Empowering Citizens. Smarter Societies.

Insight
Centre for Data Analytics

Lexical sense alignment using weighted bipartite *b*-matching

Sina Ahmadi (ULD)
Supervisors: John McCrae, Mihael Arcan

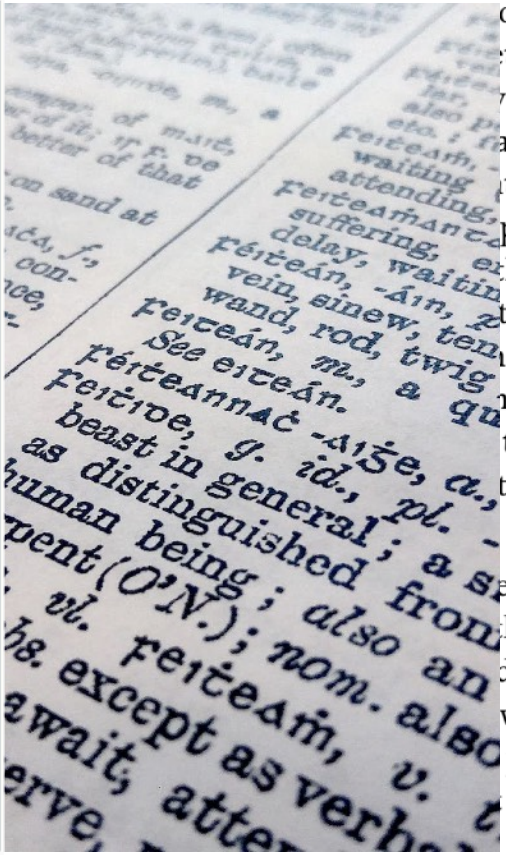
16th Ph.D. day on April 8, 2018

A World Leading SFI Research Centre

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI)under Grant Number SFI/12/RC/2289, co-funded under the European Regional Development Fund

Lexical resources

Langenscheidt dictionaries



onstrued. In chapter 6, following Durkheim, we
ties, such as Australia, hold some events, places
e attacks on them as sacrilegious Durkheim posed
al an in
t the s.
plain d
he A se
the s re
at cl of
n. In e,
the s t,
ts un d
es th er
he sa ns
d to e d
ward a
perp in
to erase from history knowledge of holocausts.

English
sacrilegious | ˌsakrɪ'lɪdʒəs | adjective
involving or committing sacrilege: *a sacrilegious act | it seems sacrilegious to say this, but it's really not that great a film.* more

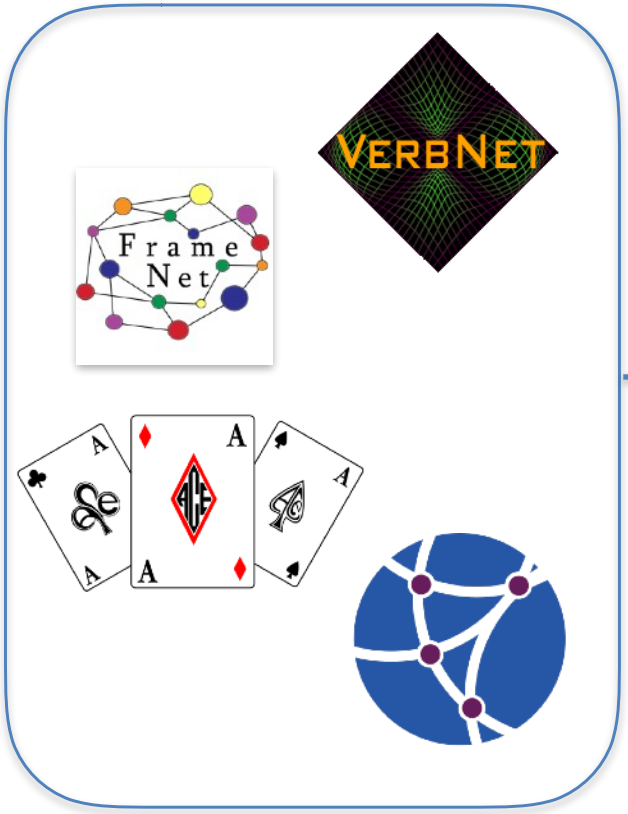
English Thesaurus
sacrilegious adjective
he condemned the book as a vicious, sacrilegious attack on their faith: PROFANE, blasphemous, impious, sinful, irreverent, irreligious, godless, more

French - English
sacrilegious | BrE ˌsakrɪ'lɪdʒəs, AmE ˌsækrə'lɪdʒəs | adjective
Religion, figurative, humorous sacrilège

