Word Re-Embedding via Manifold Learning

Souleiman Hasan

Lecturer, Maynooth University, School of Business and Hamilton Institute

souleiman.hasan@mu.ie

Maynooth University
National University of Ireland Maynooth

Machine Learning Dublin Meetup, Dublin, Ireland, 26 Feb 2018

Based On

- Souleiman Hasan and Edward Curry. "Word Re-Embedding via Manifold Dimensionality Retention." *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP* 2017)

  http://aclweb.org/anthology/D17-1033

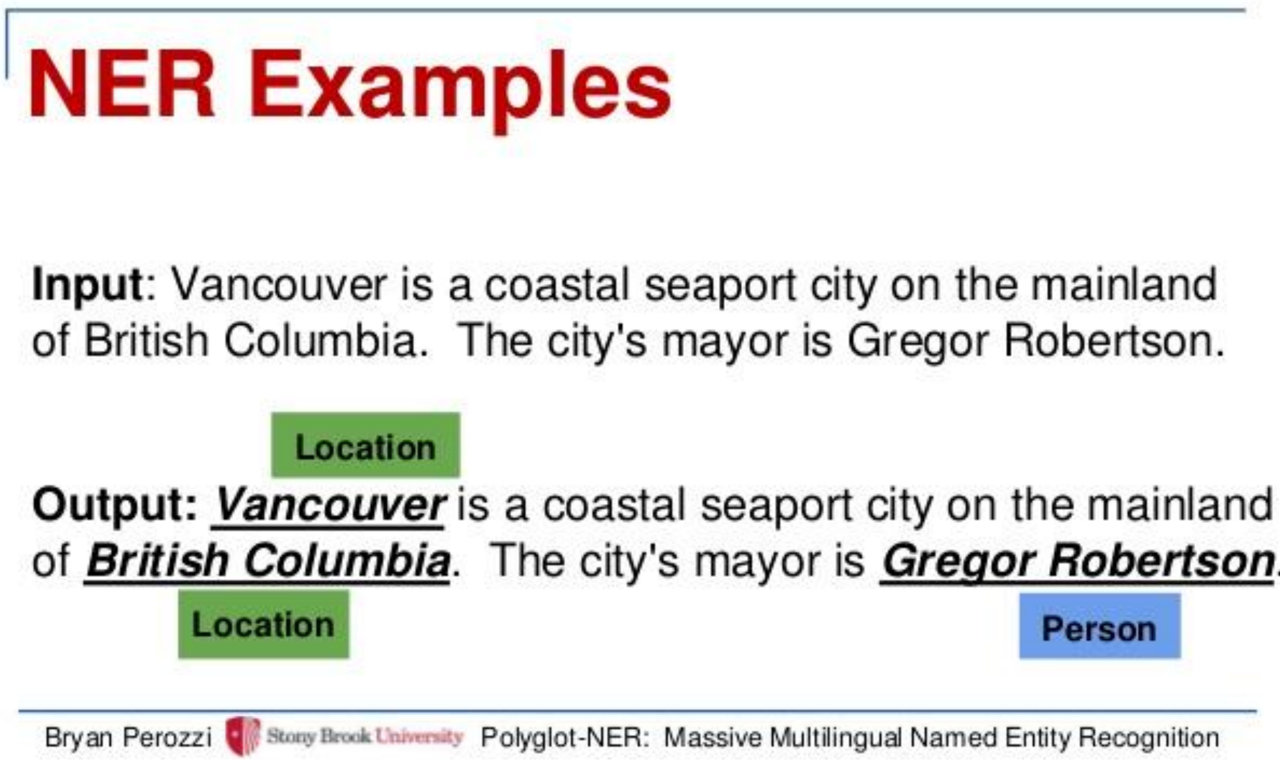
NUI Galway OÉ Gaillimh
Science Foundation Ireland sfi For what's next
lero THE IRISH SOFTWARE RESEARCH CENTRE
Ireland's European Structural and Investment Funds Programmes 2014-2020
Co-funded by the Irish Government and the European Union
European Union European Regional Development Fund
SOUTHERN Regional Assembly Promoting our Region

Outline

- Motivation
  - NLP tasks, Semantics, Word embeddings
- Background
  - Mathematical Structures, Manifold Learning
- Word Re-embedding
  - Methodology and Related Work
  - Approach
  - Results

