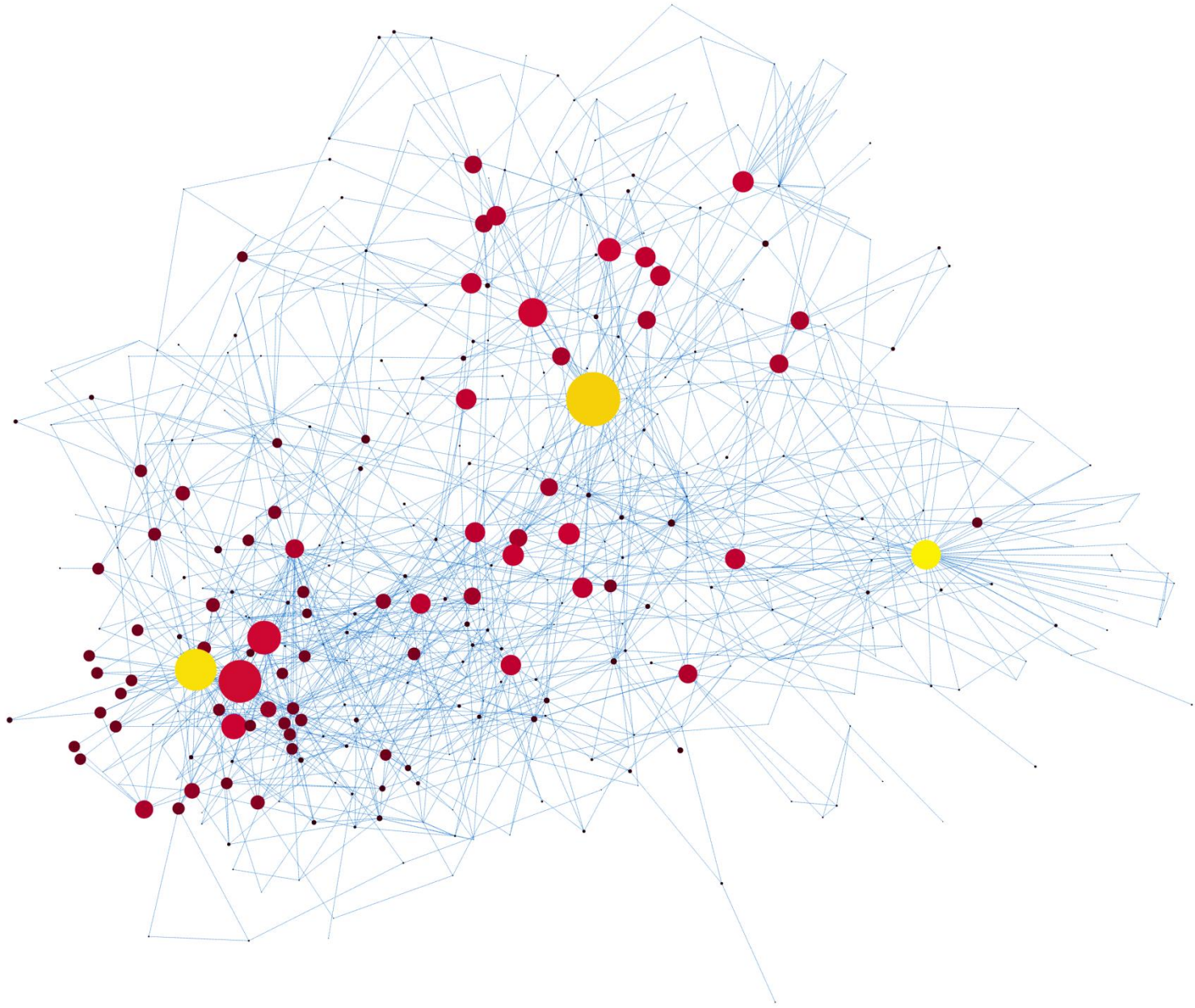
Personalized PageRank



Personalized PageRank

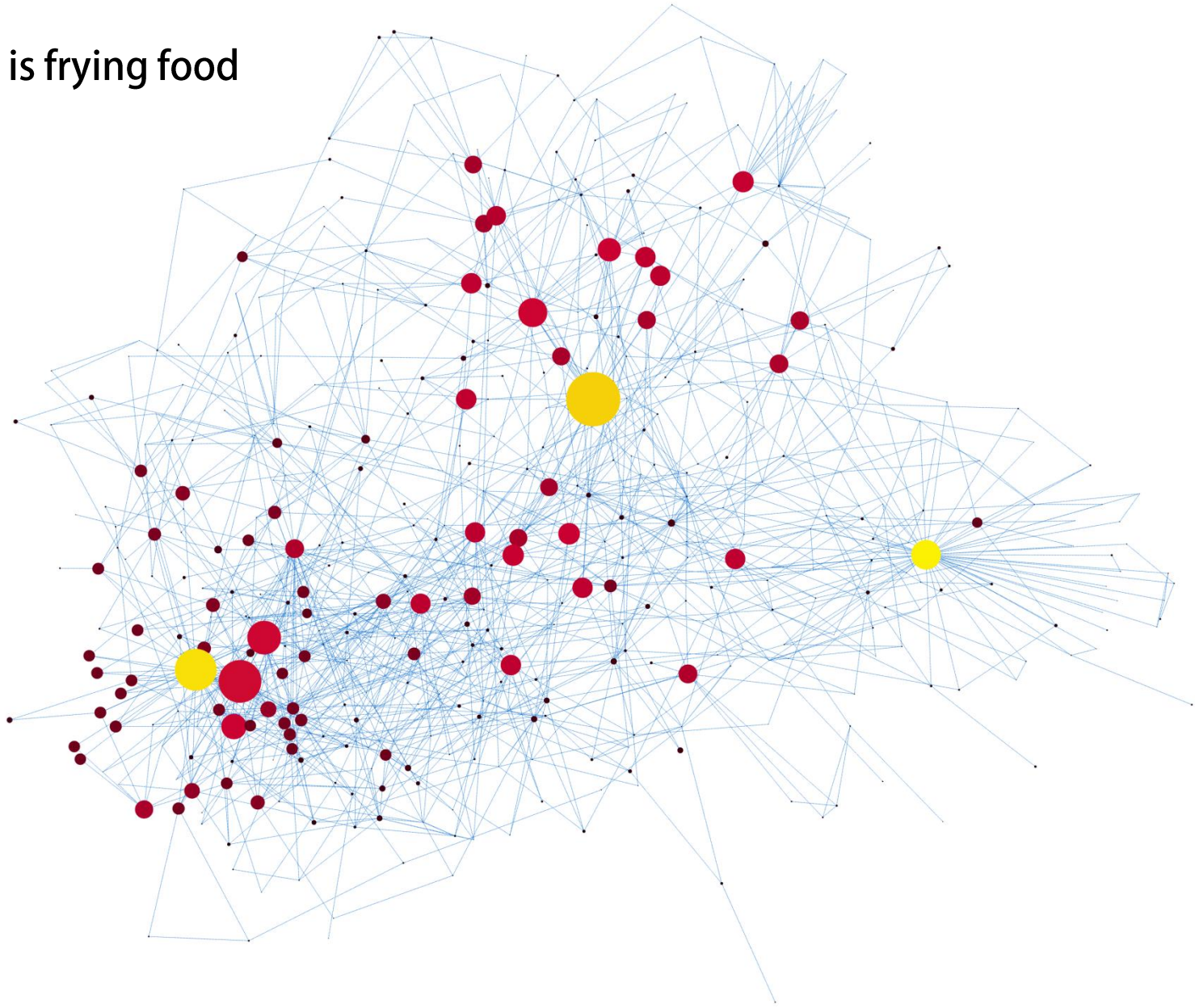


Personalized PageRank



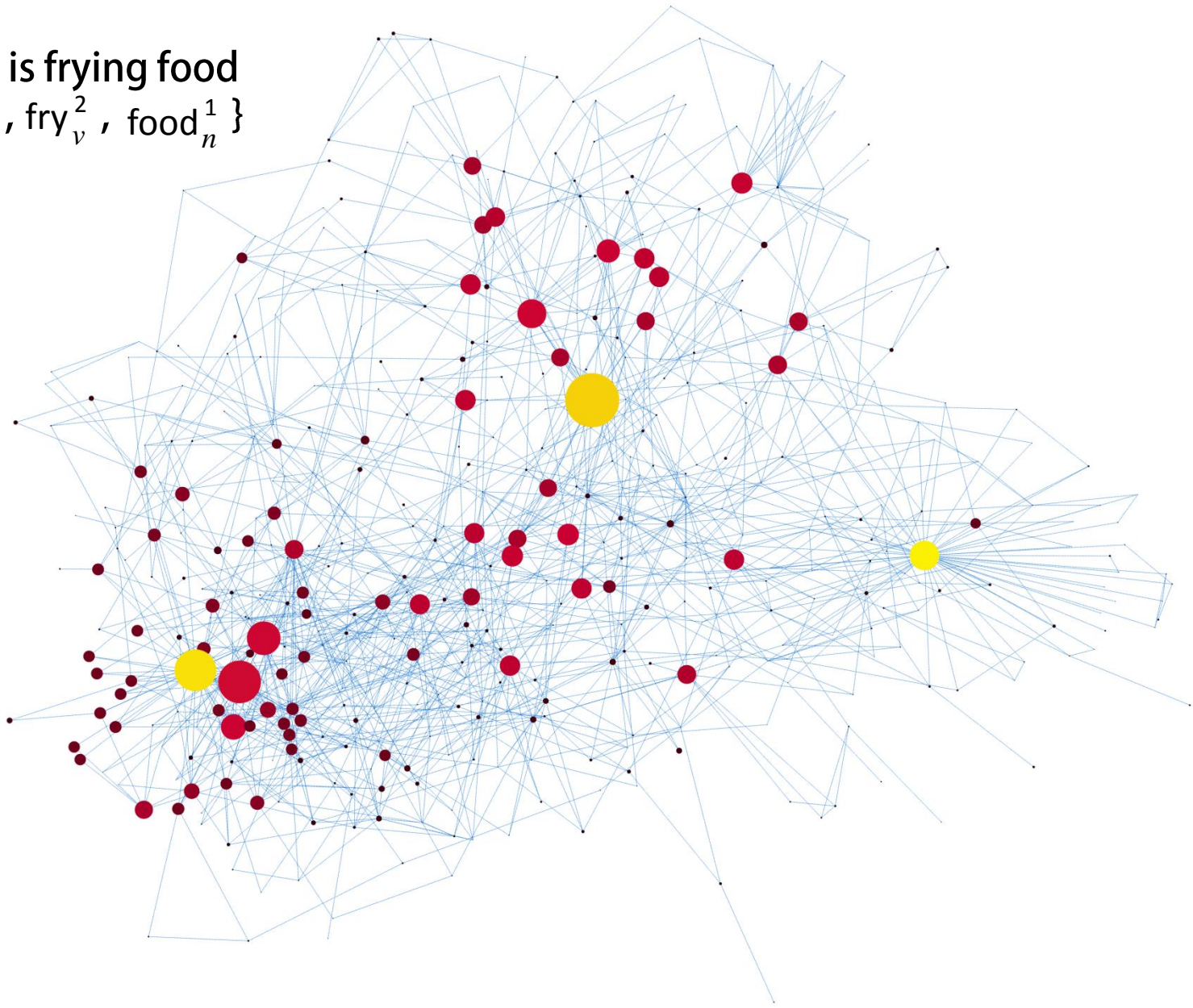
Personalized PageRank

a woman is frying food



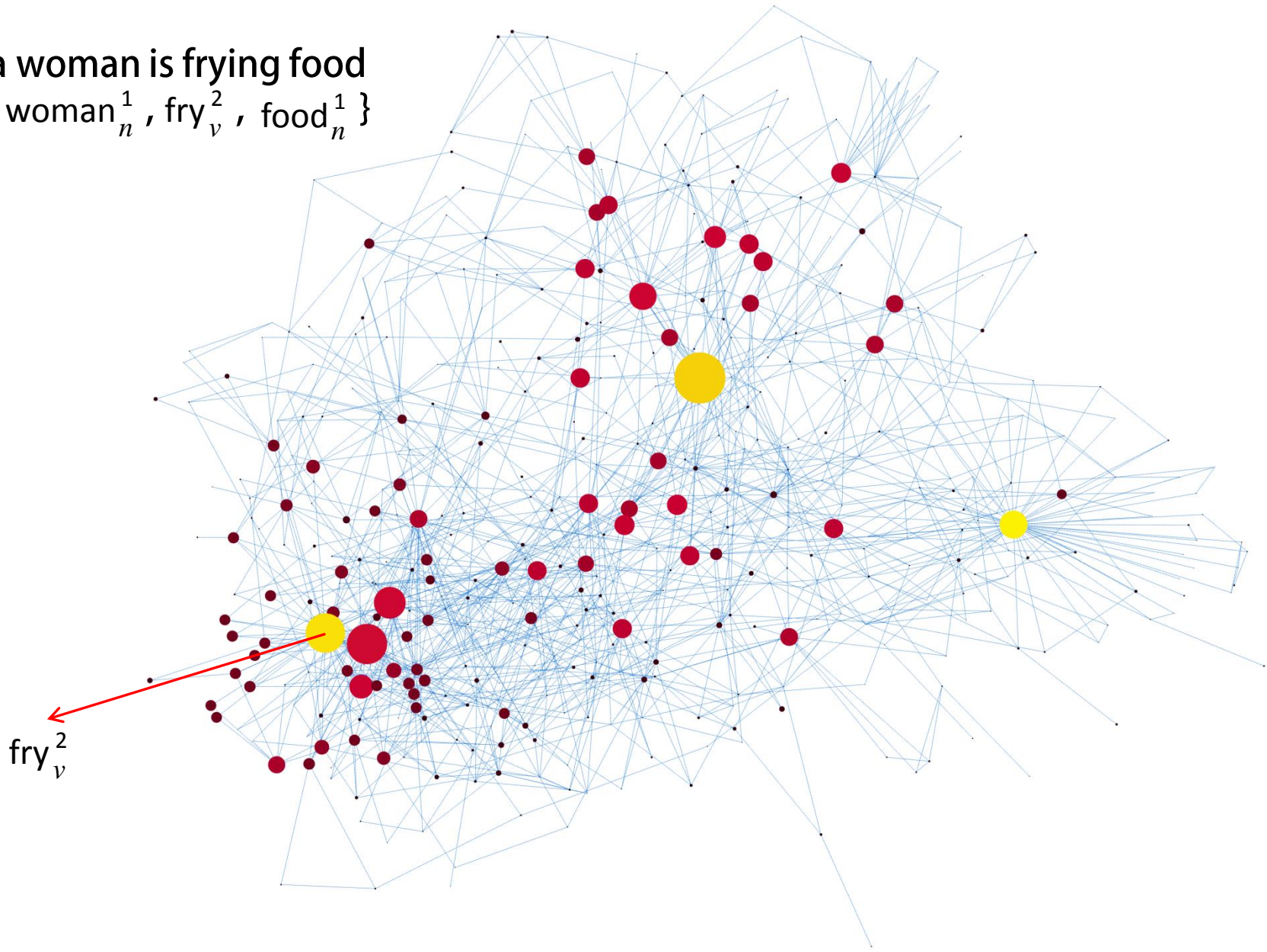
Personalized PageRank

a woman is frying food
 $\{ \text{woman}_n^1, \text{fry}_v^2, \text{food}_n^1 \}$