German - English

Dictionary Siri Knowledge

Lexical resources



Expert-made

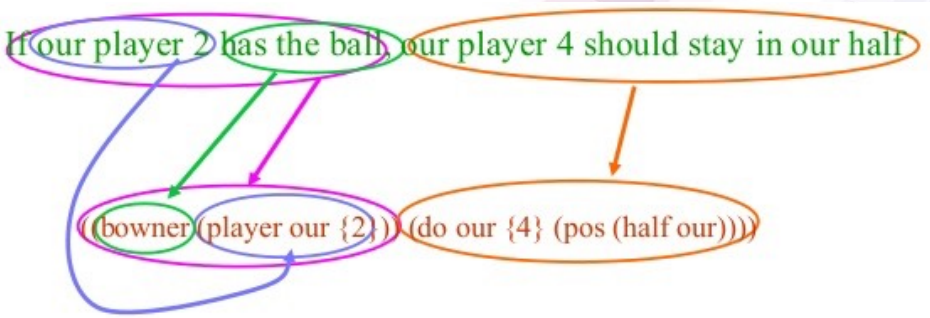
Collaboratively-curated

Why combining resources?

- To improve word and concept coverage
  - e.g., named entities, new senses
- To improve domain coverage
- To improve multilinguality
  - Creating resources for new language pairs
- To combine expert-made semantic relations
  - e.g., Hypernymy, meronymy, etc.

A few applications

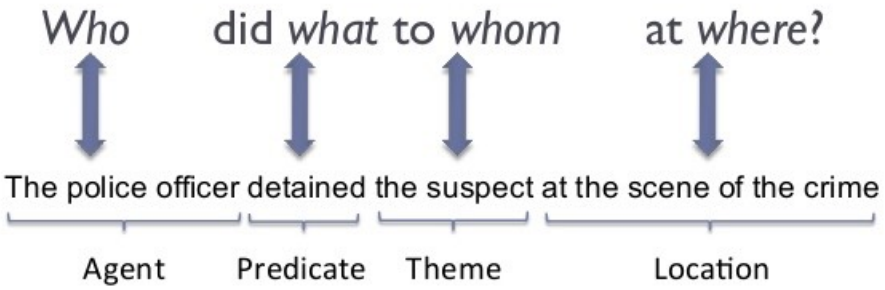
- Semantic parsing



- Word-sense disambiguation and entity linking

bat: or

- Semantic role labeling



Difficulty of resource alignment

Spring (n)

6 senses

18 senses

How does our resource alignment work?