NLP Tasks

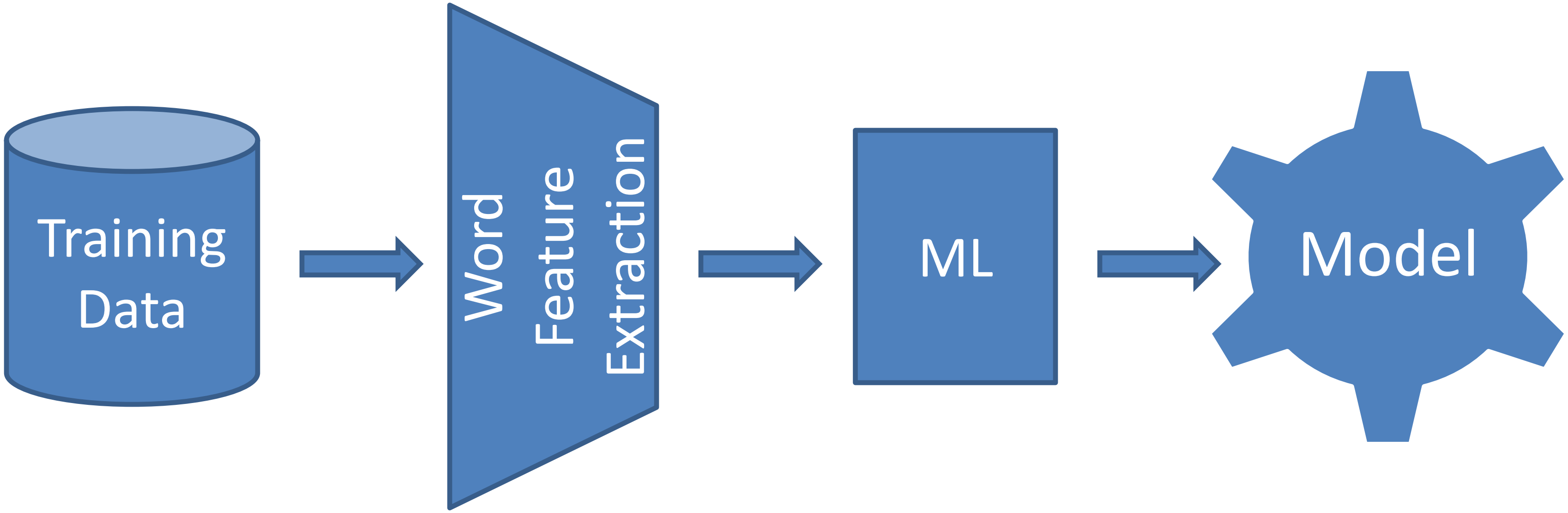
– Named entity recognition



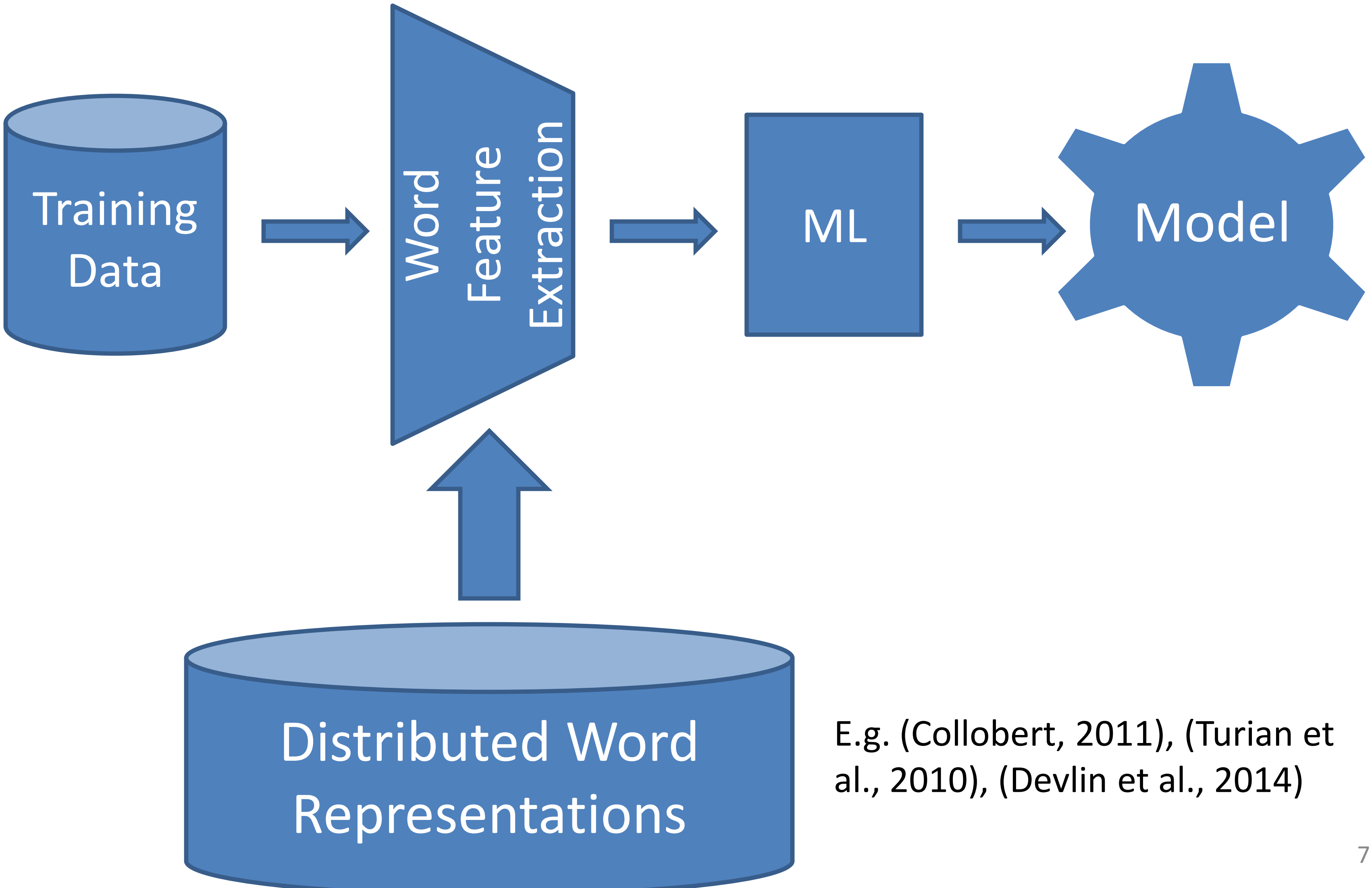
NLP Tasks

– Sentiment Analysis

Generic NLP Supervised Model



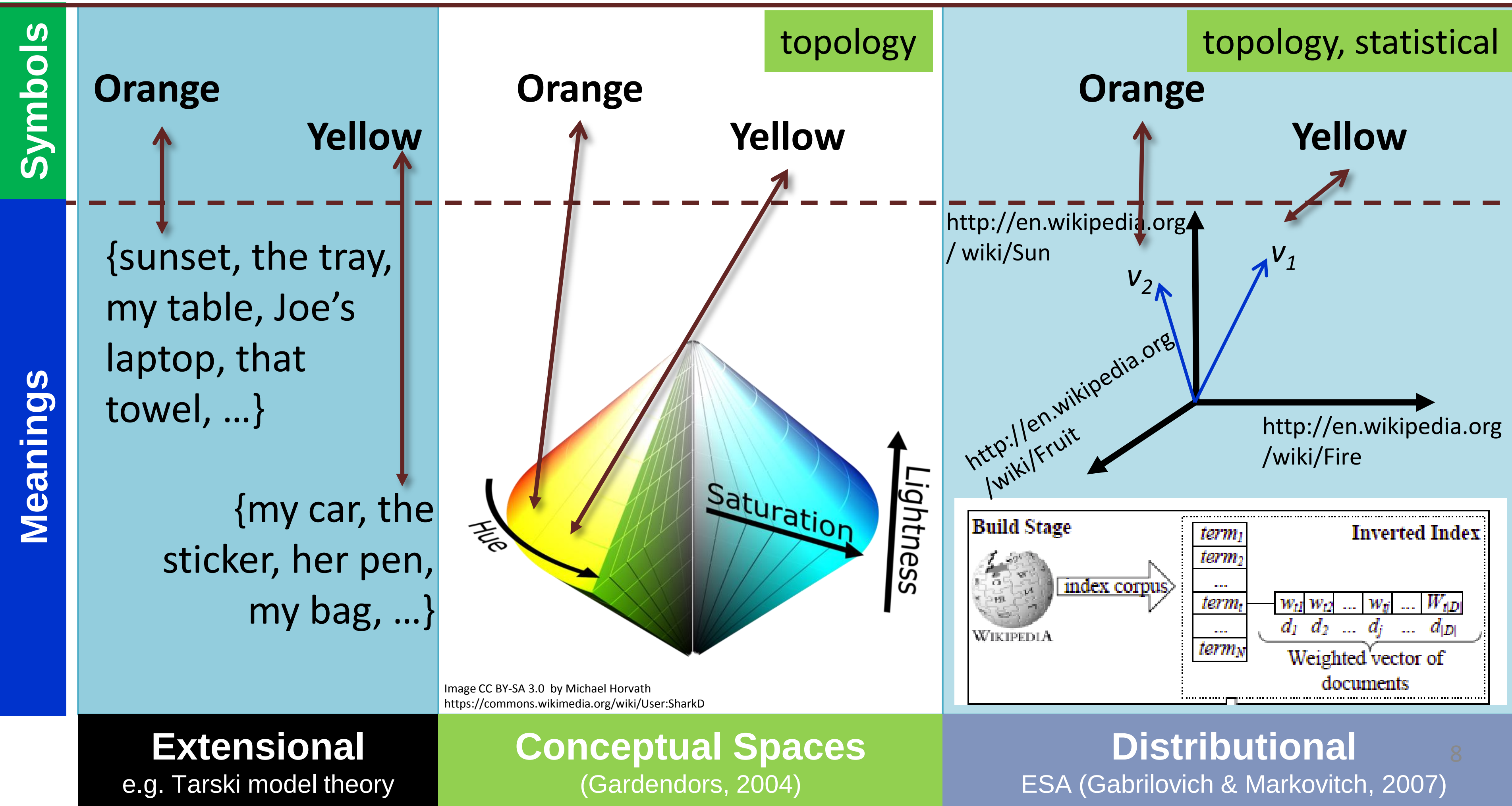
Generic NLP Supervised Model



E.g. (Collobert, 2011), (Turian et al., 2010), (Devlin et al., 2014)