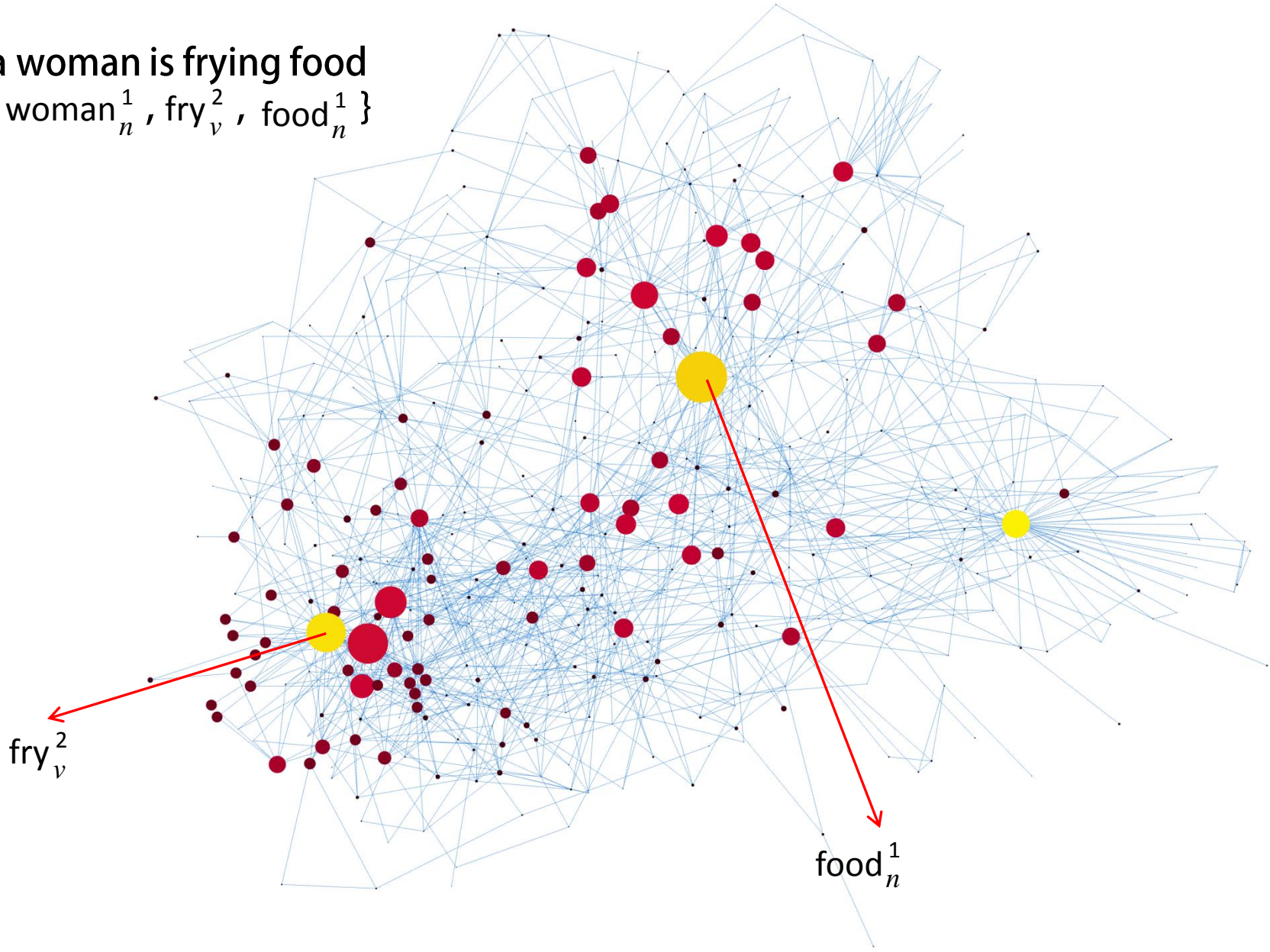
Personalized PageRank

a woman is frying food
 $\{ \text{woman}_n^1, \text{fry}_v^2, \text{food}_n^1 \}$



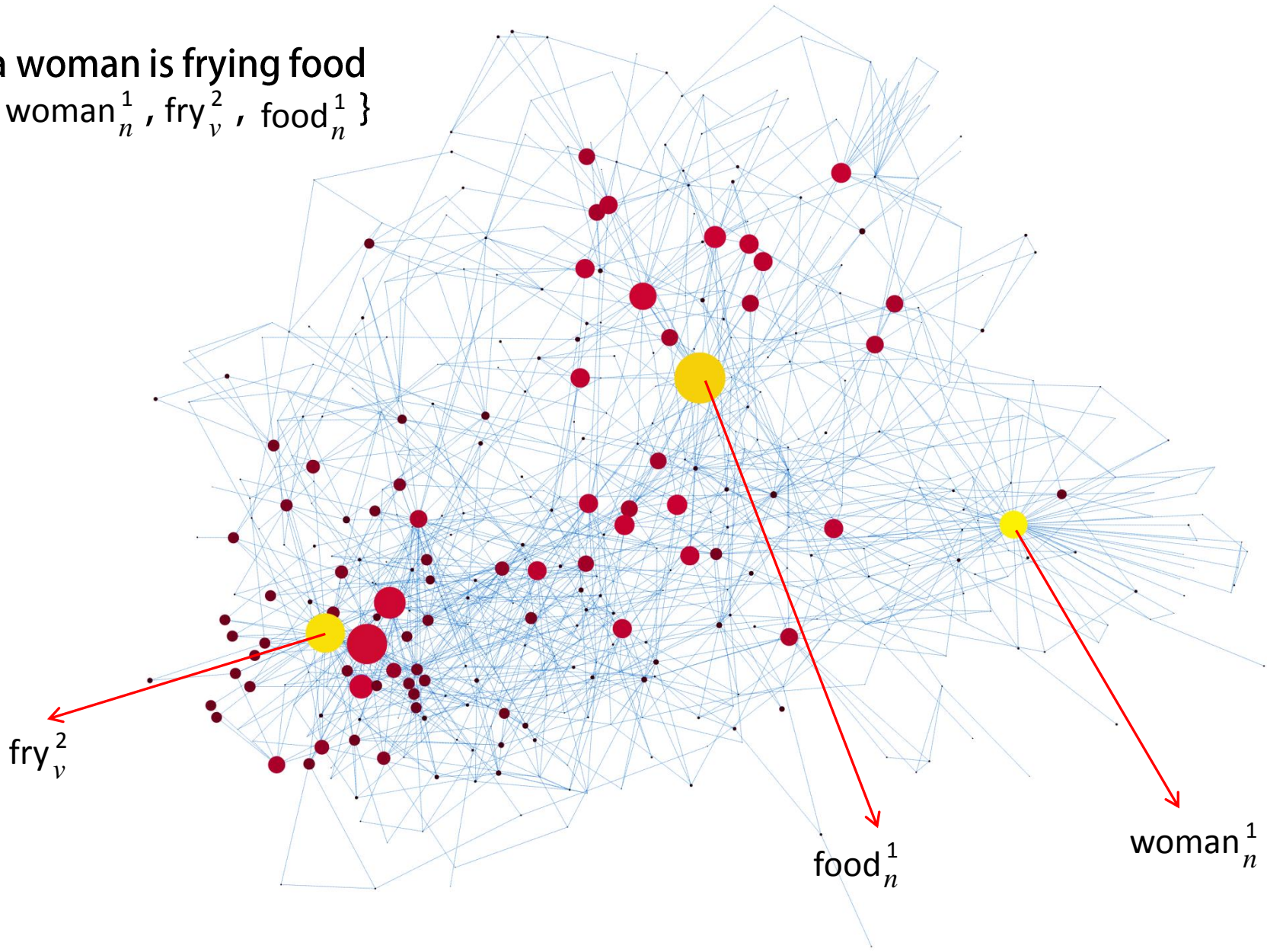
Personalized PageRank

a woman is frying food
 $\{ \text{woman}_n^1, \text{fry}_v^2, \text{food}_n^1 \}$



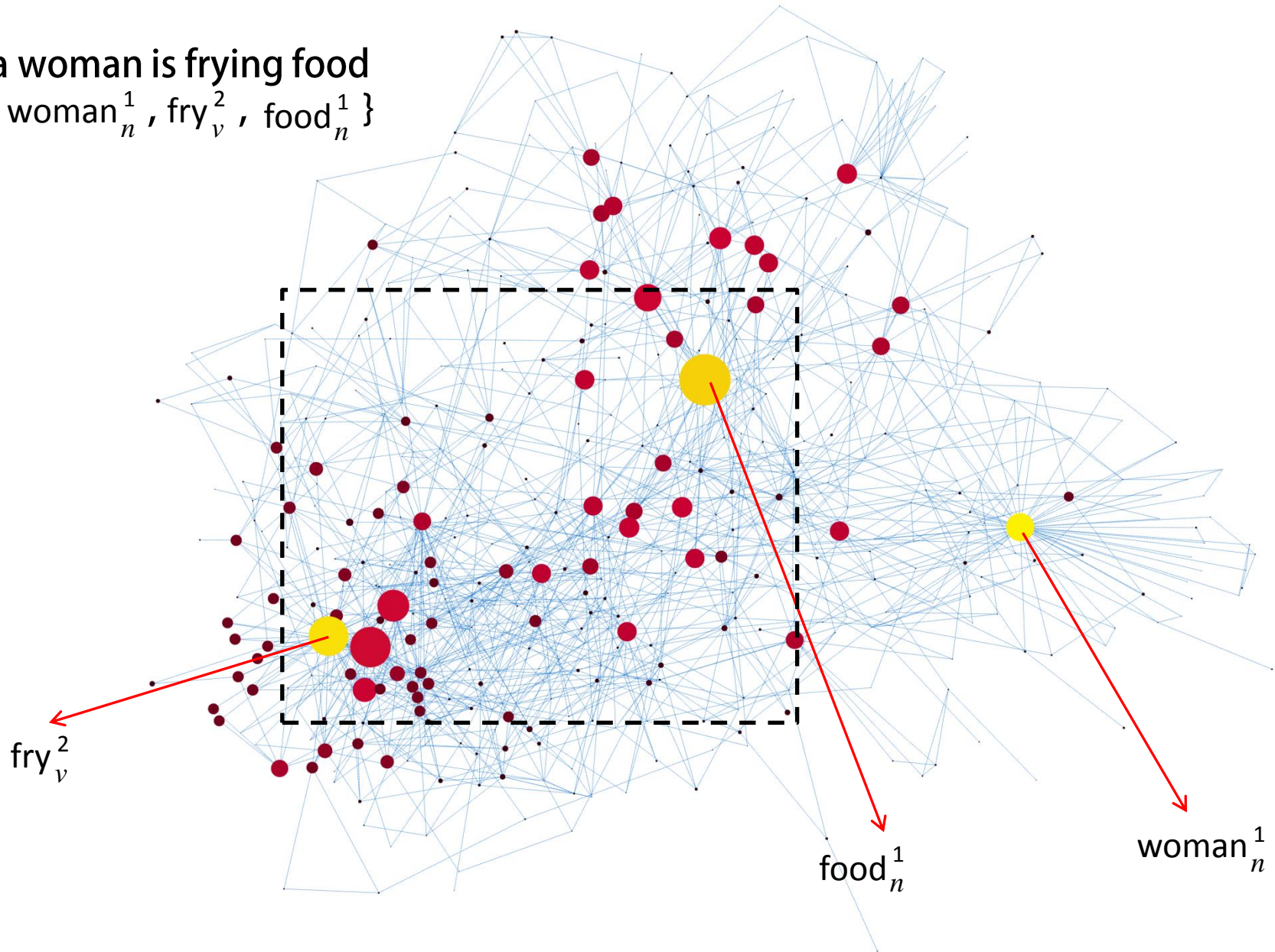
Personalized PageRank

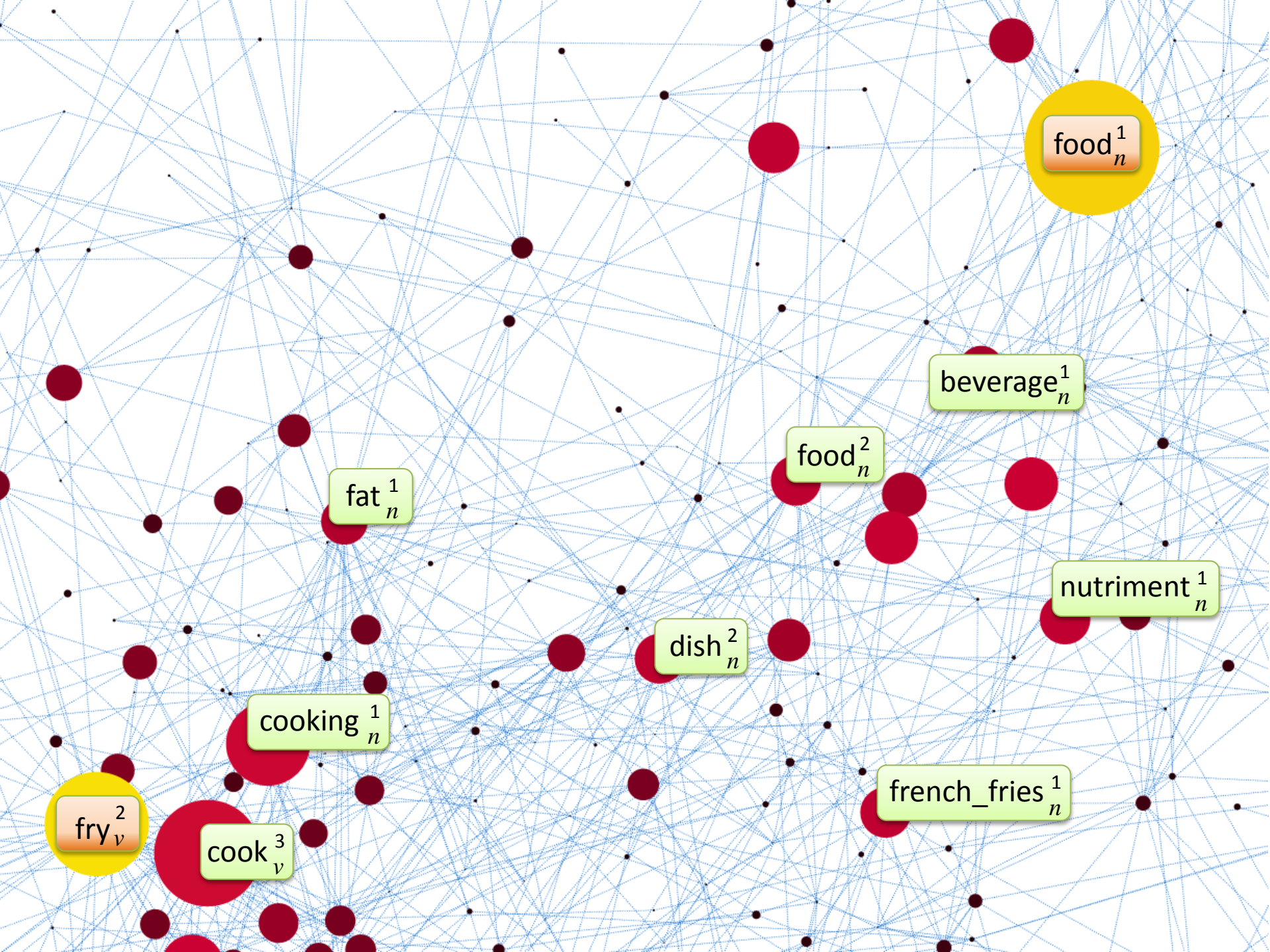
a woman is frying food
 $\{ \text{woman}_n^1, \text{fry}_v^2, \text{food}_n^1 \}$



Personalized PageRank

a woman is frying food
 $\{ \text{woman}_n^1, \text{fry}_v^2, \text{food}_n^1 \}$





A network graph with a dense web of light blue lines connecting small grey nodes. Several nodes are highlighted with larger, semi-transparent circles in shades of pink and red. One node in the upper right is highlighted with a yellow circle and contains the text 'food' with a subscript 'n' and a superscript '1'. Another node in the lower right is highlighted with a light green circle and contains the text 'beverage' with a subscript 'n' and a superscript '1'.

food_n¹

beverage_n¹

These weights form a semantic signature

A network graph similar to the one above, but with a larger number of highlighted nodes. The highlighted nodes are represented by circles of various sizes and colors (pink, red, yellow, light green). Each highlighted node contains a text label with a subscript 'n' and a superscript '1' or '2' or '3'. The labels include 'nutriment', 'dish', 'cooking', 'fry', 'cook', and 'french_fries'.