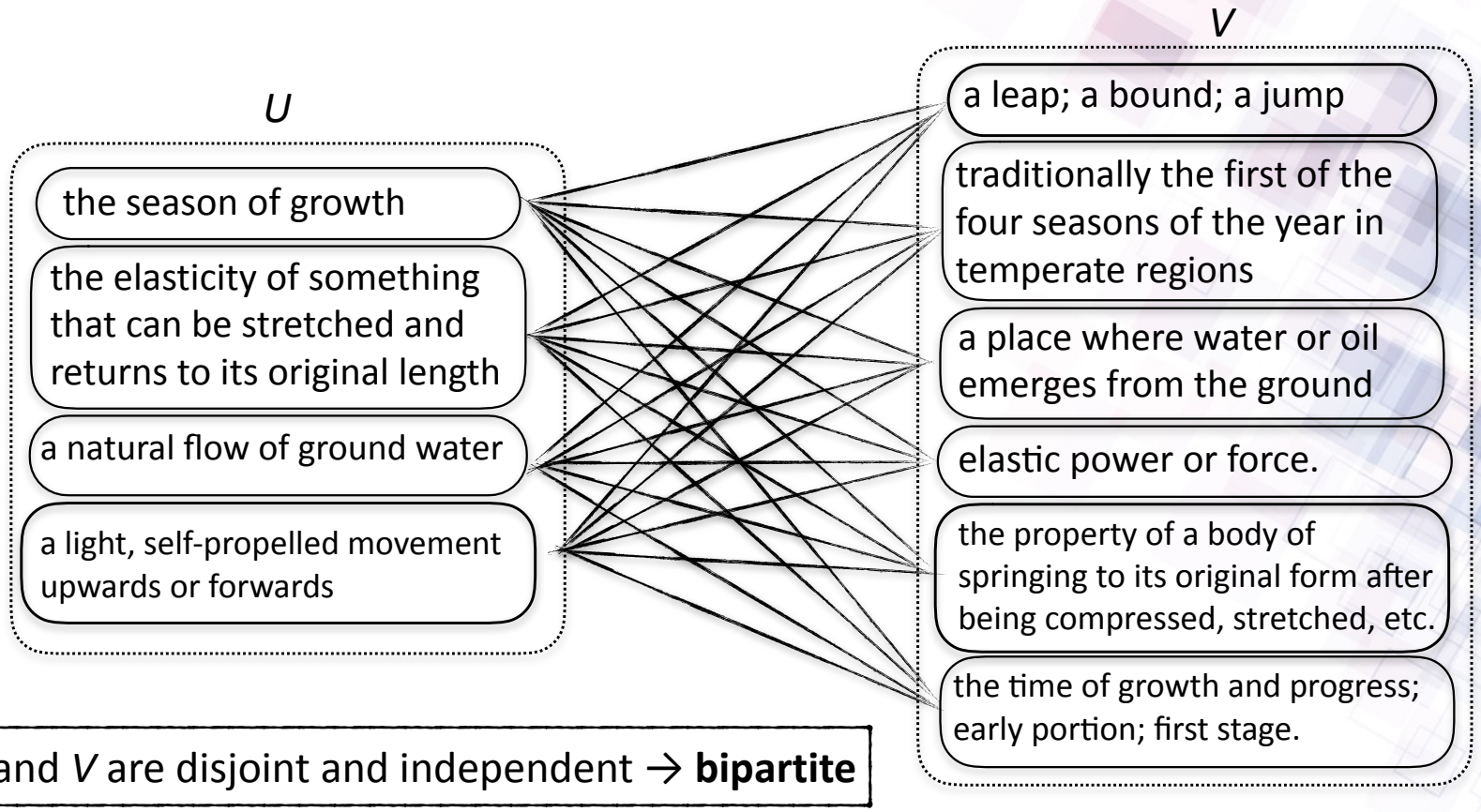
WordNet:spring

- the season of growth
- the elasticity of something that can be stretched and returns to its original length
- a natural flow of ground water
- a light, self-propelled movement upwards or forwards

Wiktionary:spring

- a leap; a bound; a jump
- traditionally the first of the four seasons of the year in temperate regions
- a place where water or oil emerges from the ground
- elastic power or force.