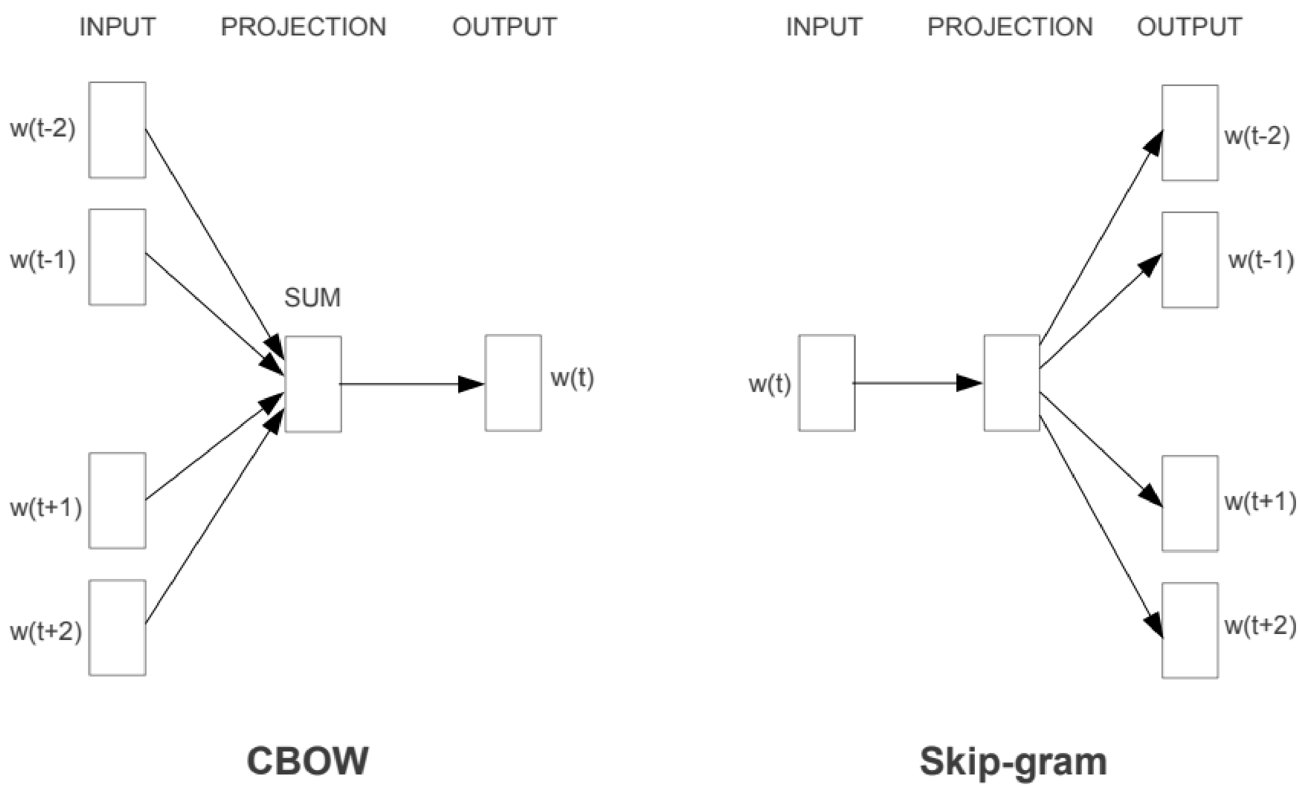
Word Distributed Representations capture Semantics

- Semantics: "The relation between the words or expressions of a language and their meaning." (Gardenfors, 2004)



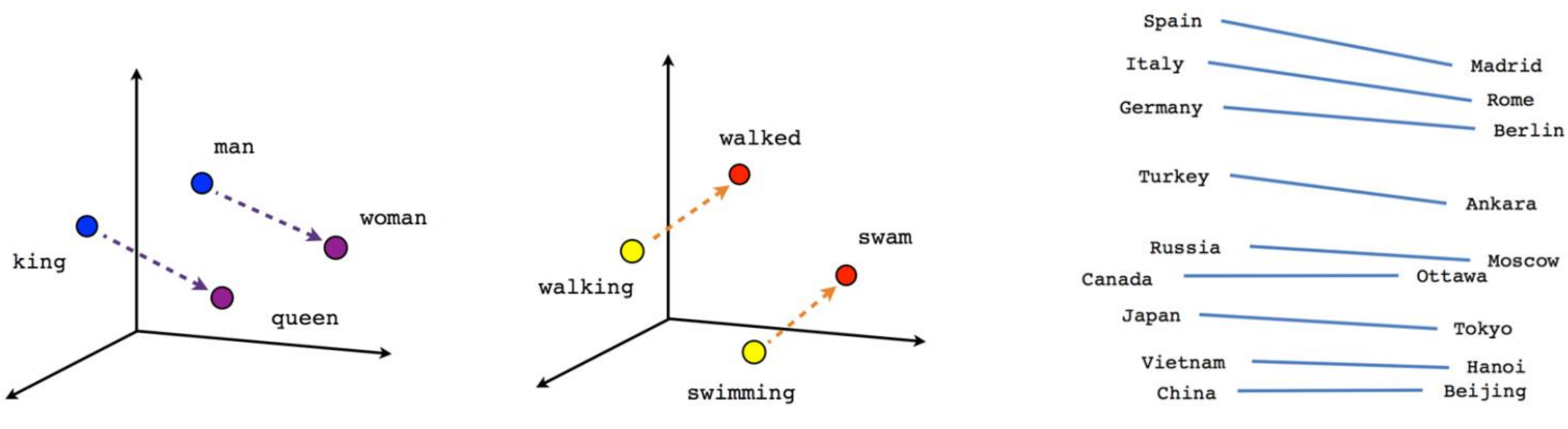
Word Embeddings

- Word2Vec (Mikolov et al., 2013)
- GloVe (Pennington et al., 2014)