nutriment_n¹

dish_n²

cooking_n¹

fry_v²

cook_v³

french_fries_n¹

Comparing Semantic Signatures

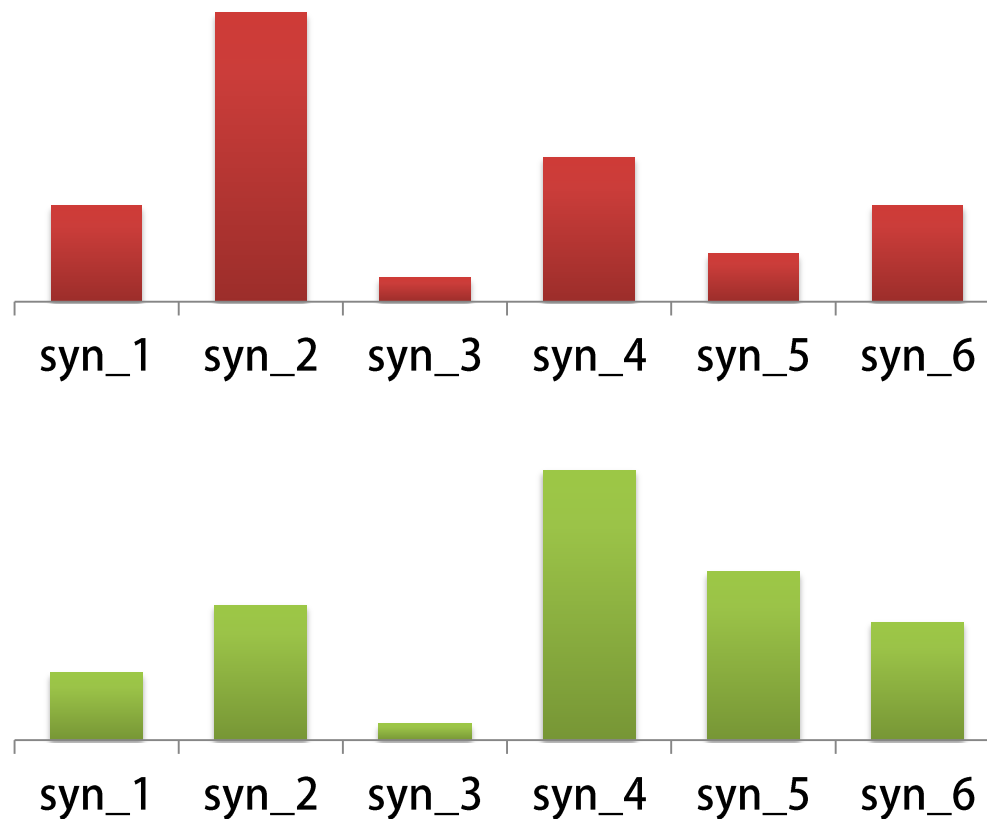


Comparing Semantic Signatures

- **Parametric**
 - Cosine
- **Non-parametric**
 - Weighted Overlap
 - Top-k Jaccard

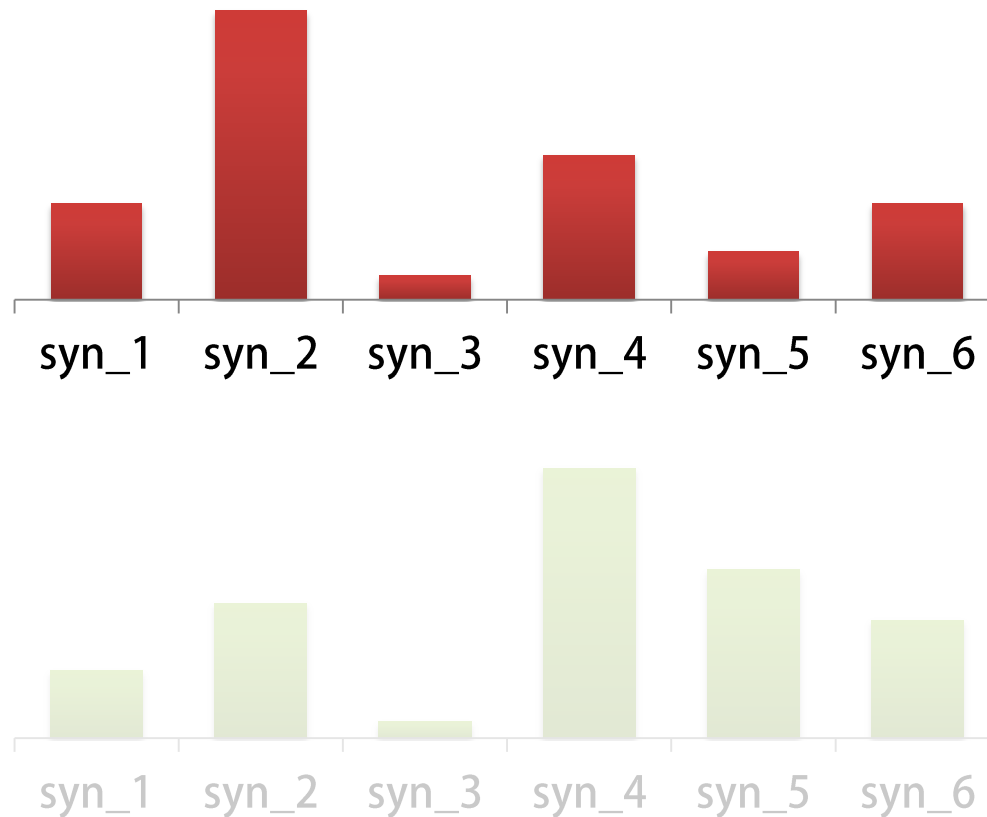
Comparing Semantic Signatures

Weighted Overlap



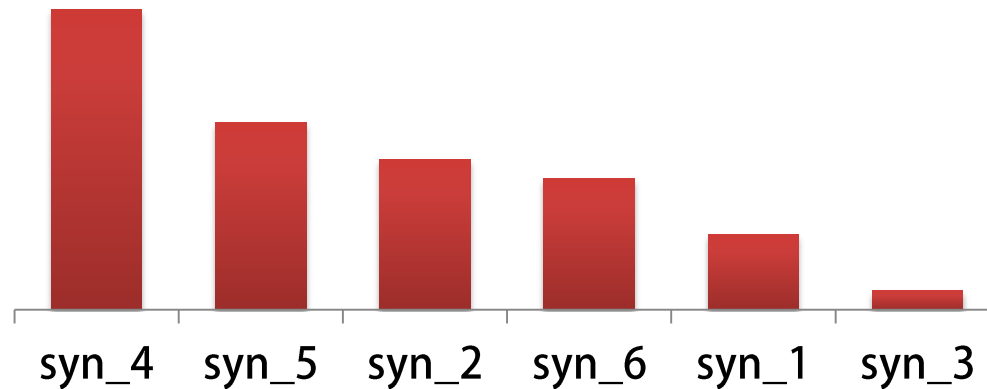
Comparing Semantic Signatures

Weighted Overlap



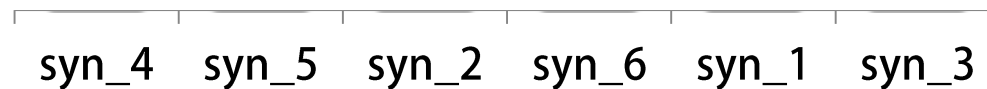
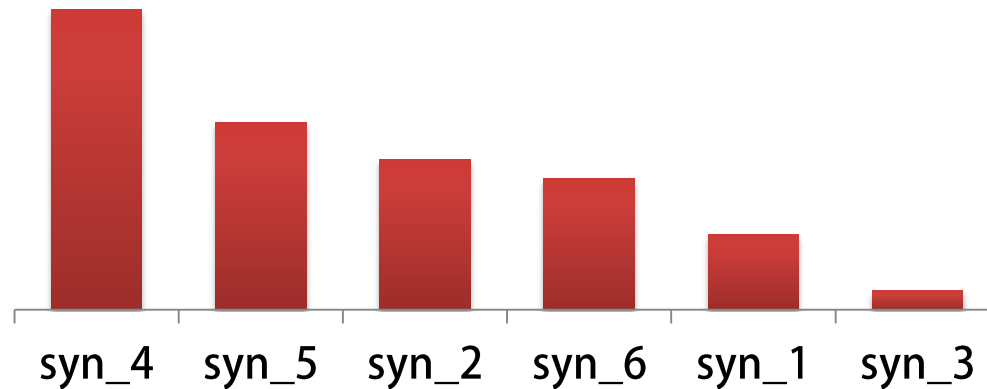
Comparing Semantic Signatures

Weighted Overlap



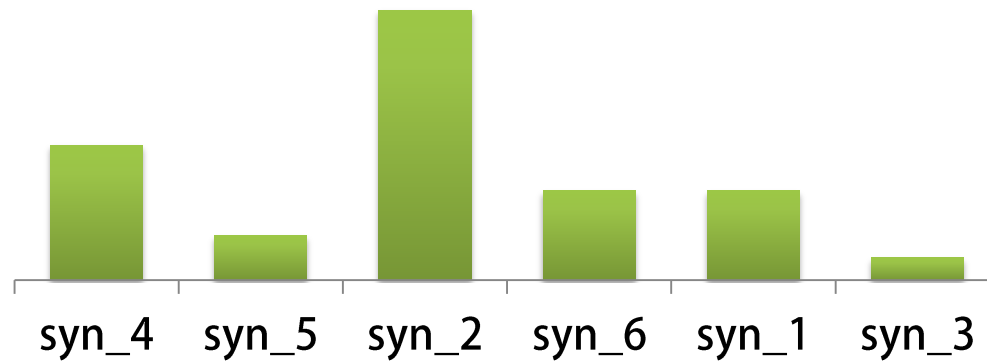
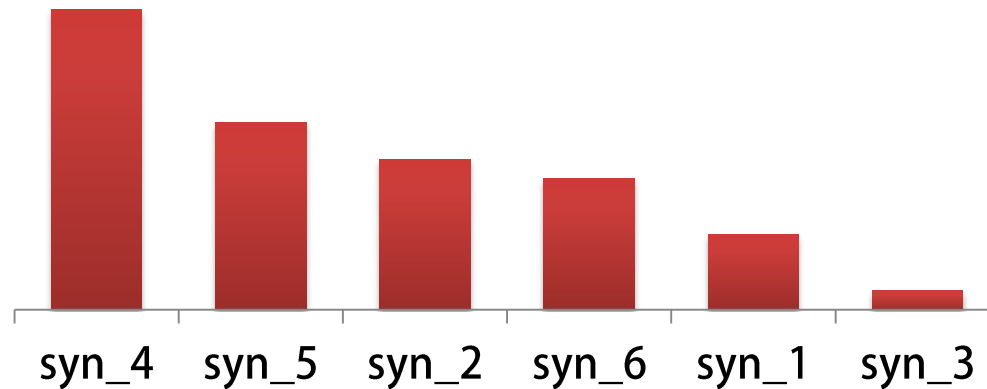
Comparing Semantic Signatures

Weighted Overlap



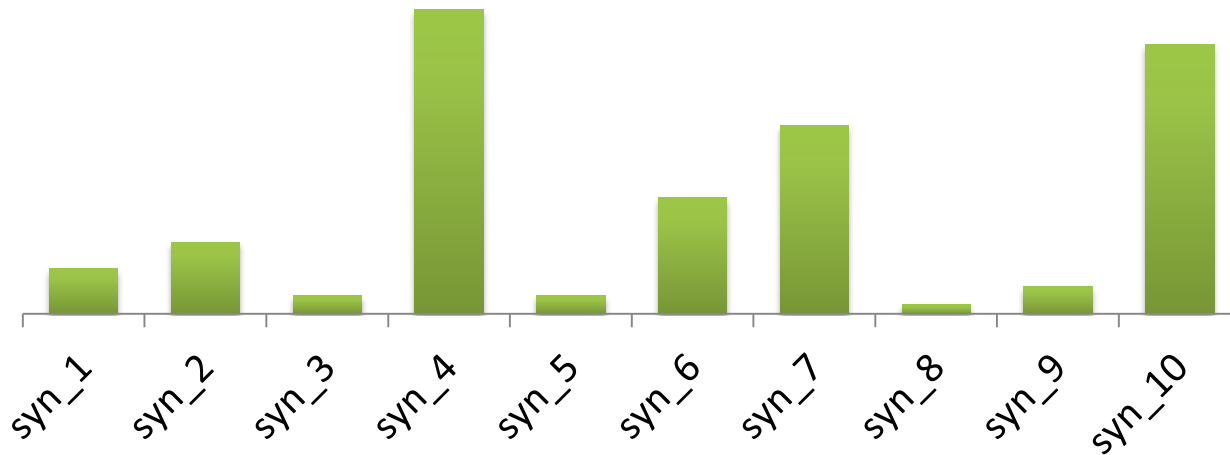
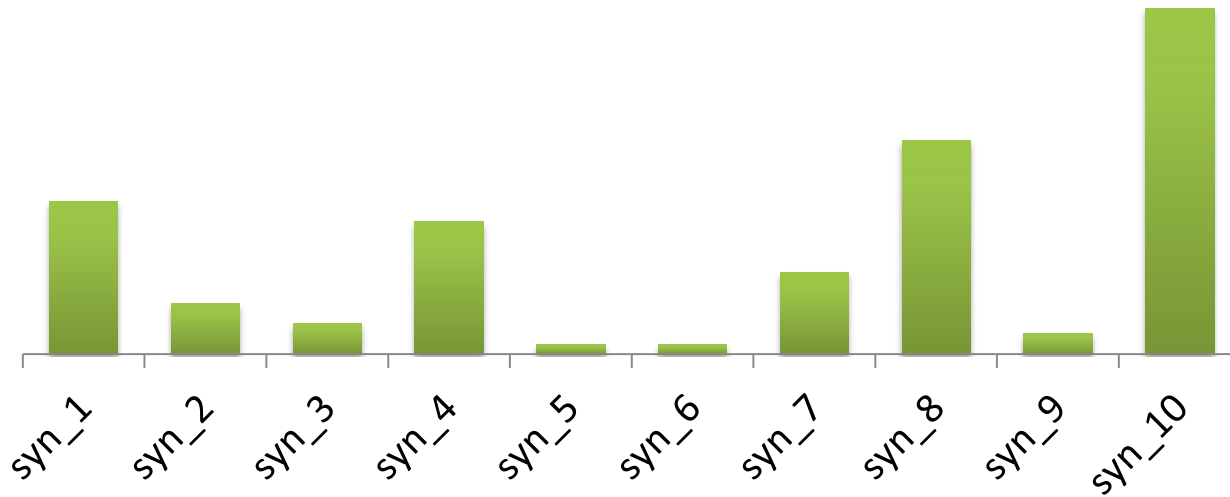
Comparing Semantic Signatures