- the property of a body of springing to its original form after being compressed, stretched, etc.
- the time of growth and progress; early portion; first stage.

Step 1. Alignment problem as a graph



U and *V* are disjoint and independent → **bipartite**

Step 2. Extract similarity scores

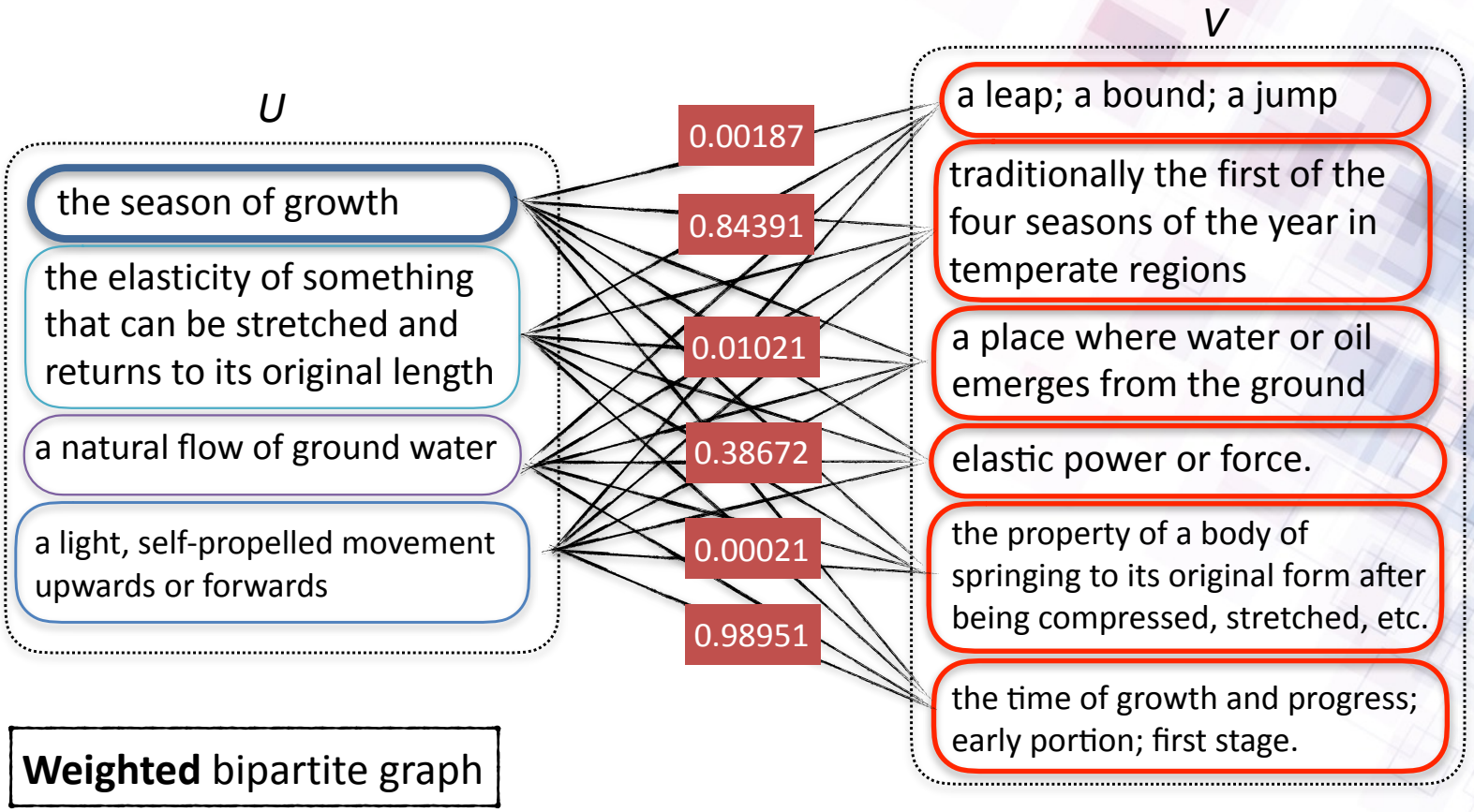
Determine how similar two senses are by training a model using textual and definitional similarity features such as:

- Word length ratio
- Longest common subsequence
- Jaccard measure
- Word embeddings
- Forward precision

*

* McCrae, John P., and Paul Buitelaar. "Linking Datasets Using Semantic Textual Similarity." *Cybernetics and Information Technologies* 18.1 (2018): 109-123.

Step 2. Extract similarity scores

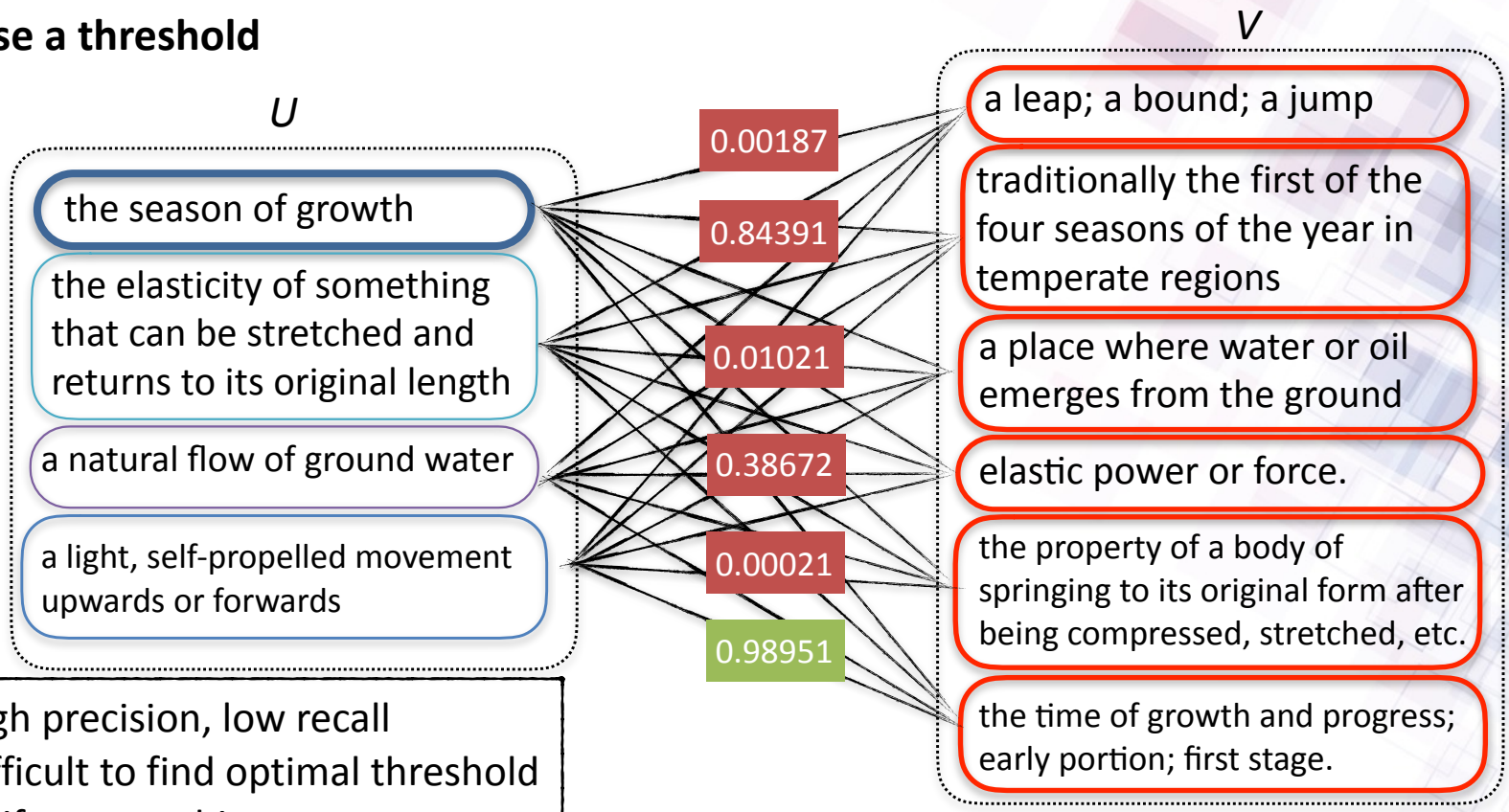