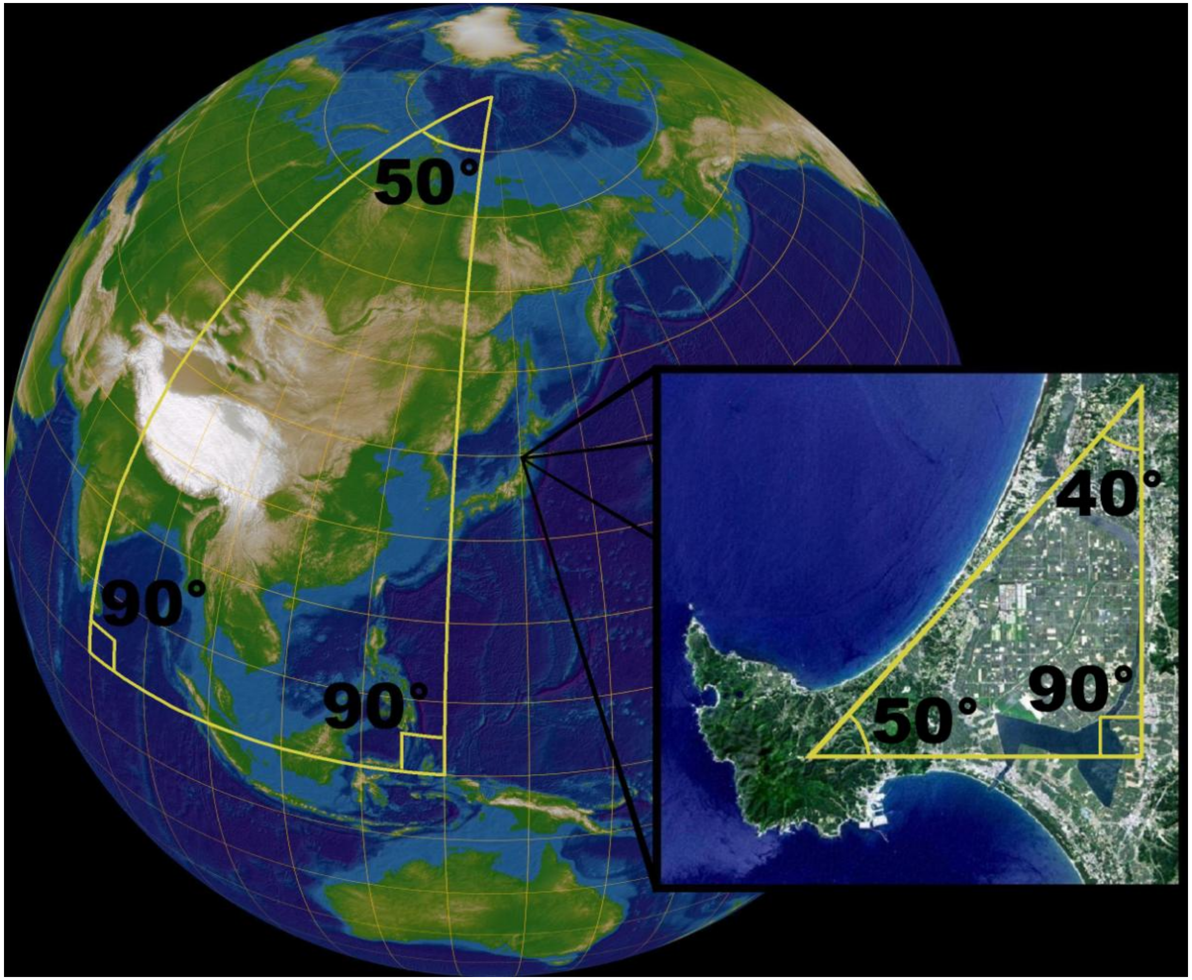
Word Embeddings

- Word2Vec (Mikolov et al., 2013)
- GloVe (Pennington et al., 2014)

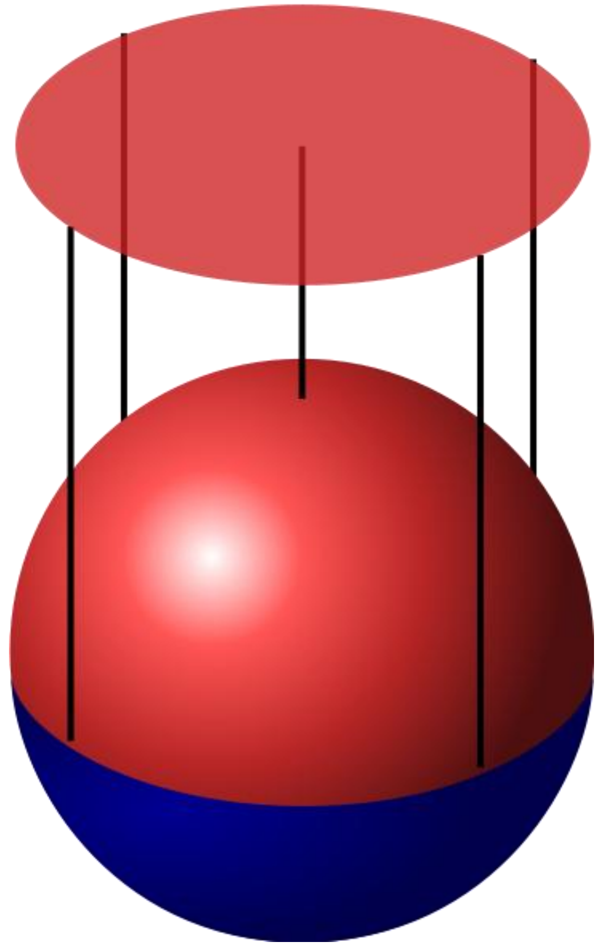


Male-Female　　Verb tense　　Country-Capital

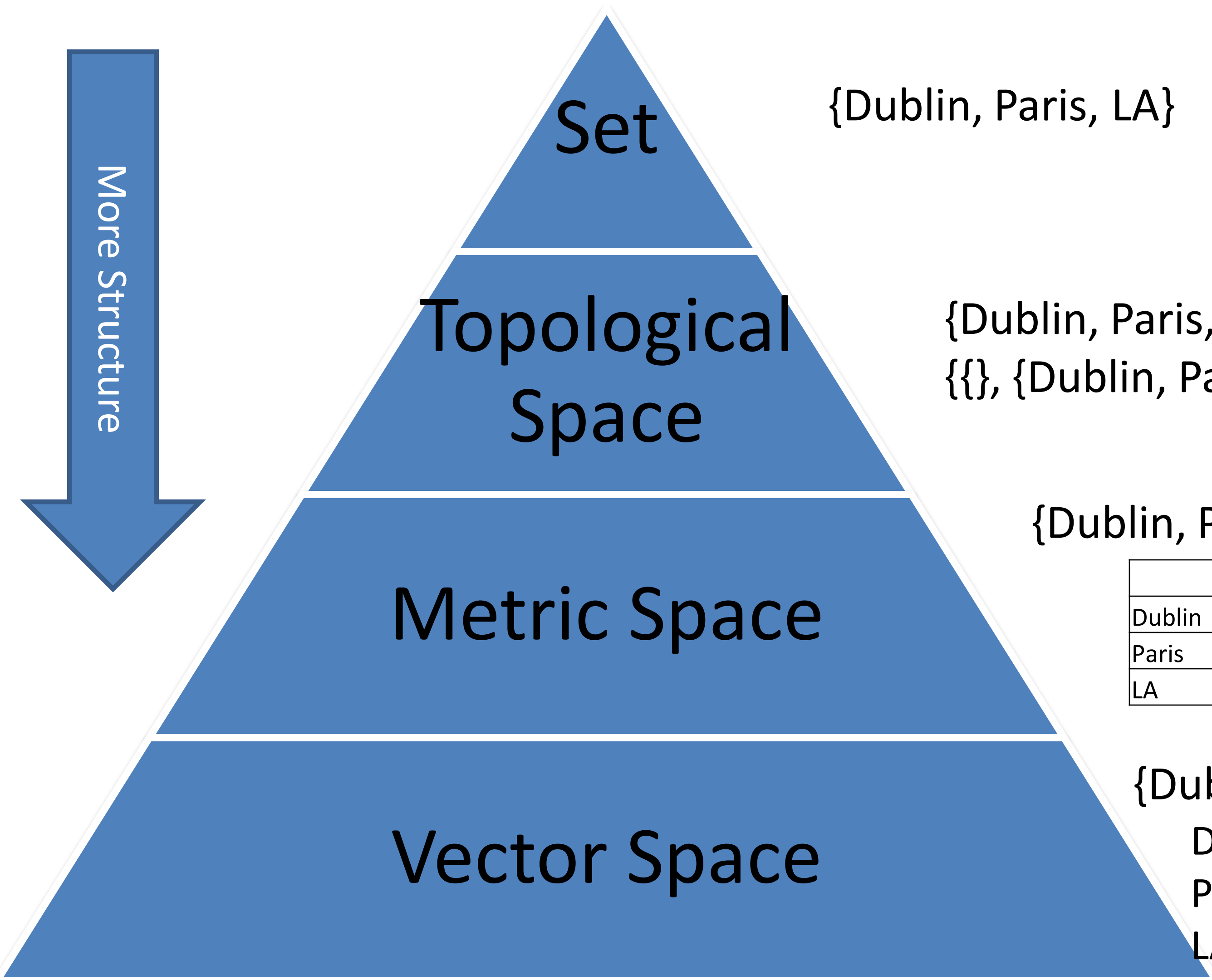
Why Called Embedding?