Weighted Overlap



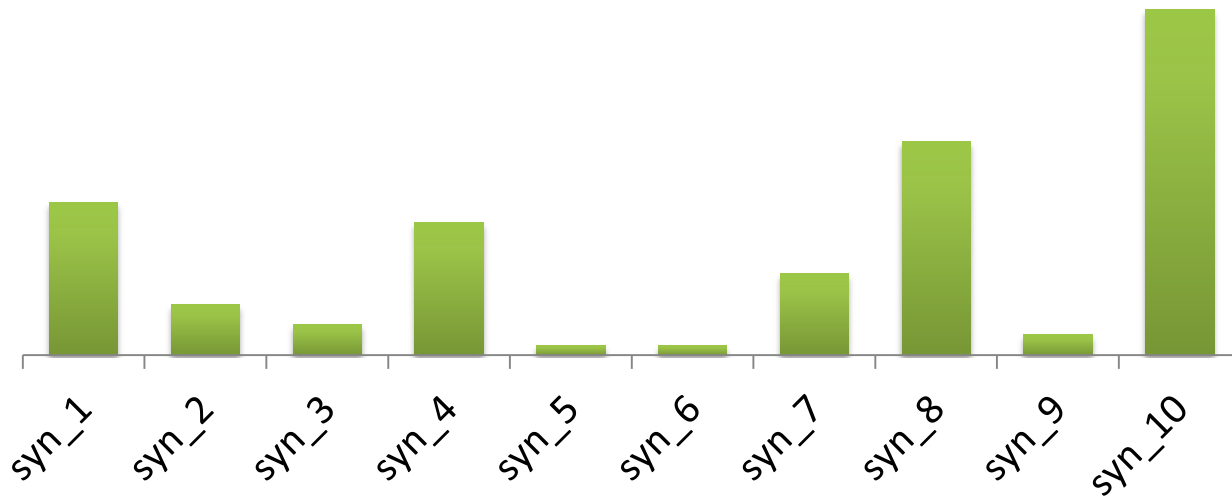
Comparing Semantic Signatures

Top- k Jaccard

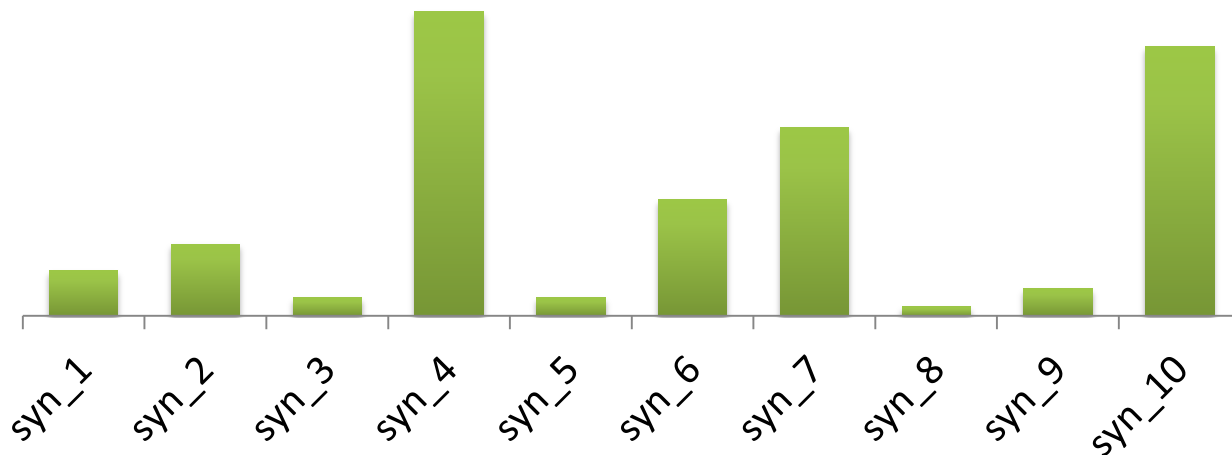


Comparing Semantic Signatures

Top- k Jaccard

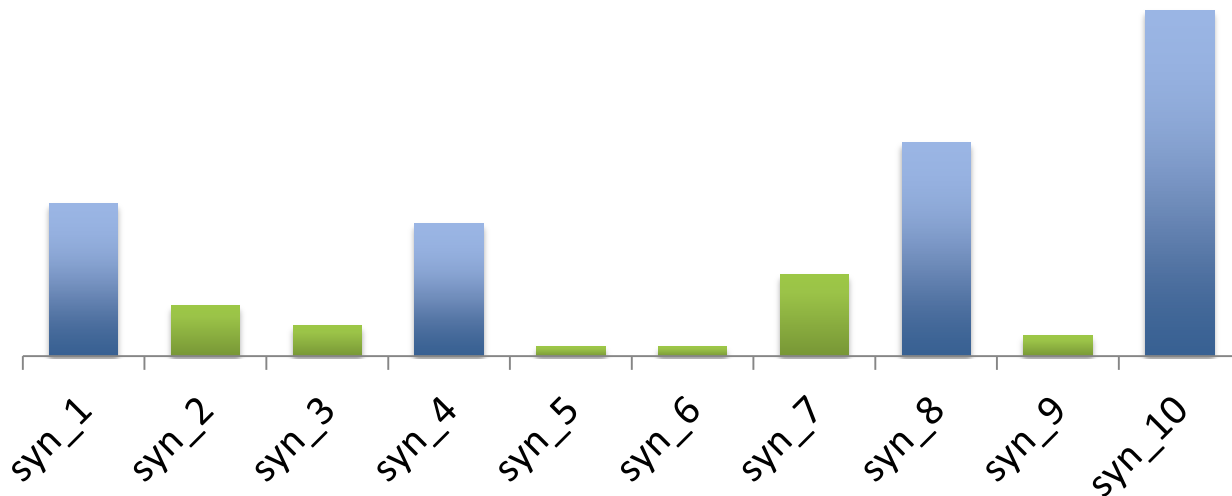


$k = 4$

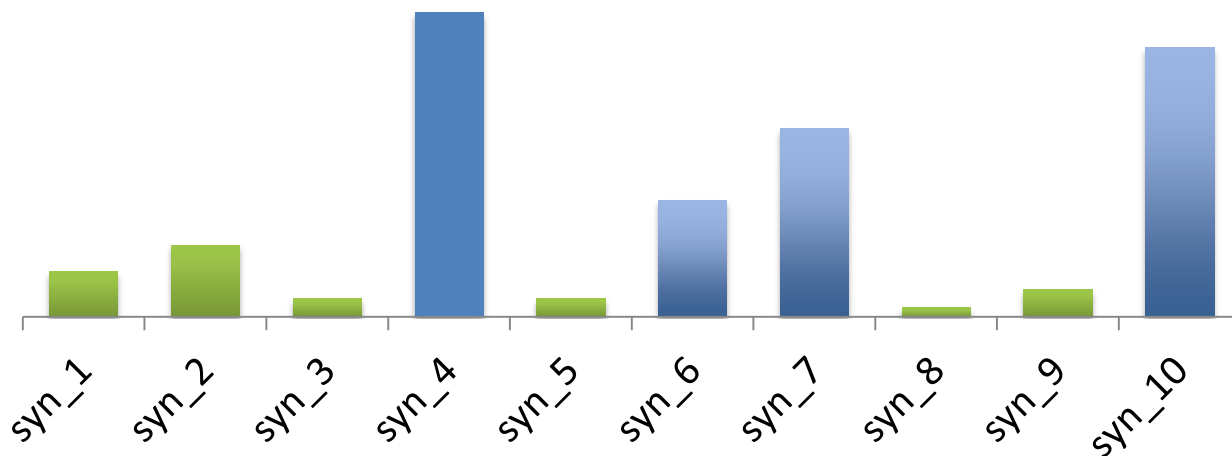


Comparing Semantic Signatures

Top- k Jaccard

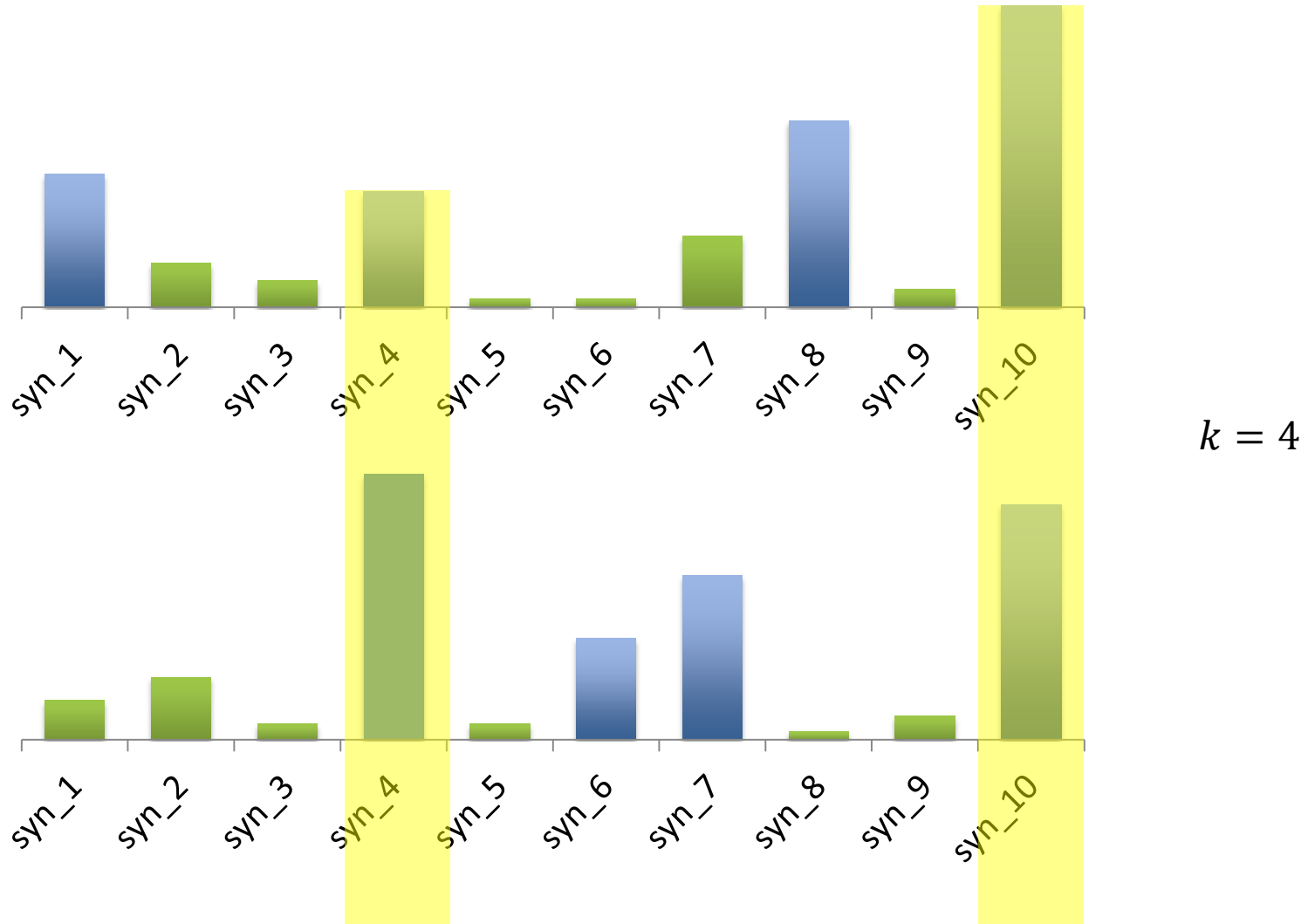


$k = 4$

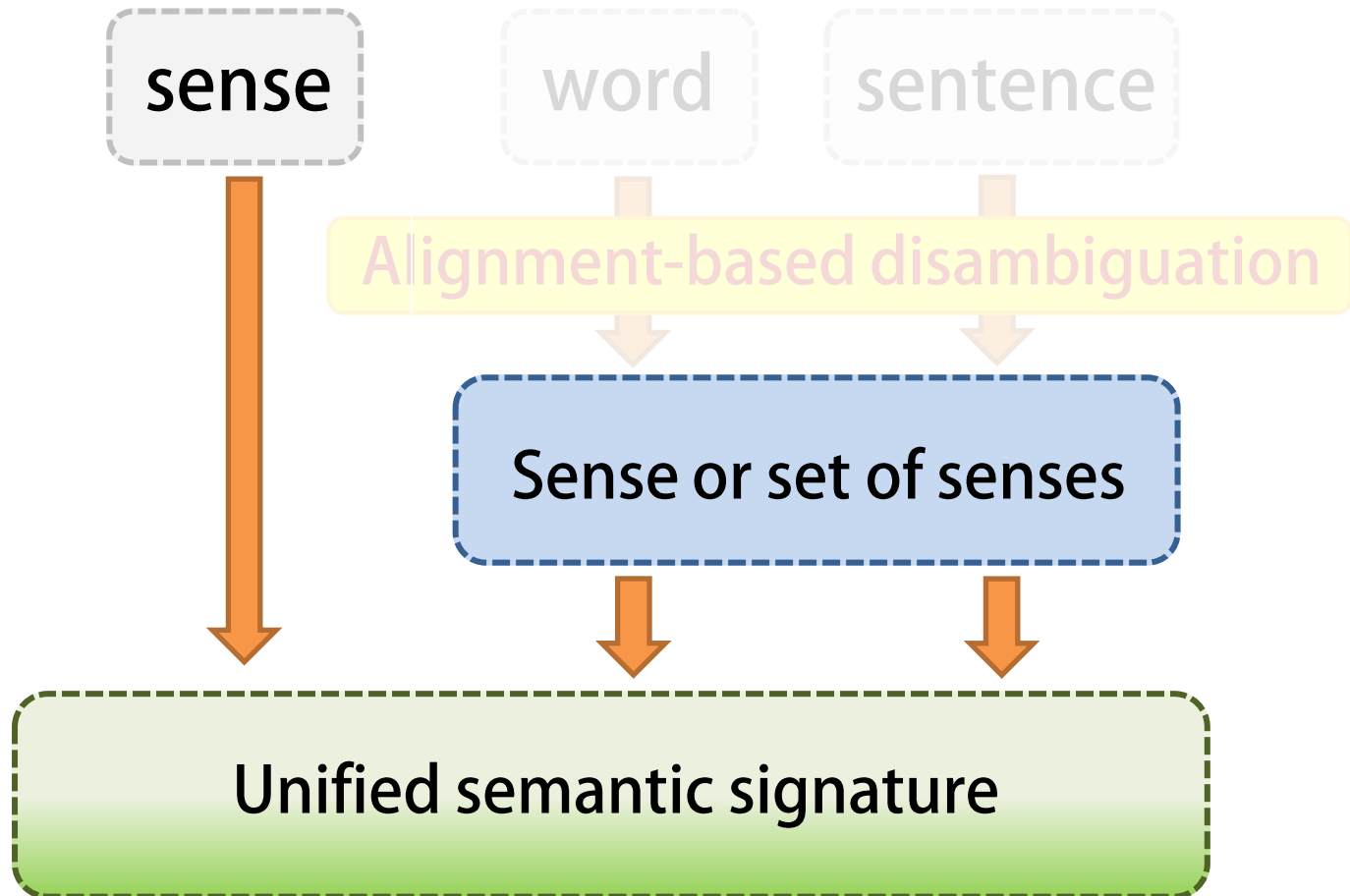


Comparing Semantic Signatures

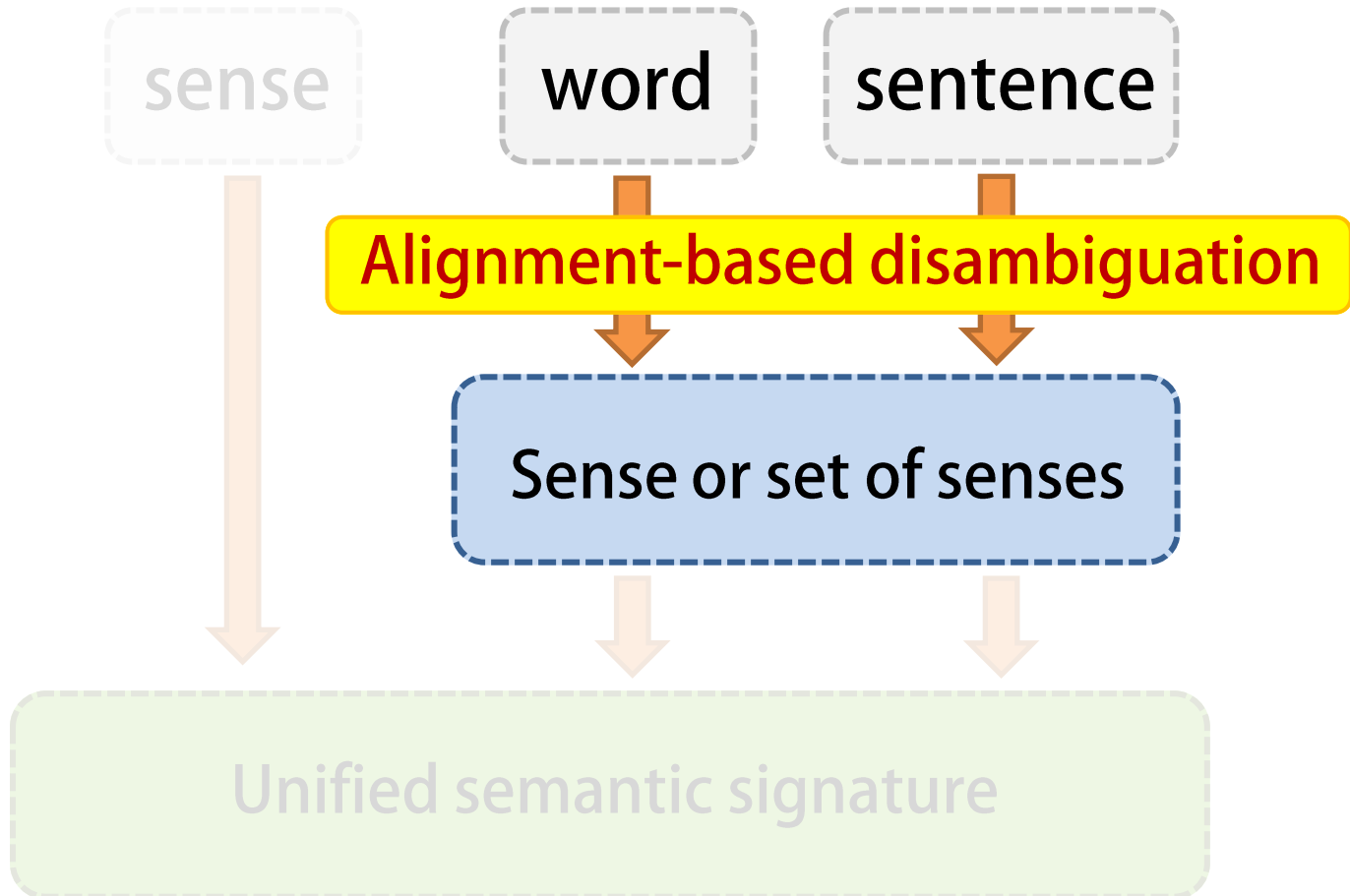
Top- k Jaccard



Alignment-based disambiguation



Alignment-based disambiguation



Why is disambiguation needed?