Weighted bipartite graph

Step 3. Graph matching

Exhaustive matching

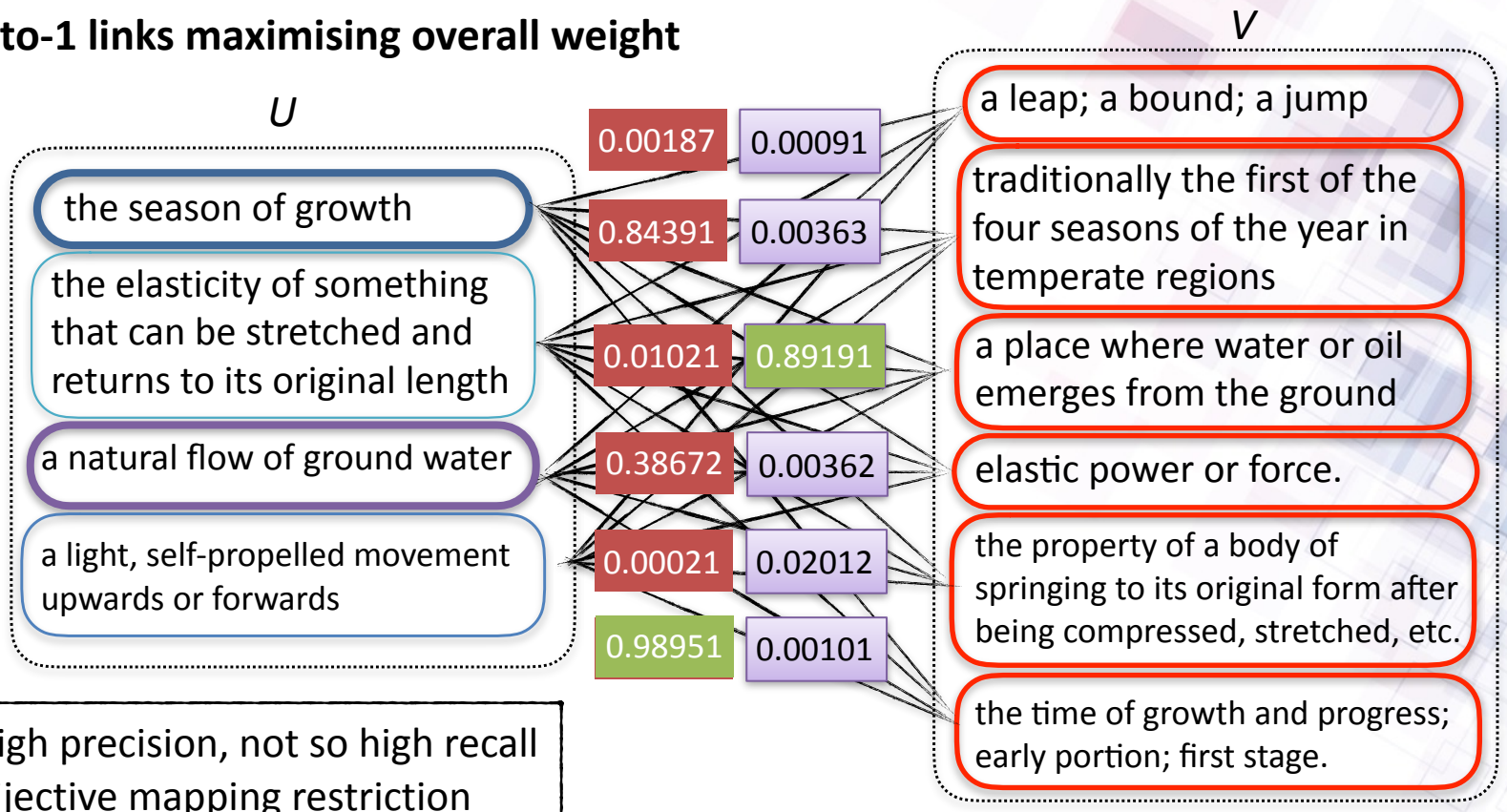
→ **Use a threshold**



- High precision, low recall
- Difficult to find optimal threshold
- Uniform matching

Exact matching

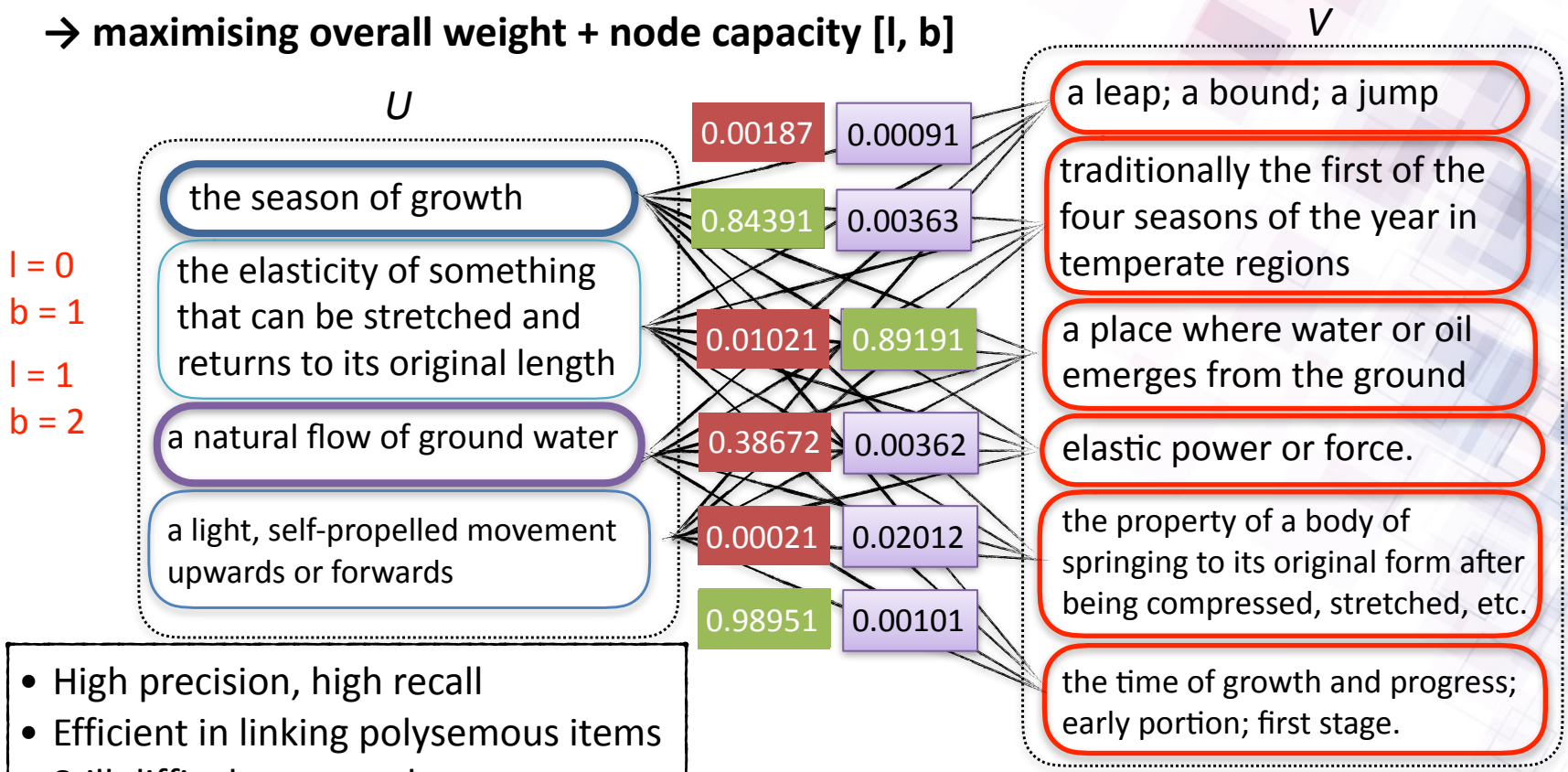
→ **1-to-1 links maximising overall weight**