Why Called Embedding?



Mathematical Structures



{Dublin, Paris, LA}

{Dublin, Paris, LA}
{{}, {Dublin, Paris}, {Dublin, Paris, LA}}

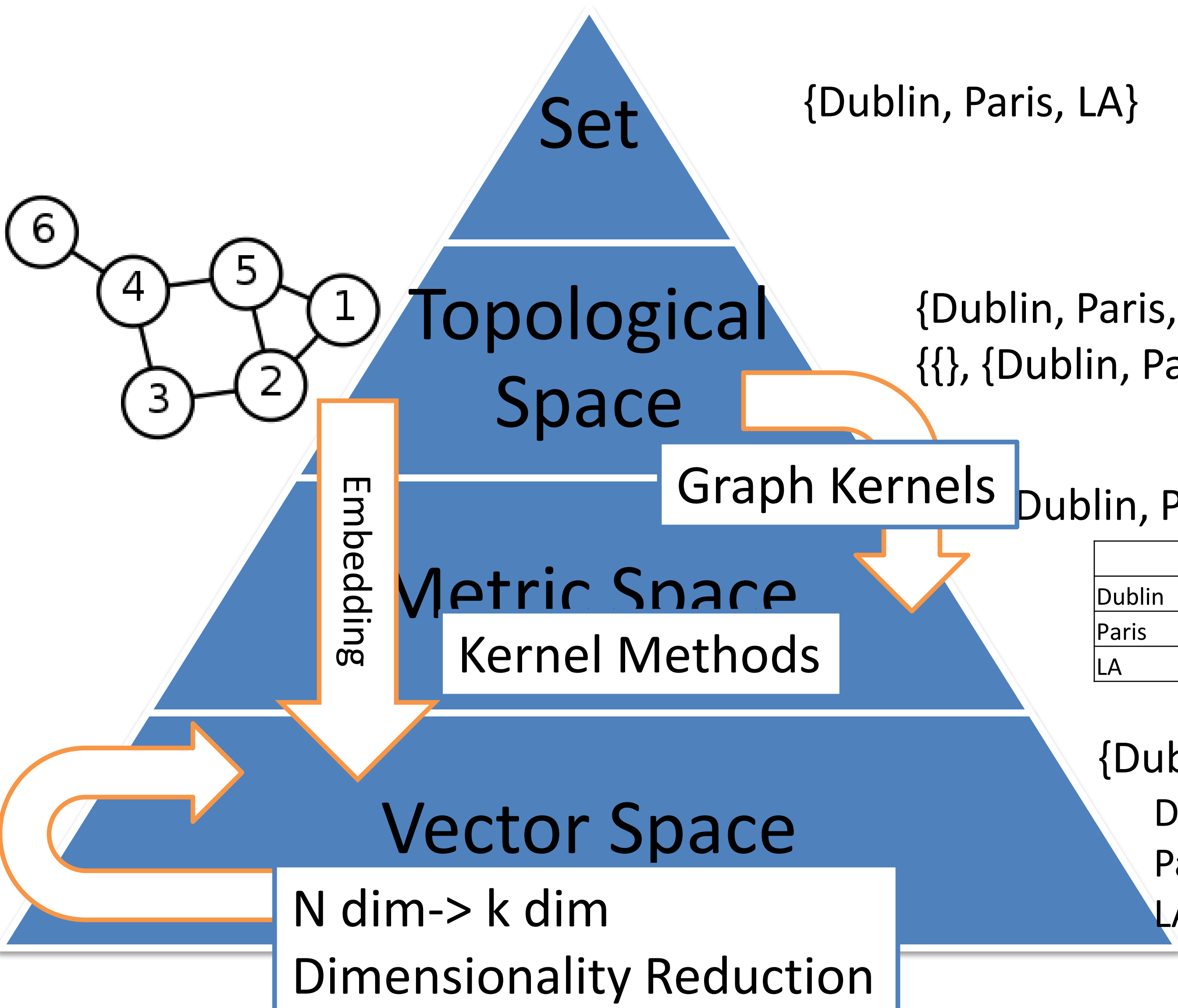
{Dublin, Paris, LA}

	Dublin	Paris	LA
Dublin	0	779	8314
Paris	779	0	9096
LA	8314	9096	0

{Dublin, Paris, LA}

Dublin=53.3498° N+6.2603° W
Paris= 48.8566° N+ 2.3522° E
LA= 34.0522° N+ 118.2437° W

Mathematical Structures and ML



{Dublin, Paris, LA}

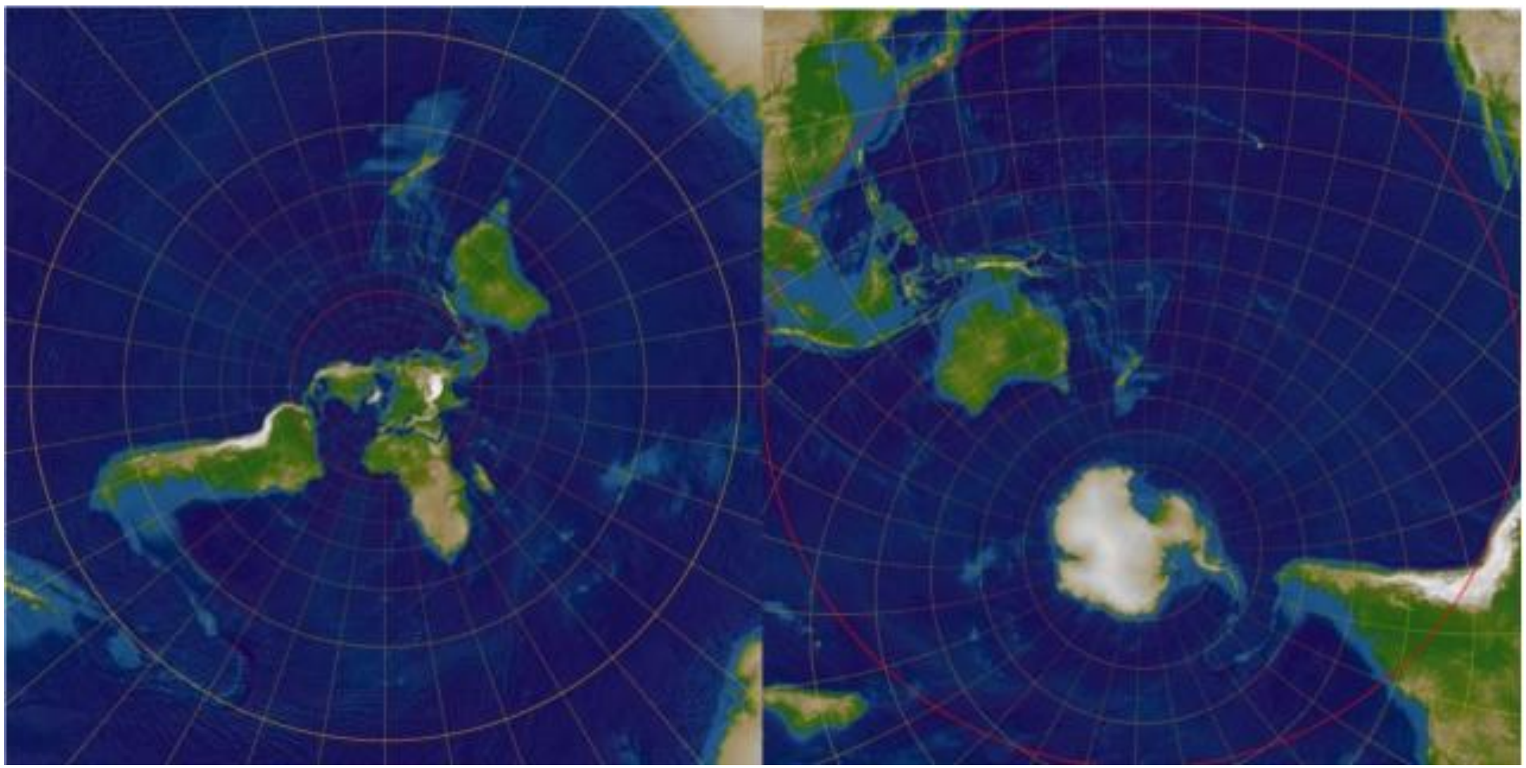
{Dublin, Paris, LA}
{{}, {Dublin, Paris}, {Dublin, Paris, LA}}