Why is disambiguation needed?

The worker was fired

He was terminated



Why is disambiguation needed?

The worker was fired

He was terminated



Why is disambiguation needed?

The worker was fired

He was terminated



Why is disambiguation needed?

The worker was fired

He was terminated



Alignment-based disambiguation

A manager fired the worker.

An employee was terminated from work by his boss.

Alignment-based disambiguation

A **manager** fired the **worker**.

An **employee** was **terminated** from **work** by his **boss**.

Alignment-based disambiguation

manager_n fire_v worker_n

employee_n terminate_v work_n boss_n

Alignment-based disambiguation

manager_n fire_v worker_n

employee_n terminate_v work_n boss_n

manager_n¹ fire_v¹ worker_n¹

manager_n² fire_v² worker_n²

fire_v³ ⋮

⋮

Sentence 1

employee_n¹ terminate_v¹ work_n¹ boss_n¹

terminate_v² work_n² boss_n²

terminate_v³ work_n³ ⋮

terminate_v⁴ ⋮

Sentence 2

Alignment-based disambiguation

manager_n fire_v worker_n

employee_n terminate_v work_n boss_n

manager_n¹ fire_v¹ worker_n¹

manager_n² fire_v² worker_n²

fire_v³ ⋮

⋮

Sentence 1

employee_n¹ terminate_v¹ work_n¹ boss_n¹

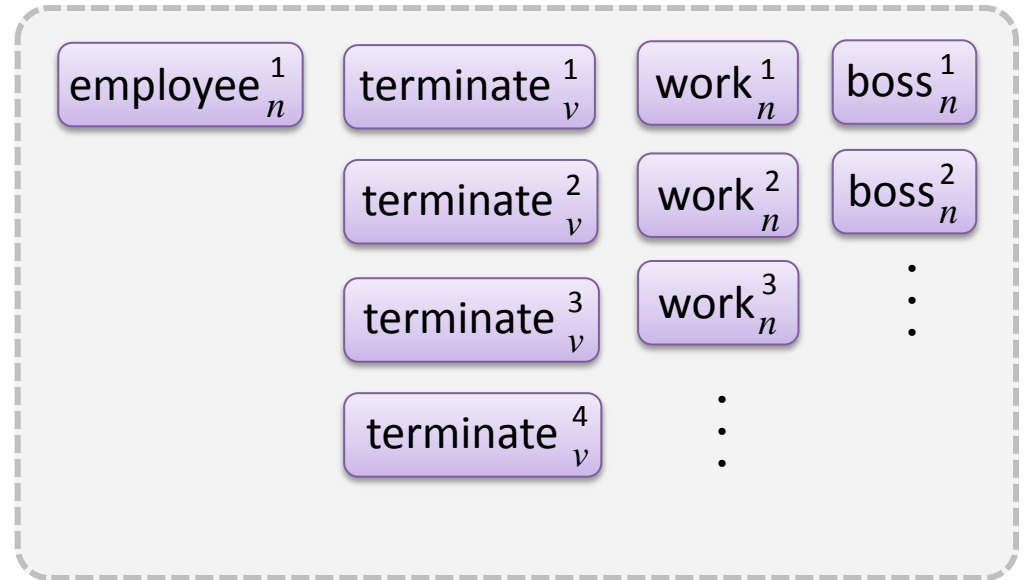
terminate_v² work_n² boss_n²

terminate_v³ work_n³ ⋮

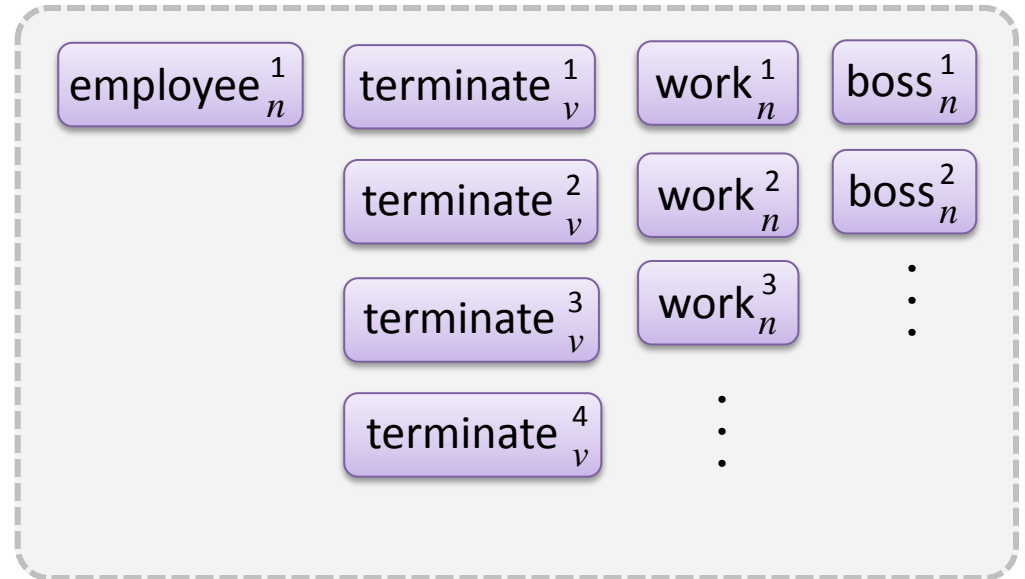
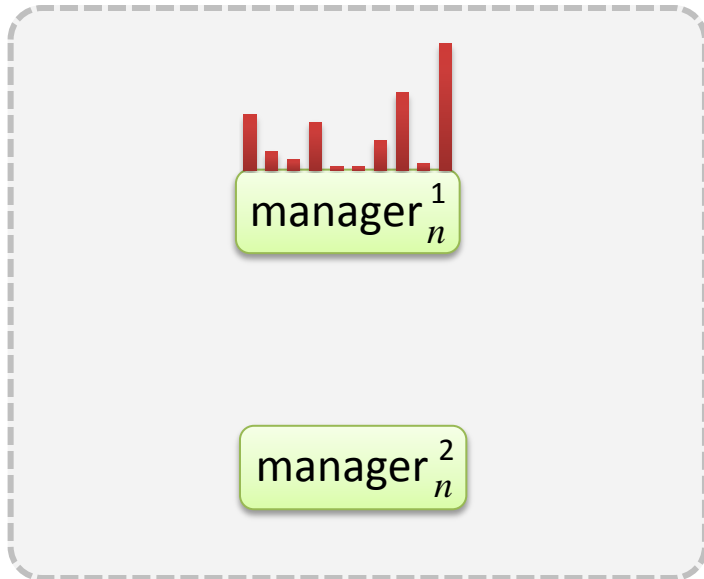
terminate_v⁴ ⋮