- High precision, not so high recall
- Bijective mapping restriction

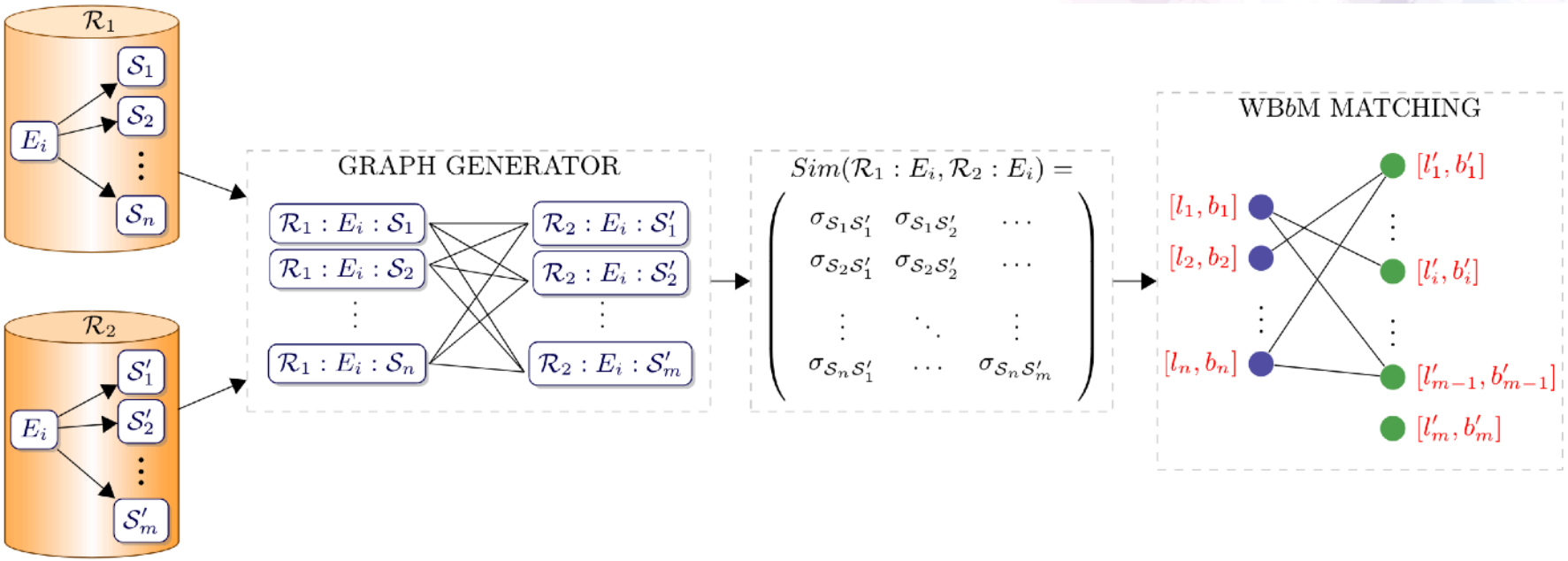
Weighted bipartite *b*-matching (WB*b*M)