Dublin, Paris, LA}

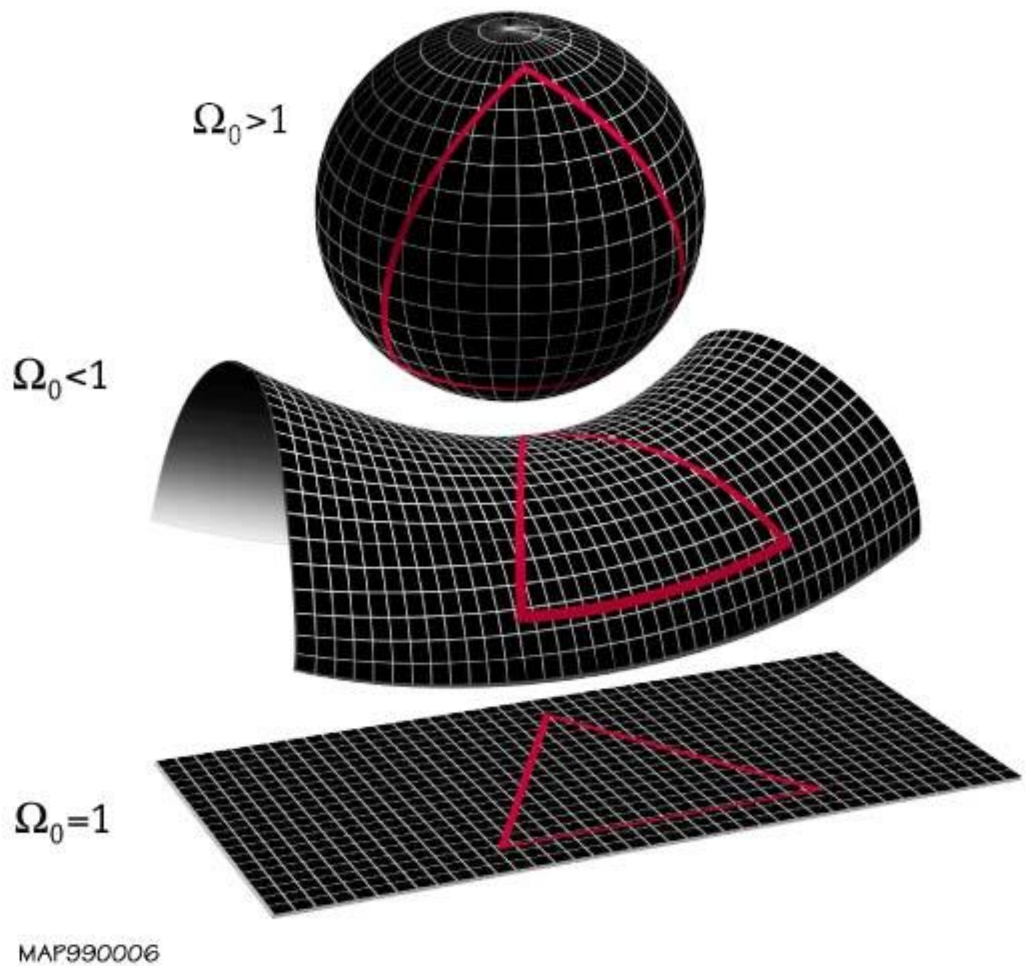
	Dublin	Paris	LA
Dublin	0	779	8314
Paris	779	0	9096
LA	8314	9096	0

{Dublin, Paris, LA}
Dublin=53.3498° N+6.2603° W
Paris= 48.8566° N+ 2.3522° E
LA= 34.0522° N+ 118.2437° W