Sentence 2

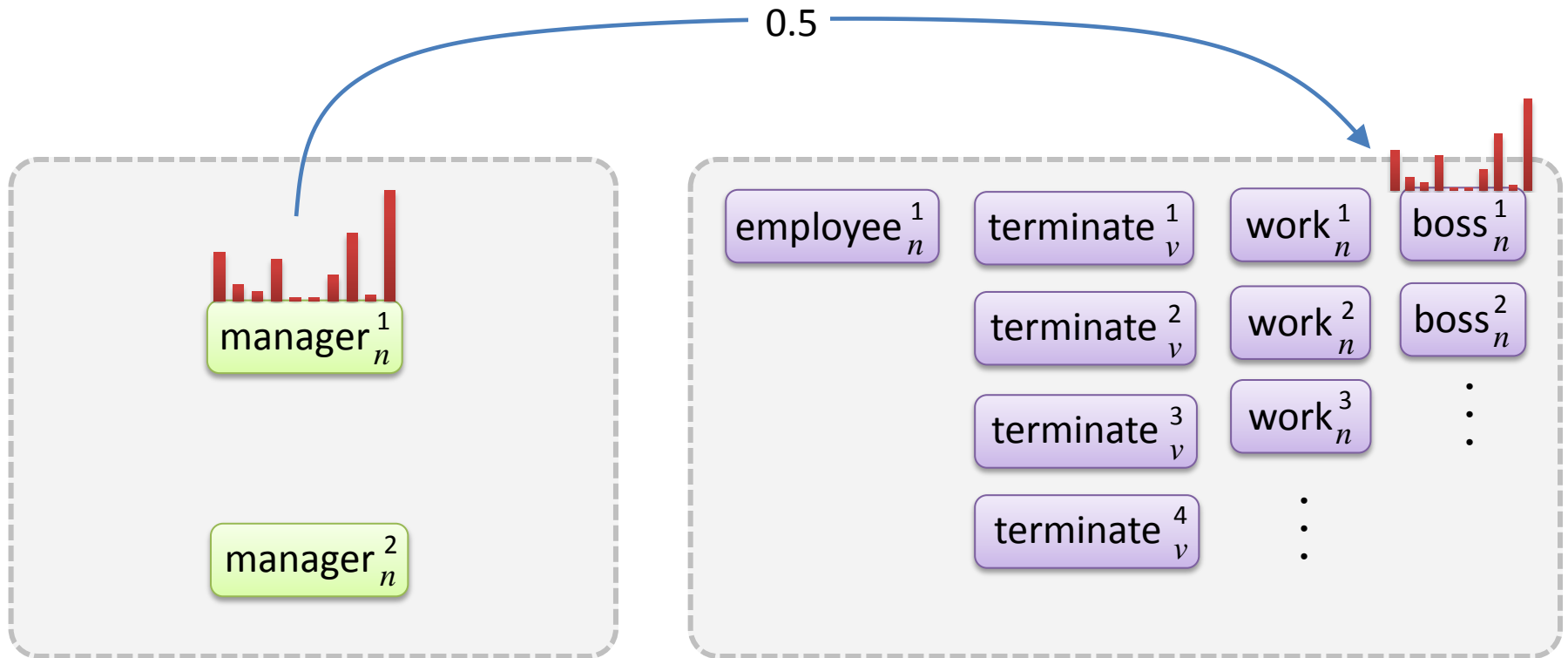
Alignment-based disambiguation



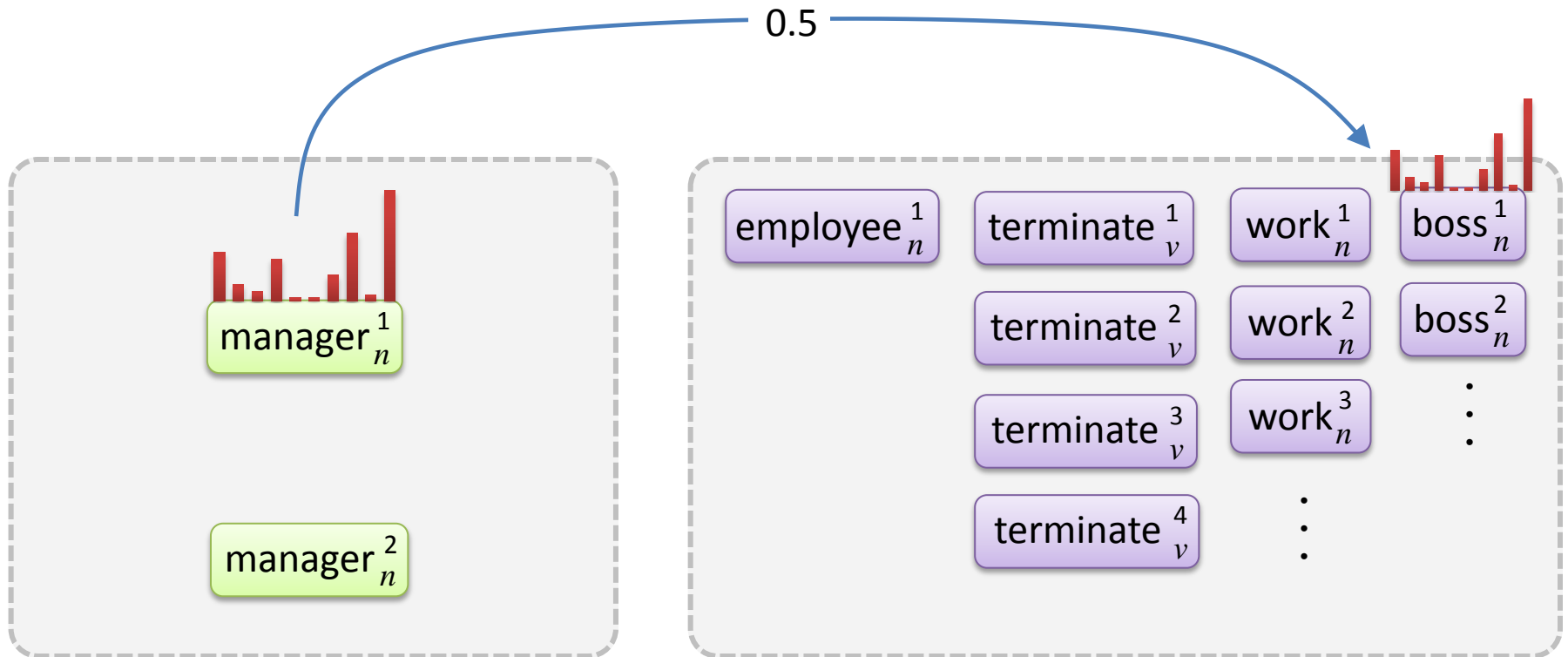
Alignment-based disambiguation



Alignment-based disambiguation

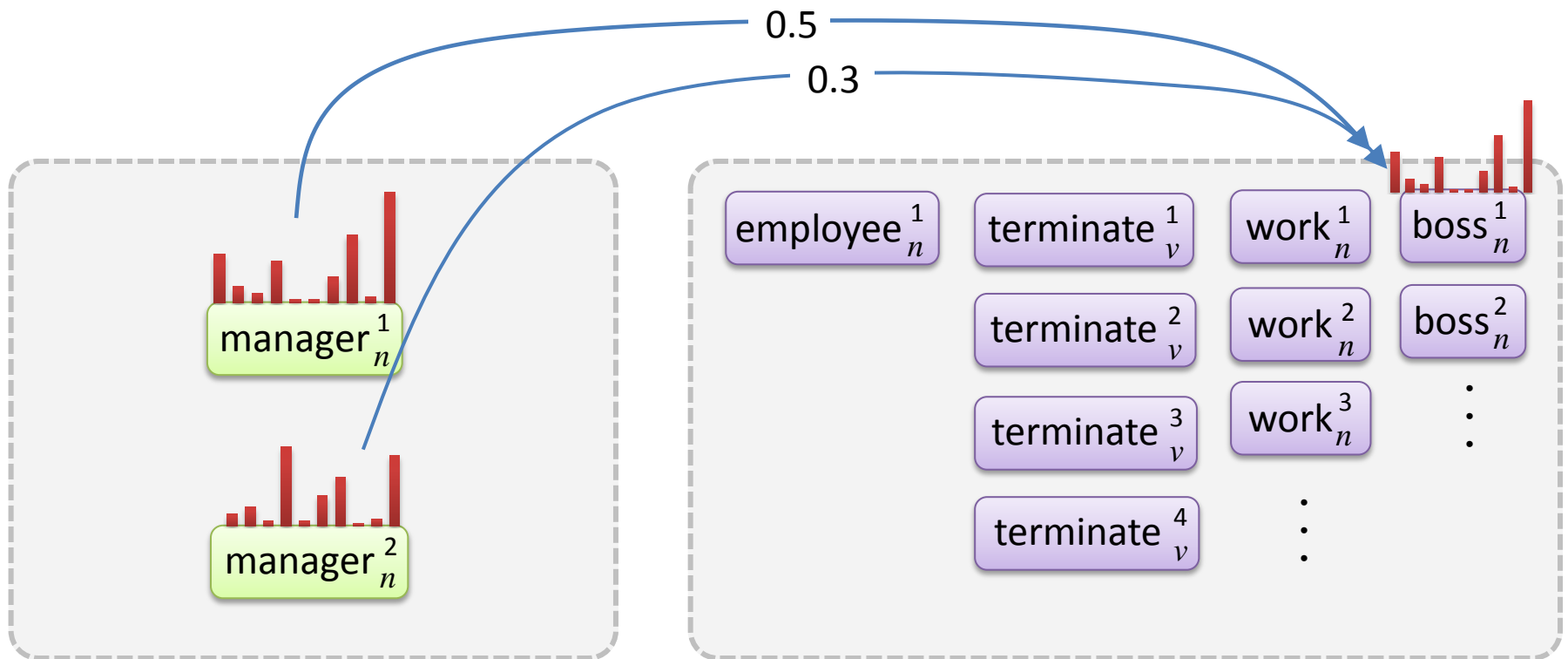


Alignment-based disambiguation



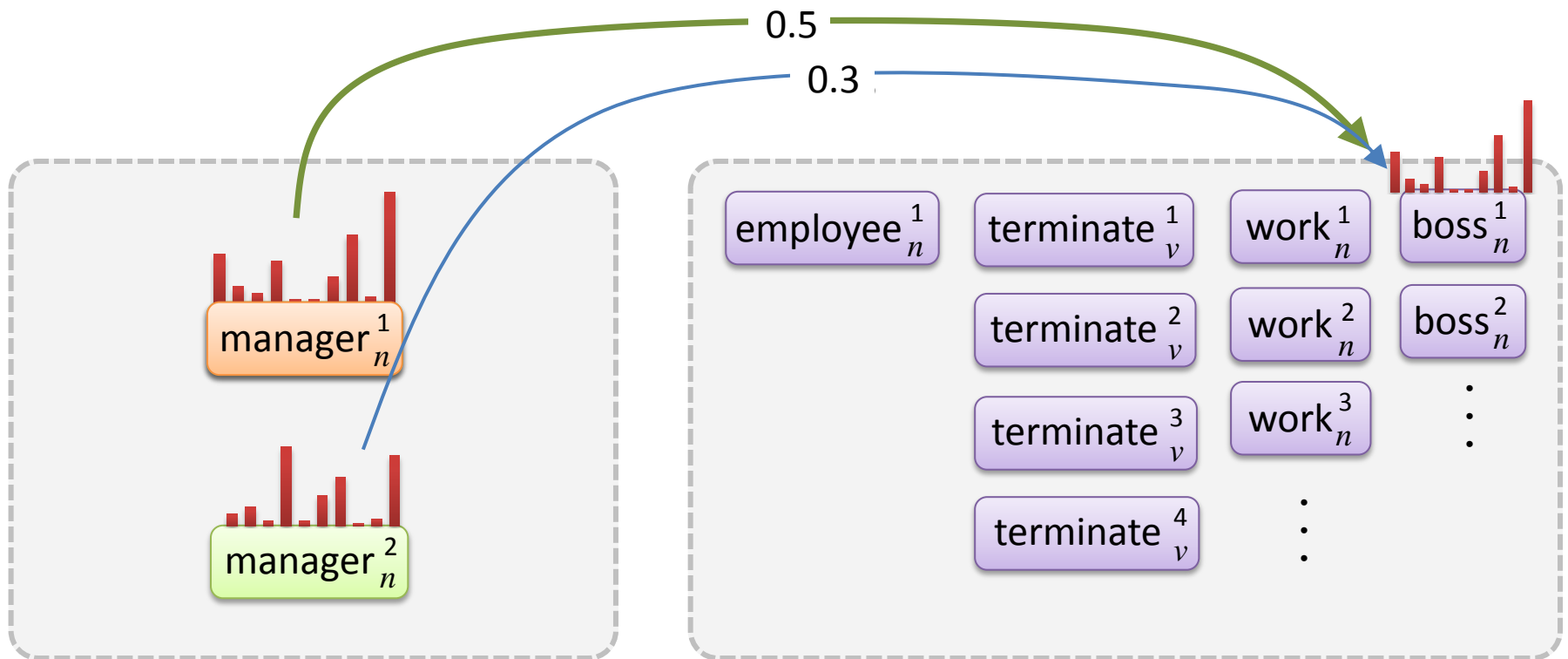
Tversky (1977)
Markman and Gentner (1993)

Alignment-based disambiguation



Tversky (1977)
Markman and Gentner (1993)

Alignment-based disambiguation



Alignment-based disambiguation

manager_n fire_v worker_n

employee_n terminate_v work_n boss_n

manager¹_n fire¹_v worker¹_n

manager²_n fire²_v worker²_n

fire³_v ⋮

fire⁴_v

Sentence 1

⋮

employee¹_n terminate¹_v work¹_n boss¹_n

terminate²_v work²_n boss²_n

terminate³_v work³_n ⋮

terminate⁴_v ⋮

Sentence 2

Alignment-based disambiguation

manager_n fire_v worker_n

employee_n terminate_v work_n boss_n

manager¹_n fire¹_v worker¹_n

manager²_n fire²_v worker²_n

fire³_v ⋮

fire⁴_v

Sentence 1

⋮

employee¹_n terminate¹_v work¹_n boss¹_n

terminate²_v work²_n boss²_n

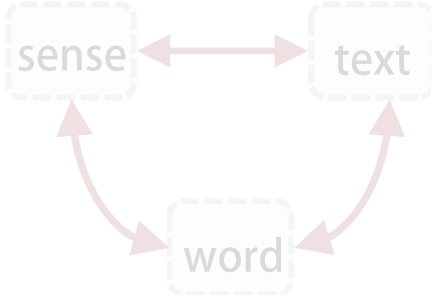
terminate³_v work³_n ⋮

terminate⁴_v ⋮

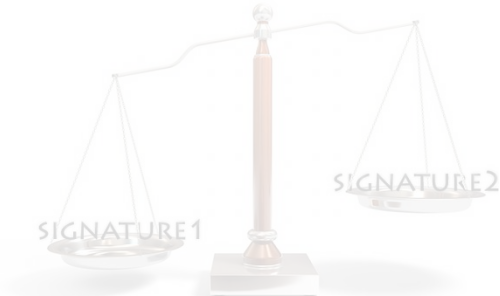
Sentence 2

Outline

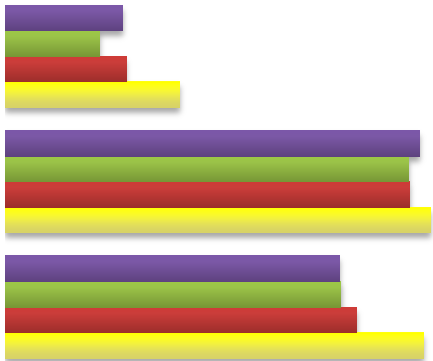
Introduction



Methodology



Experiments



Experiments

- Sentence level
 - Semantic Textual Similarity (SemEval-2012)

Experiments

- Sentence level
 - Semantic Textual Similarity (SemEval-2012)
- Word level
 - Synonymy recognition (TOEFL dataset)
 - Correlation-based (RG-65 dataset)

Experiments

- **Sentence level**
 - Semantic Textual Similarity (SemEval-2012)
- **Word level**
 - Synonymy recognition (TOEFL dataset)
 - Correlation-based (RG-65 dataset)
- **Sense level**
 - Coarsening WordNet sense inventory

Experiment I

Similarity at Sentence level

- **Semantic Textual Similarity (STS-12)**
 - 5 datasets
 - Three evaluation measures
 - ALL, ALLnrm, and Mean

Experiment I

Similarity at Sentence level

- **Semantic Textual Similarity (STS-12)**
 - 5 datasets
 - Three evaluation measures
 - ALL, ALLnrm, and Mean
- **Top-ranking systems**
 - UKP2 (Bär et al., 2012)
 - TLSim and TLSyn (Šarić et al., 2012)

Experiment I

Similarity at Sentence level

Features

- Main features
 - Cosine
 - Weighted Overlap
 - Top-k Jaccard

Experiment I

Similarity at Sentence level

Features

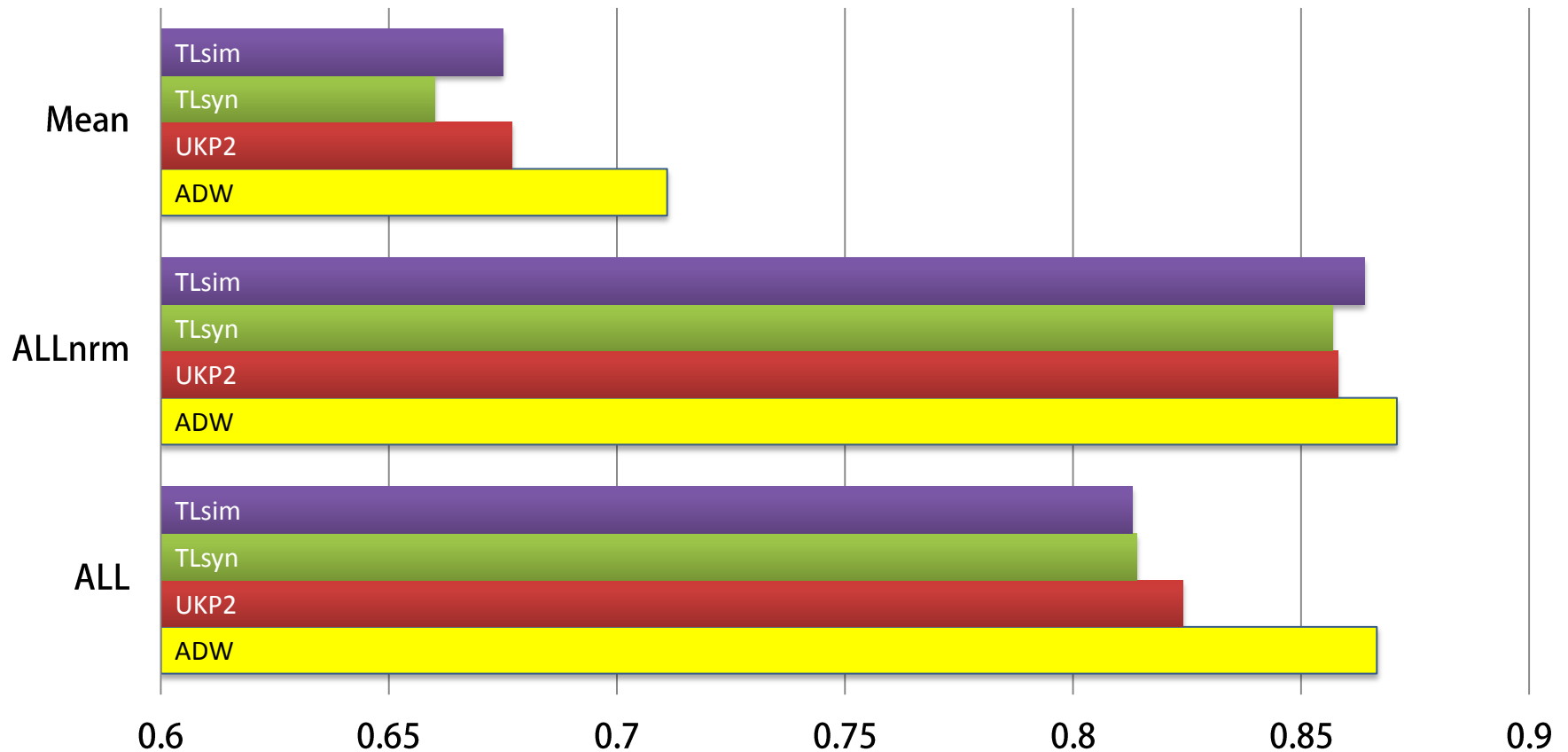
- Main features
 - Cosine
 - Weighted Overlap
 - Top-k Jaccard

- String-based features
 - Longest common substring
 - Longest common subsequence
 - Greedy string tiling
 - Character/word n-grams

Experiment I

Similarity at Sentence level

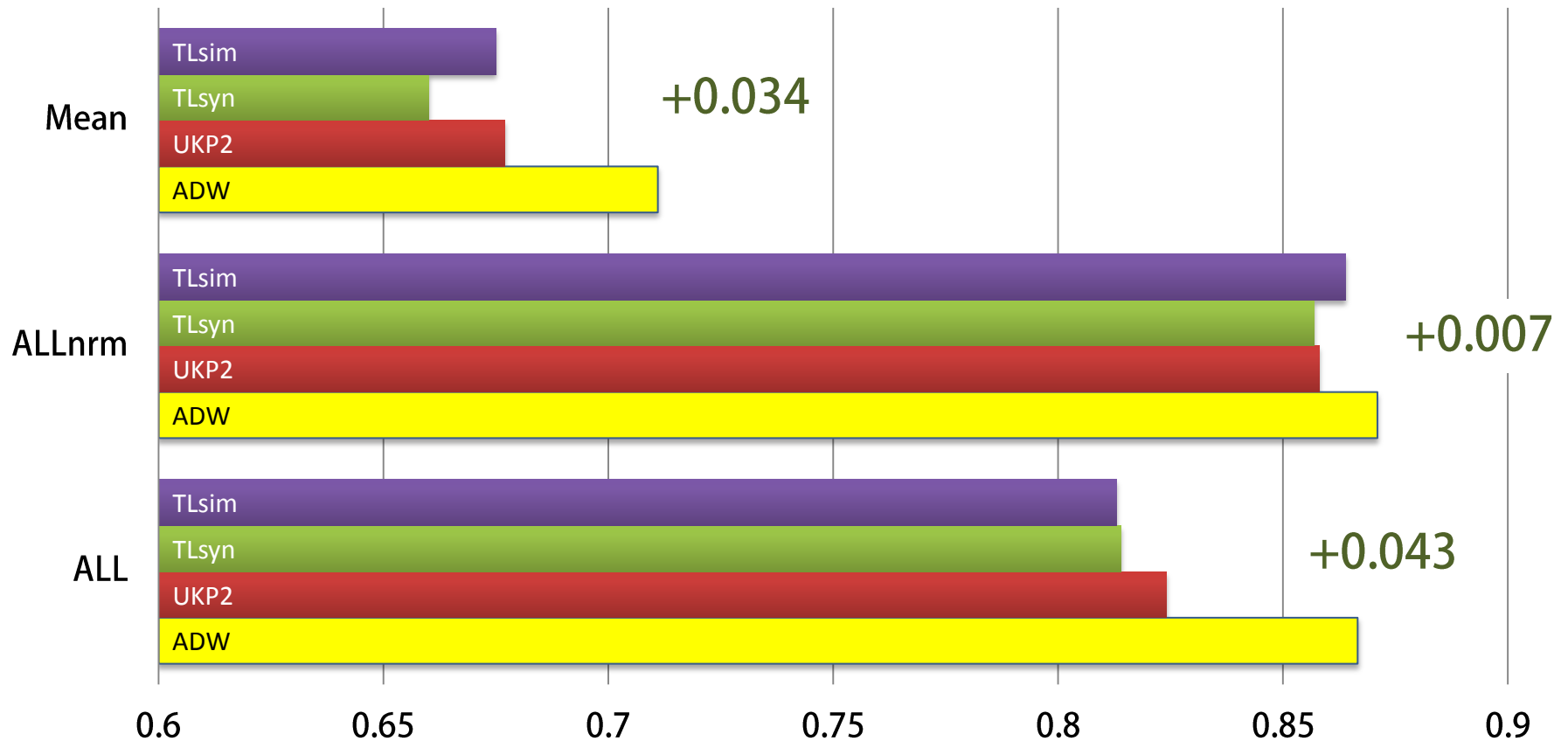
STS Results



Experiment I

Similarity at Sentence level

STS Results



Experiments

- Sentence level
 - Semantic Textual Similarity (Semeval-12)
- **Word level**
 - Synonymy recognition (TOEFL dataset)
 - Correlation-based (RG-65 dataset)
- Sense level
 - Coarsening WordNet sense inventory

Experiment 2

Similarity at Word Level

Experiment 2

Similarity at Word Level

Synonymy recognition



Experiment 2

Similarity at Word Level

Synonymy recognition

enormous?