→ **maximising overall weight + node capacity [l, b]**



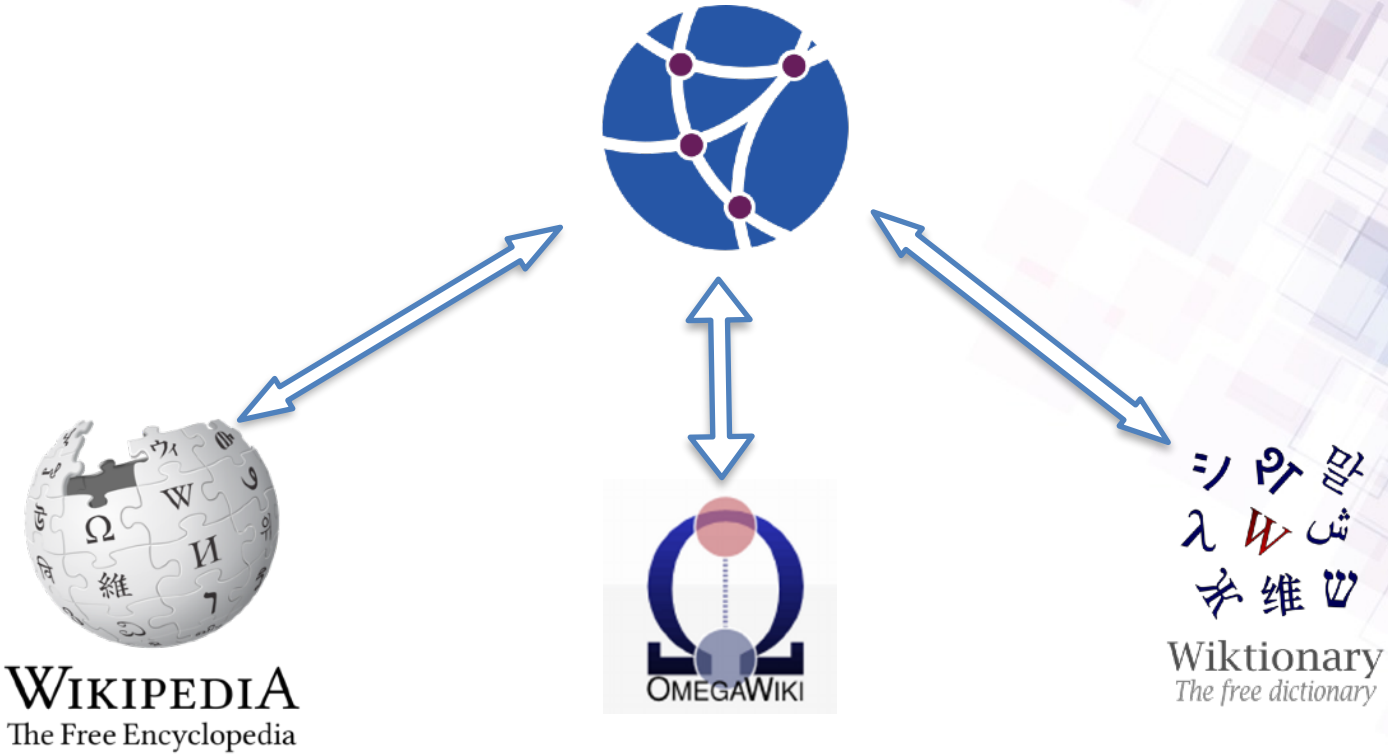
- High precision, high recall
- Efficient in linking polysemous items
- Still difficult to tune the parameters

Our resource alignment mechanism: schema



Alignment Experiments: Datasets

WordNet synsets manually mapped to their corresponding concepts *



* Matuschek, Michael, and Iryna Gurevych. "Dijkstra-wsa: A graph-based approach to word sense alignment." *Transactions of the Association for Computational Linguistics* 1 (2013): 151-164.

Experiments: WordNet-Wiktionary alignment

Previous work:

	Precision	Recall	F-1	Accuracy
SemAlign	75.40	62.70	68.40	93.10
SB+DWSA	68.00	71.00	69.00	92.00

Our current method:

Left bound, right bound	Ave. Precision	Ave. Recall	Ave. F-measure	Ave. Accuracy
[0, 1], [0, 1]	81.86	61.83	68.51	69.48
[0, 2], [0, 1]	78.13	70.74	73.28	76.57
[0, 3], [0, 1]	77.88	71.38	73.59	77.13
[1, 2], [1, 2]	81.21	74.17	76.59	79.49
[1, 3], [1, 3]	**81.26**	75.02	77.12	80.14
[1, 5], [0, 1]	81.25	**75.25**	**77.28**	**80.33**
[1, 5], [1, 2]	81.25	75.23	77.26	80.32

Conclusion

- Graph matching algorithms can be efficiently applied to lexical alignment problems
- **WB*b*M**
  - includes more possible linking combinations by defining capacity
  - efficient in lexical alignment providing high precision and recall
  - difficult to find optimal parameters
  - highly dependent on the textual and definitional similarities

Future directions

- Exploring link prediction methods for lexical alignment
- Extend our researches to multi-lingual resources
- Using graph neural networks (GNNs)