e.g. Earth Surface 2D Embedding

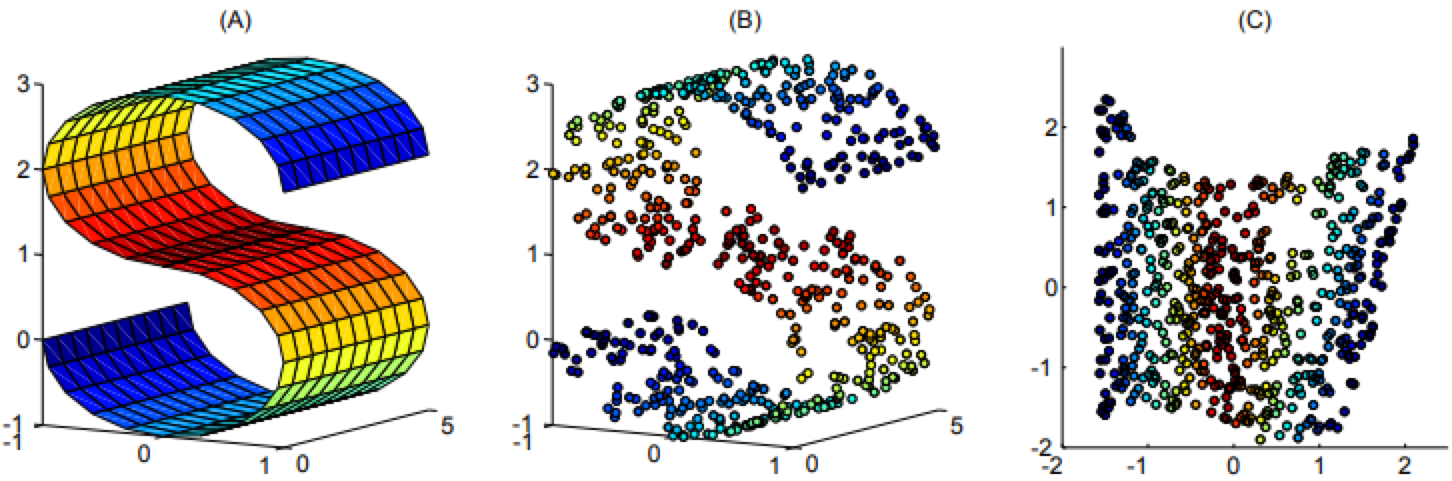


e.g. Shape of the Universe



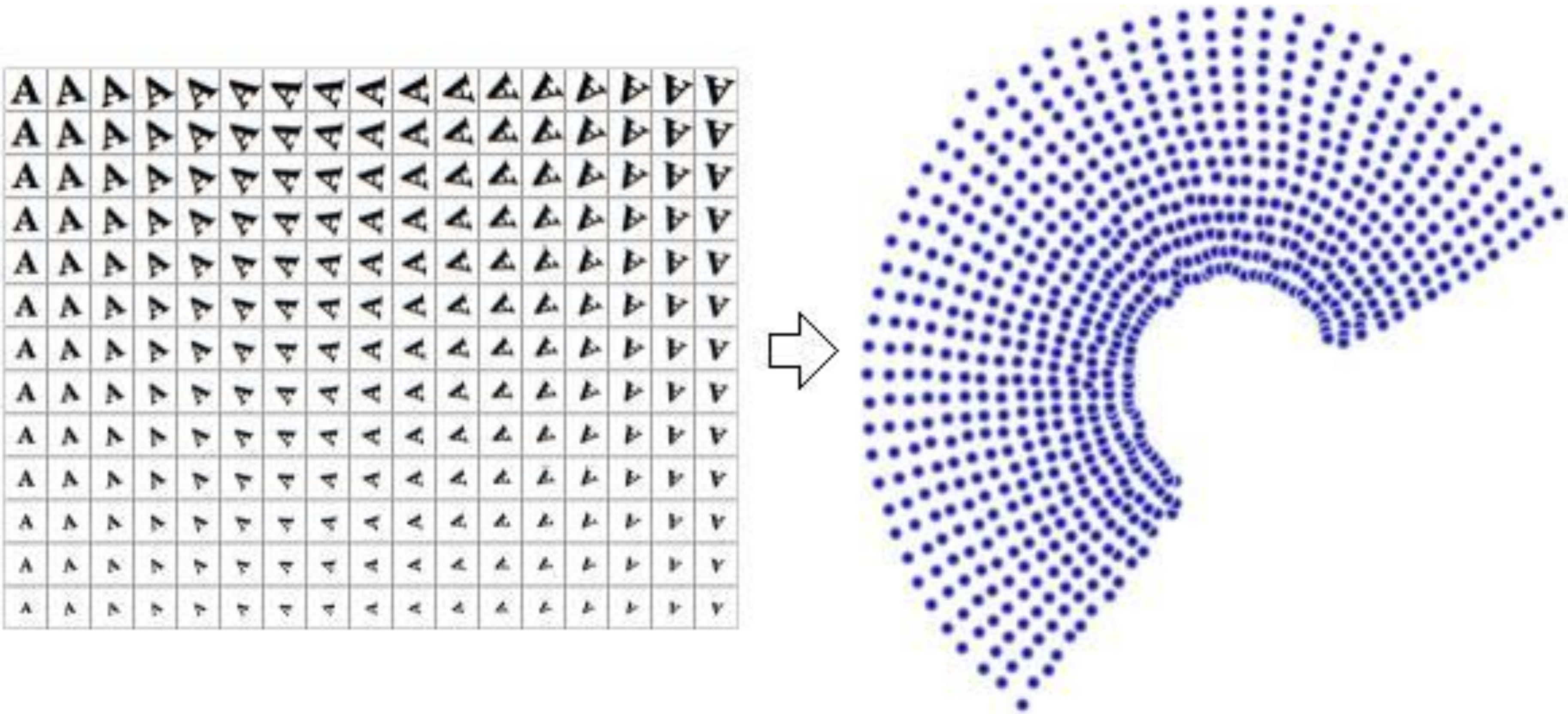
Manifold Learning

- Embedding while preserving the neighbourhood

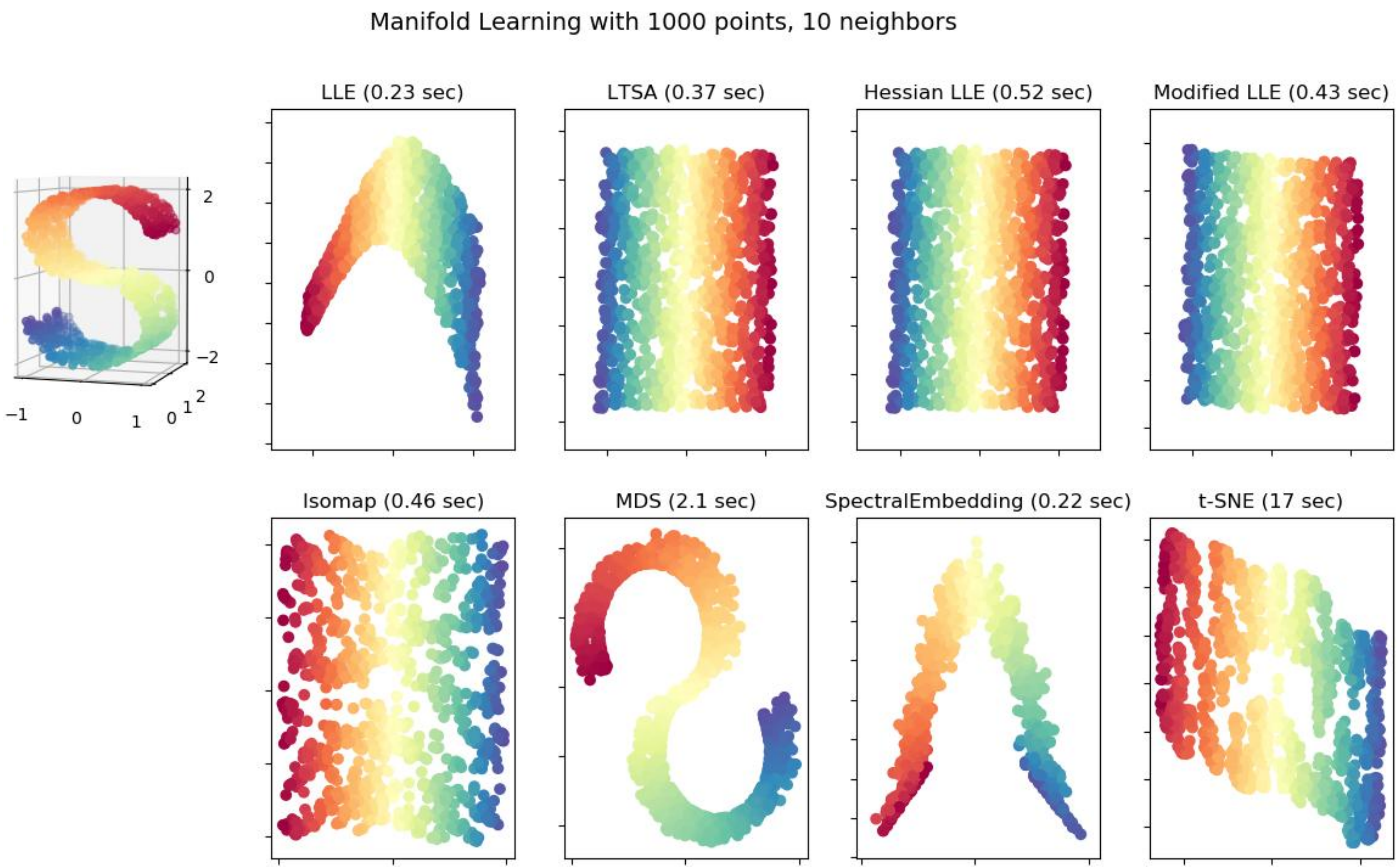


Saul, Lawrence K., and Sam T. Roweis. "An introduction to locally linear embedding.“ *Available at: http://www. cs. toronto. edu/~ roweis/lle/publications. html*(2000).