- appropriate
- unique
- tremendous
- decided

Correlating word similarity judgments

autograph	shore	0.06
coast	forest	0.85
midday	noon	3.94

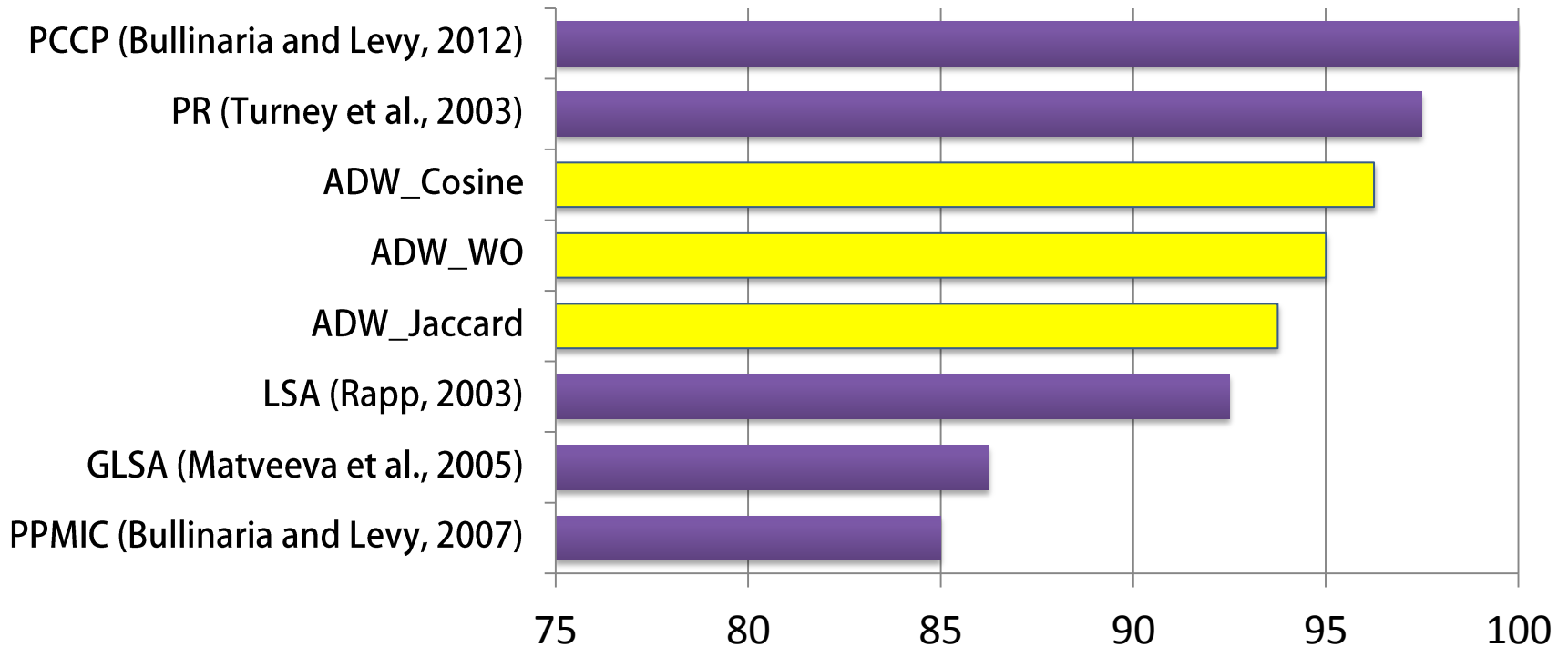
Similarity at Word Level

Synonym Recognition

- TOEFL dataset (Landauer and Dumais, 1997)
 - 80 multiple choice questions
 - Human test takers: 64.5% only

Similarity at Word Level Synonym Recognition

Accuracy on TOEFL dataset



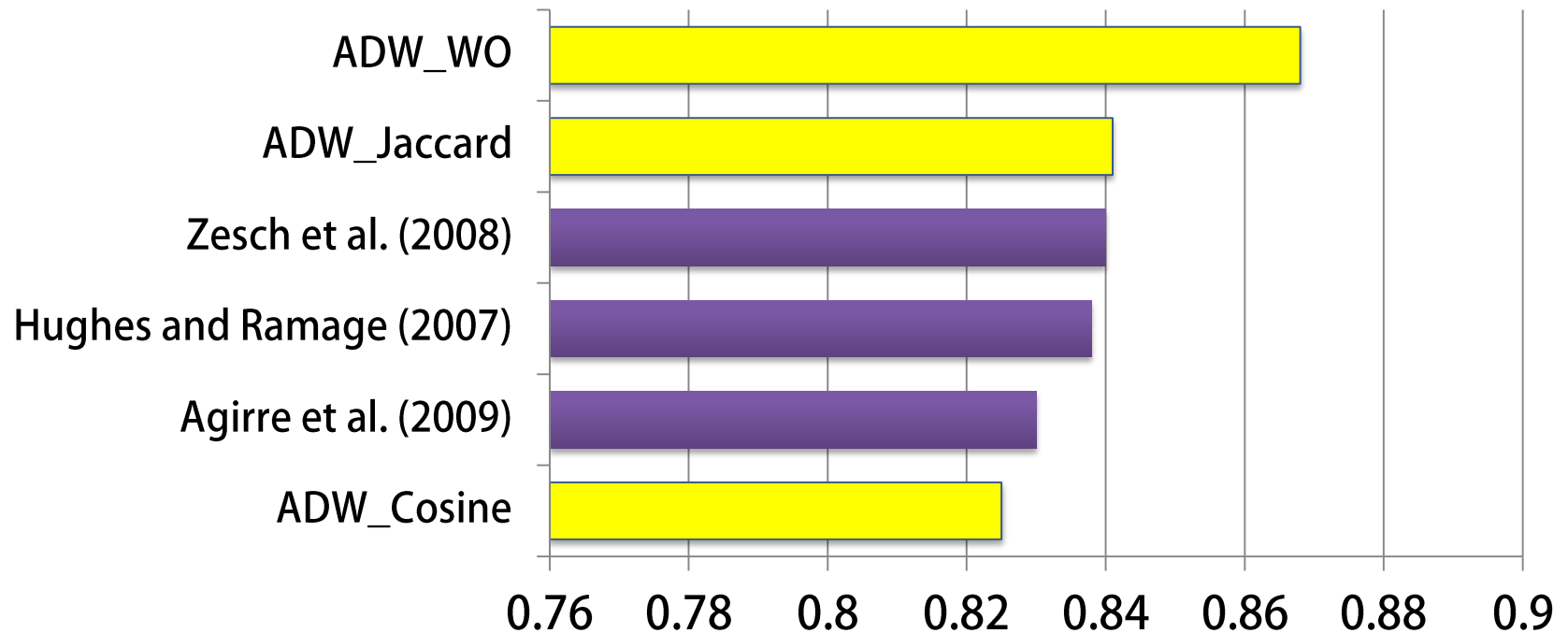
Similarity at Word Level

Judgment Correlation

- Dataset: RG-65 (Rubenstein and Goodenough, 1965)
 - 65 word pairs
 - judged by 51 human subjects
 - Scale of 0 → 4

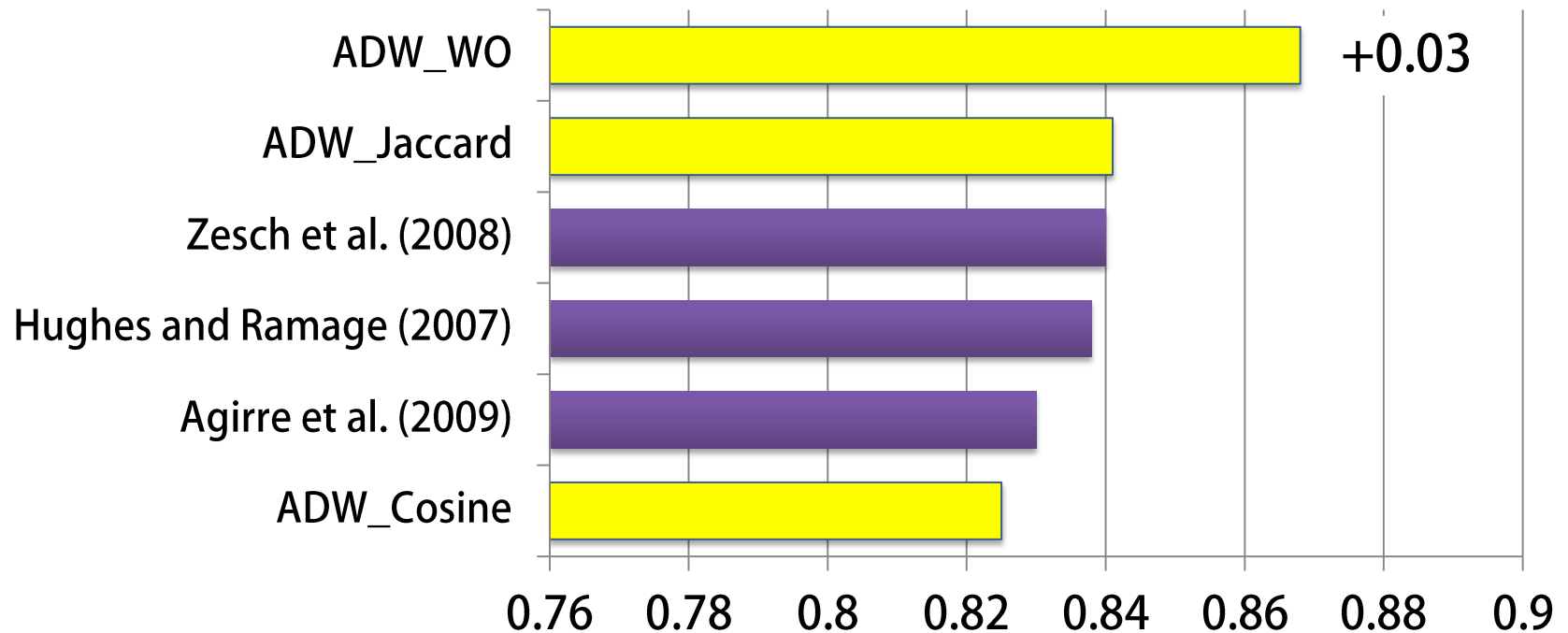
Similarity at Word Level Judgment Correlation

Spearman correlation, RG-65 dataset



Similarity at Word Level Judgment Correlation

Spearman correlation, RG-65 dataset



Experiments

- Text level
 - Semantic Textual Similarity (Semeval-12)
- Word level
 - Synonymy recognition (TOEFL dataset)
 - Correlation-based (RG-65 dataset)
- Sense level
 - Coarsening WordNet sense inventory

Experiment 3

Similarity at Sense Level

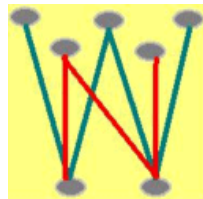
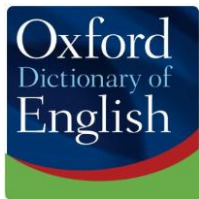
- Coarse-graining WordNet

Experiment 3

Similarity at Sense Level

- Coarse-graining WordNet

Navigli (2006)

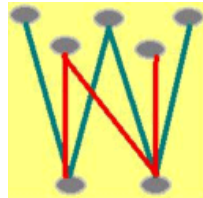
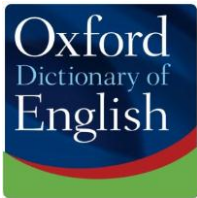


Experiment 3

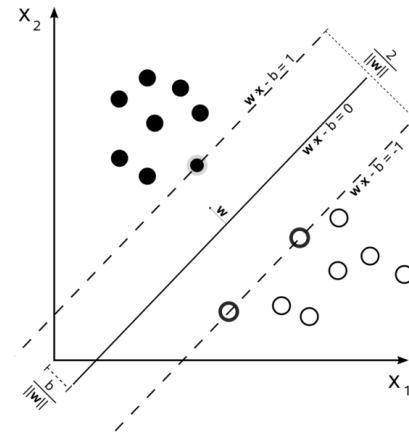
Similarity at Sense Level

- Coarse-graining WordNet

Navigli (2006)



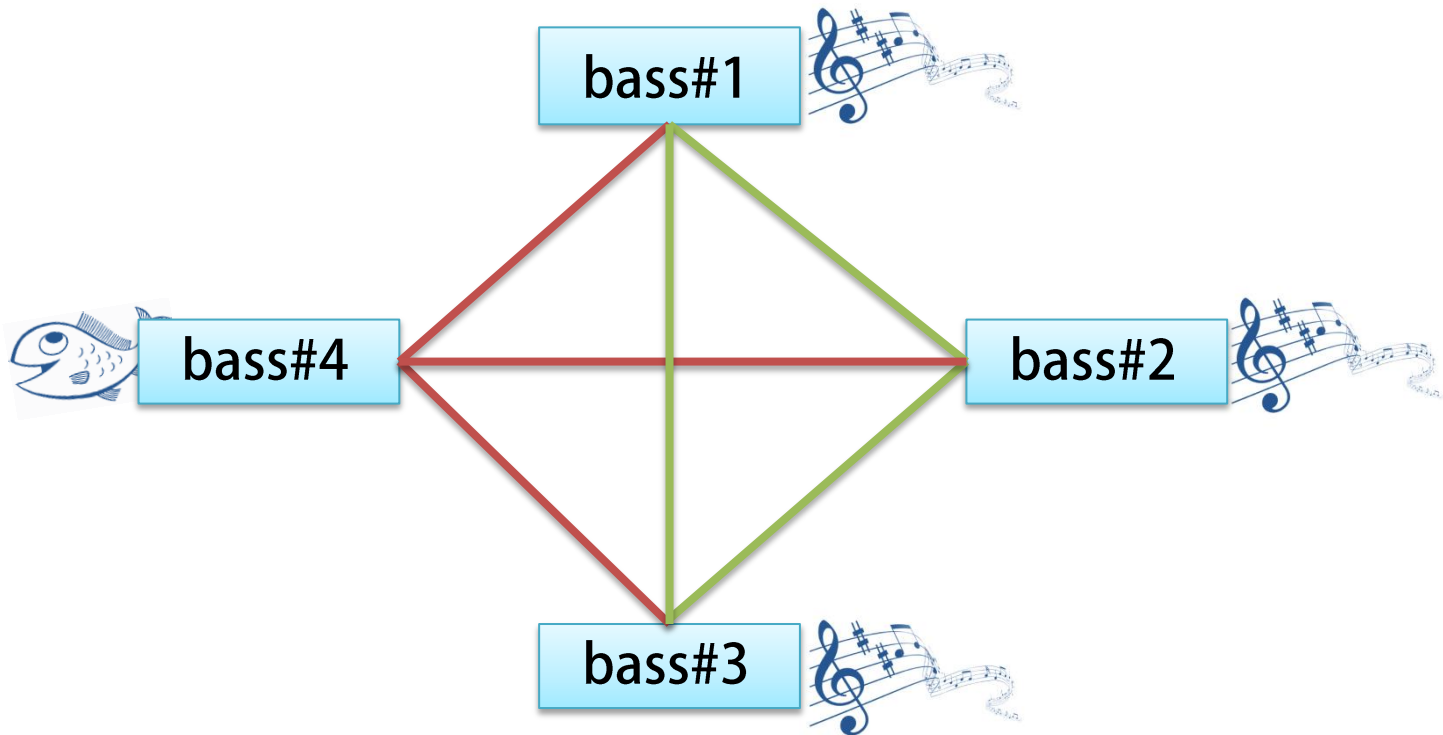
Snow et al. (2007)



Experiment 3

Similarity at Sense Level

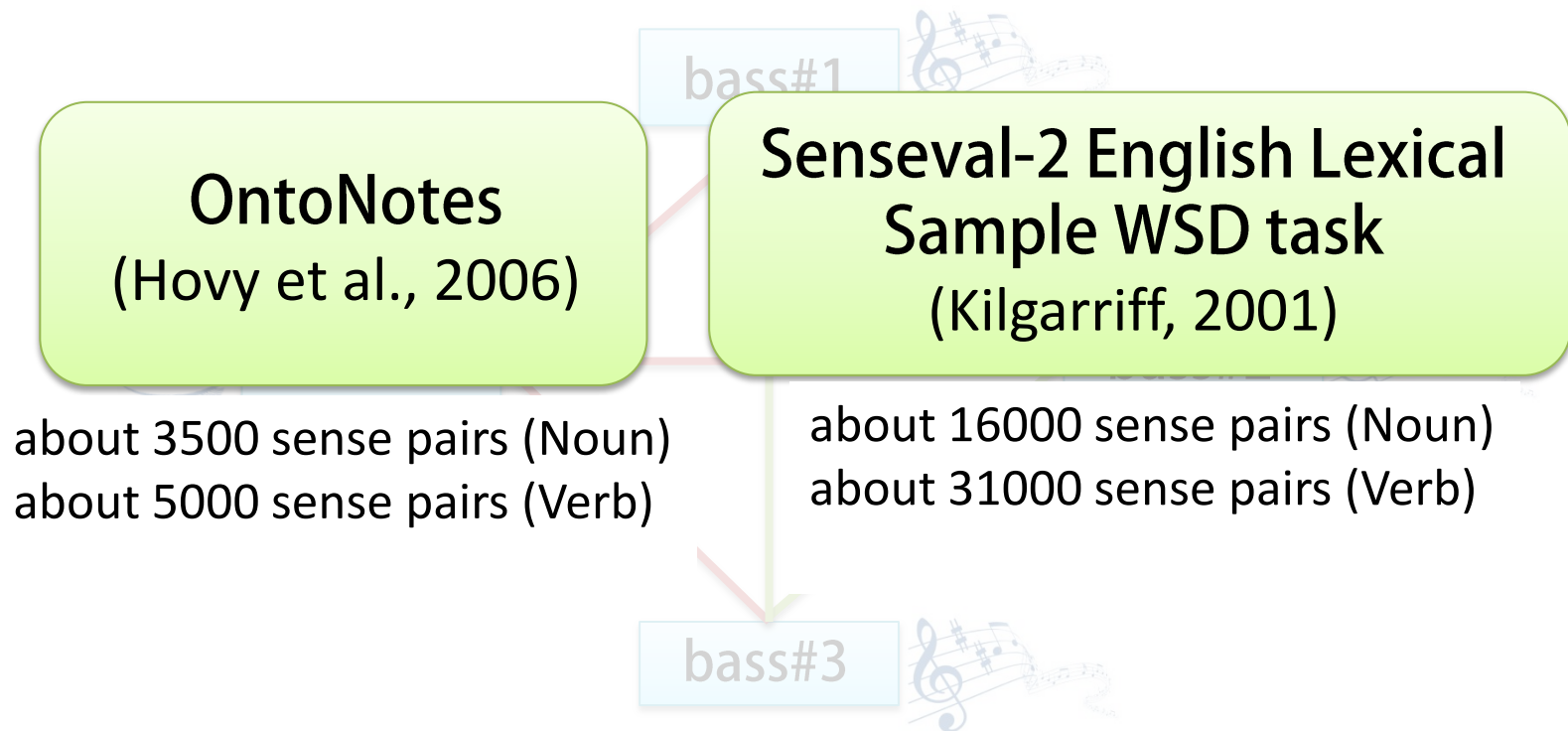
- Binary classification: Merged or not-merged



Experiment 3

Similarity at Sense Level

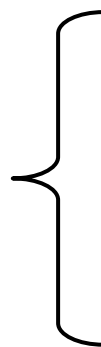
- Binary classification: Merged or not-merged



Experiment 3

Similarity at Sense Level

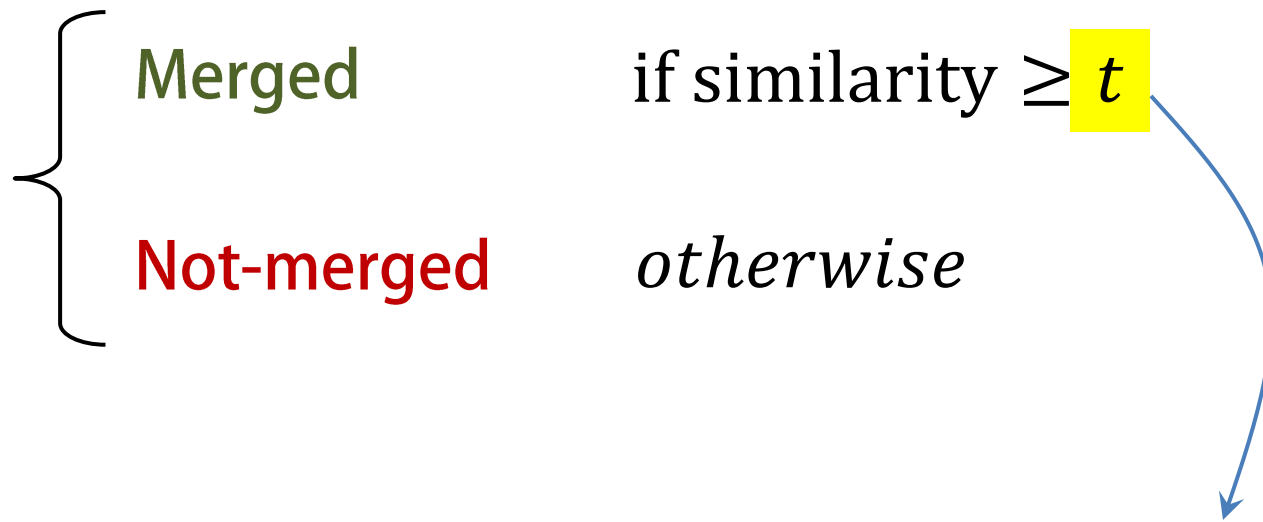
- Binary classification: Merge or not-merged

 **Merged** if similarity $\geq t$
Not-merged *otherwise*

Experiment 3

Similarity at Sense Level

- Binary classification: Merge or not-merged

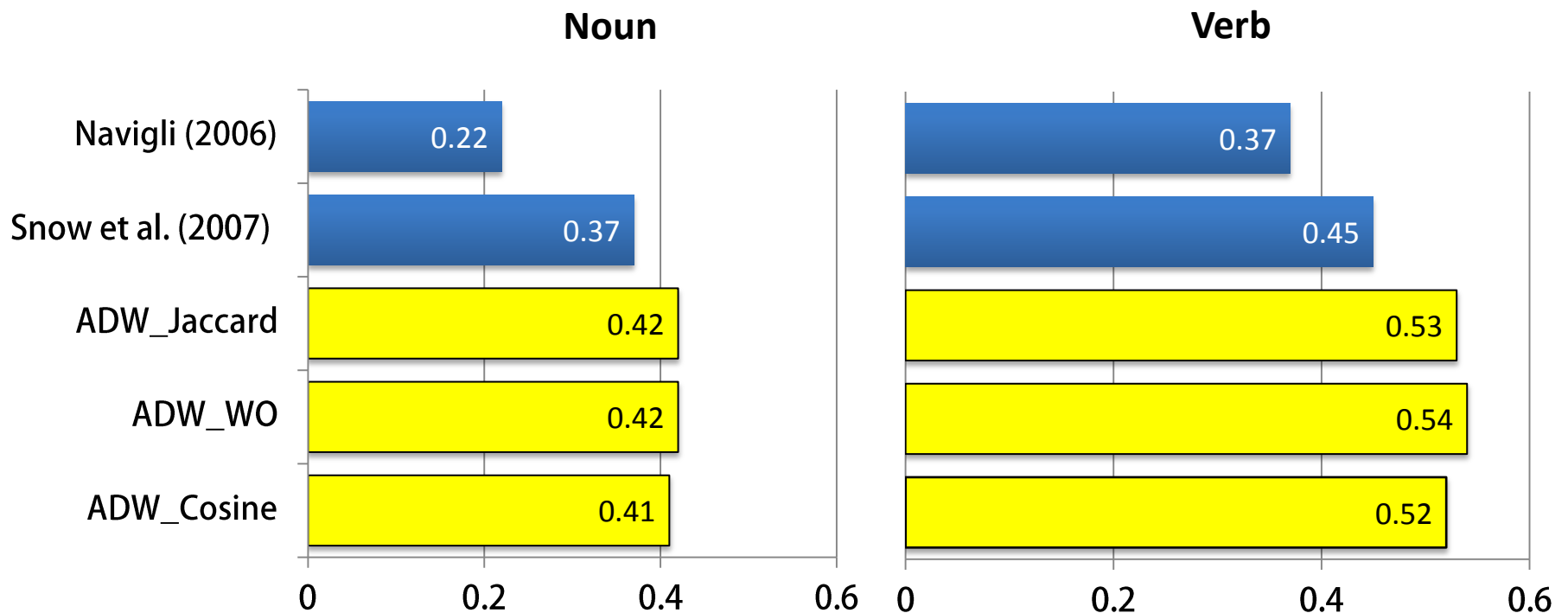


Tuned using 10% of the dataset

Experiment 3

Similarity at Sense Level

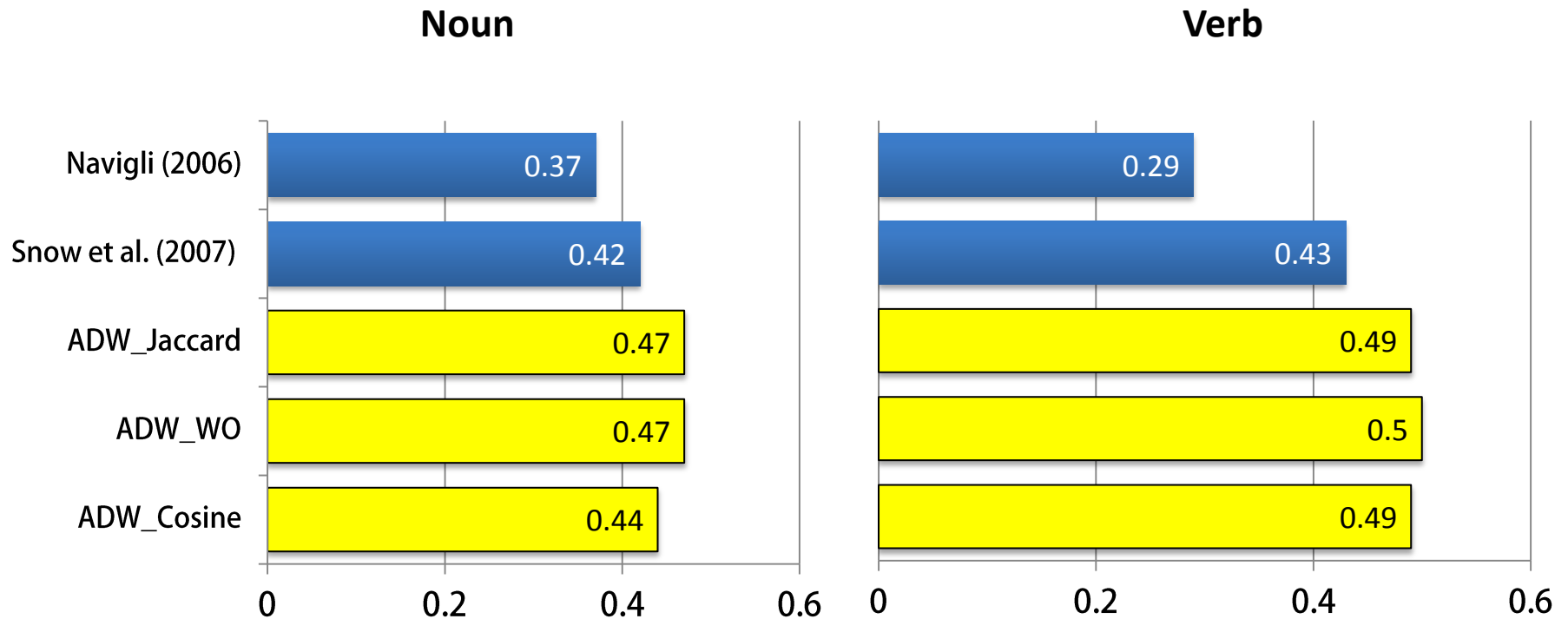
F-score on OntoNotes dataset



Experiment 3

Similarity at Sense Level

F-score on Senseval-2 dataset



Conclusions

- A unified approach for computing semantic similarity for any pair of lexical items
- Experiments with SOA performance
 - Sense level (sense coarsening)
 - Word level (synonymy recognition and judgment)
 - Sentence level (Semantic Textual Similarity)

Future Direction

- Larger sense inventories (e.g., BabelNet)

Future Direction

- Larger sense inventories (e.g., BabelNet)
- Cross-level semantic similarity

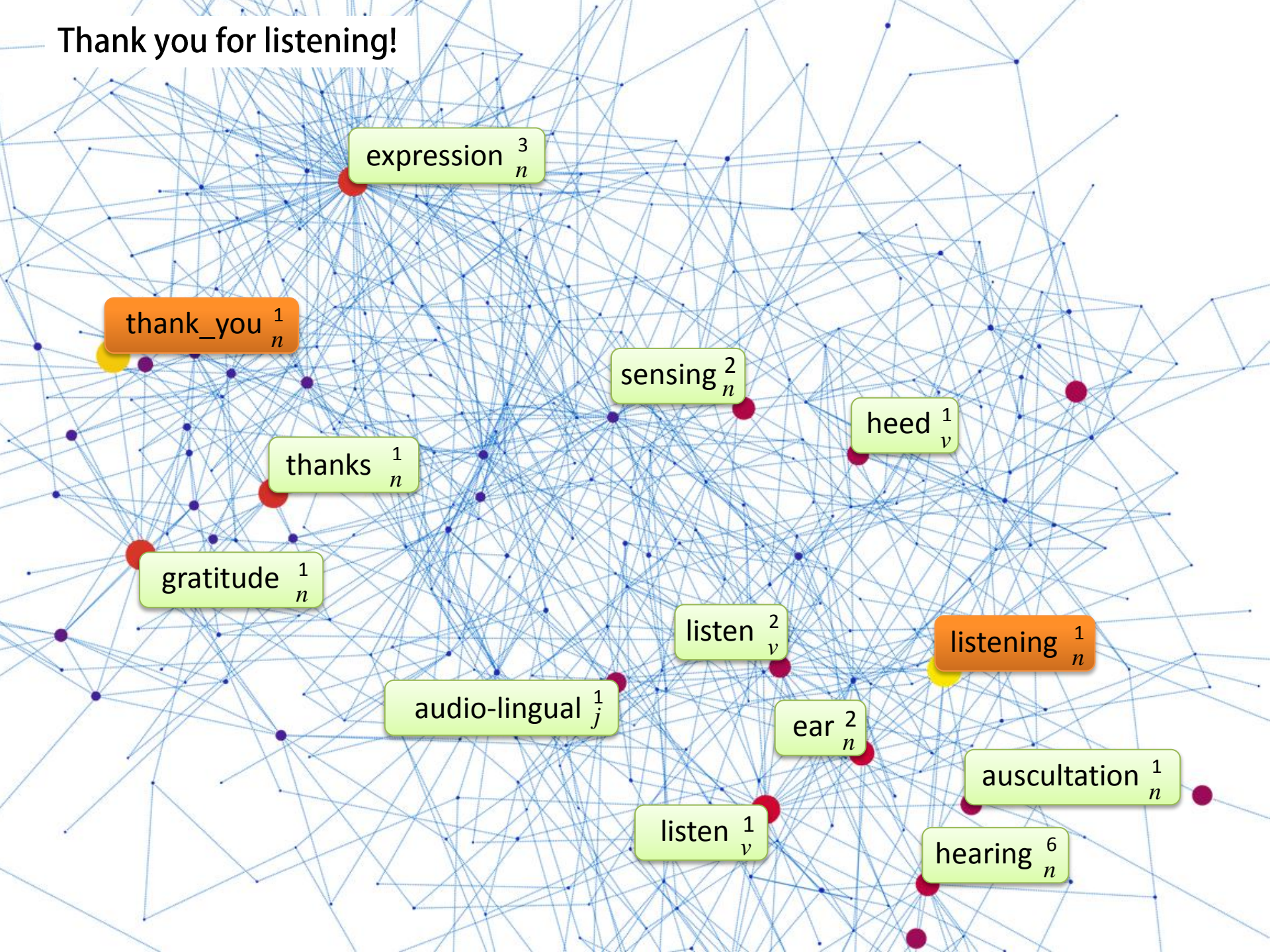
Future Direction

- Larger sense inventories (e.g., BabelNet)
- Cross-level semantic similarity

Create datasets for cross-level similarity

– Future Semeval task?

Thank you for listening!



fanciful	familiar	apparent	imaginative	logical
verbal	oral	overt	fitting	verbose
resolved	settled	forgotten	publicized	examined
percentage	volume	sample	proportion	profit
figure	list	solve	divide	express
highlight	alter	imitate	accentuate	restore

STS-13

System	HDL	OnWN	FNWN	SMT	mean	Rank
Dkpro	0.735	0.735	0.341	0.323	0.565	6
TakeLab	0.486	0.633	0.269	0.279	0.434	58
ADW (STS-13)	0.621	0.511	0.446	0.384	0.502	34
ADW (All) GP	0.717	0.697	0.411	0.272	0.538	20
ADW (All) LR	0.667	0.735	0.409	0.374	0.565	6