Manifold Learning for Dimensionality Reduction

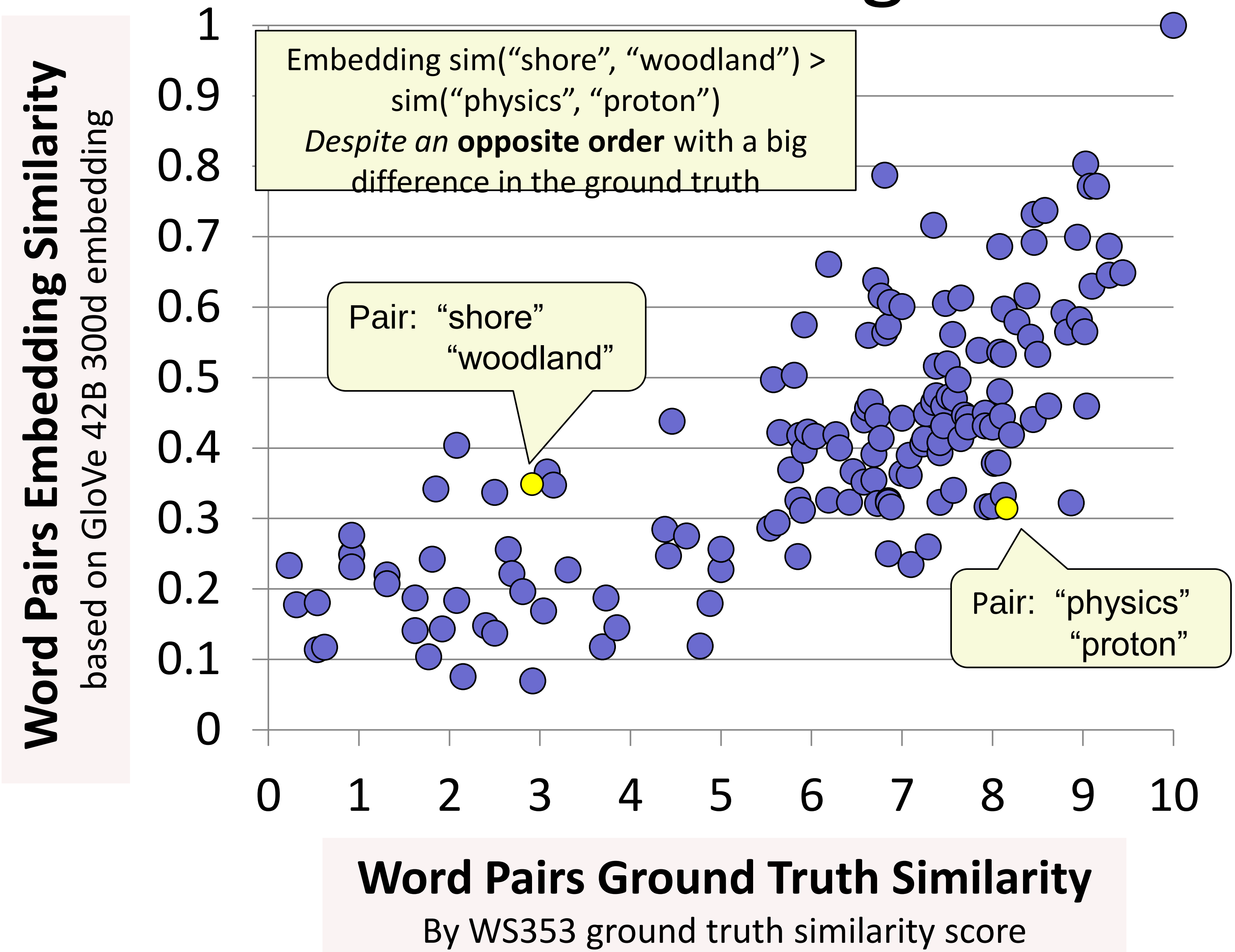


Manifold Learning- Various Algorithms

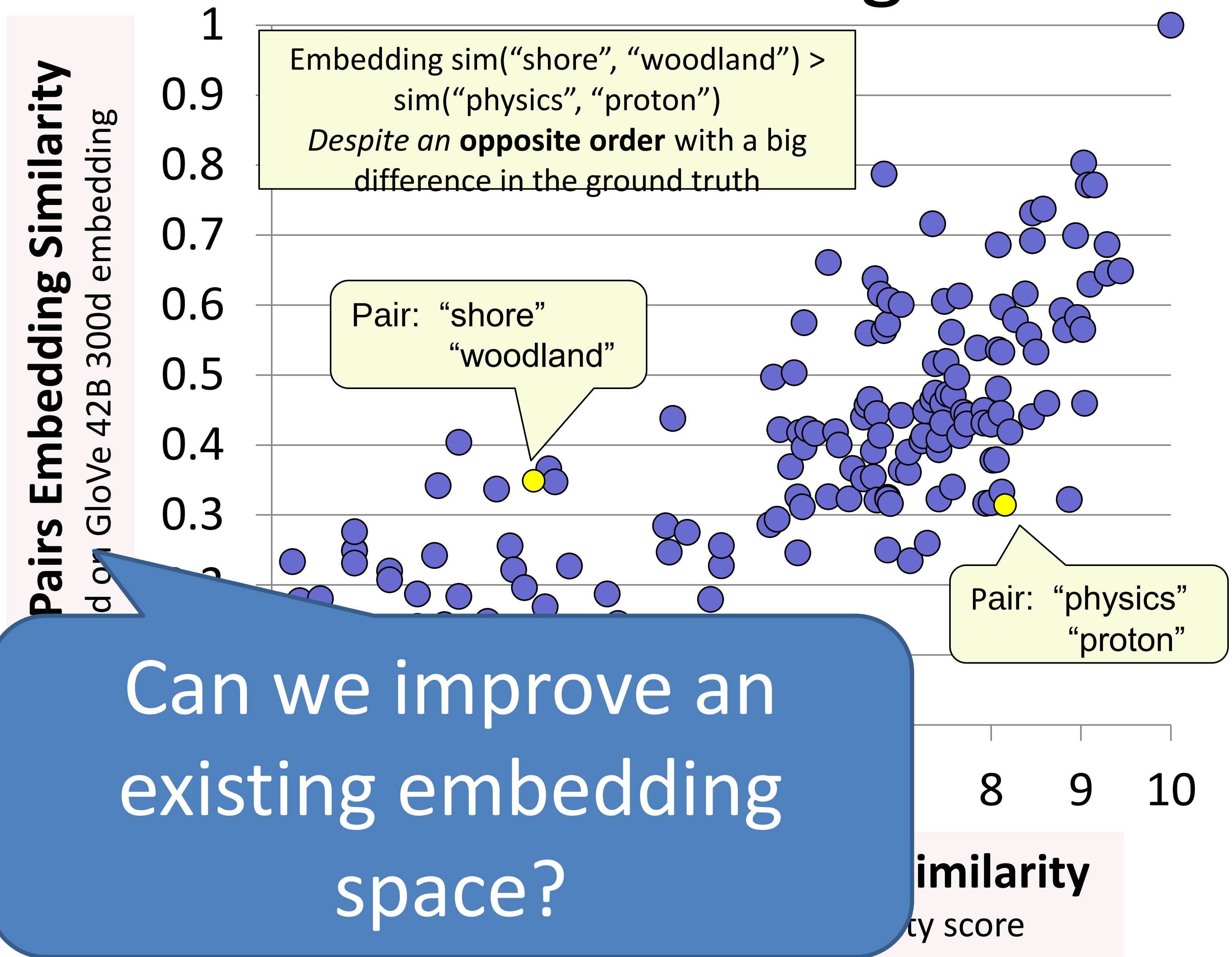


http://scikit-learn.org/stable/auto_examples/manifold/plot_compare_methods.html

Word Re-Embedding: Problem

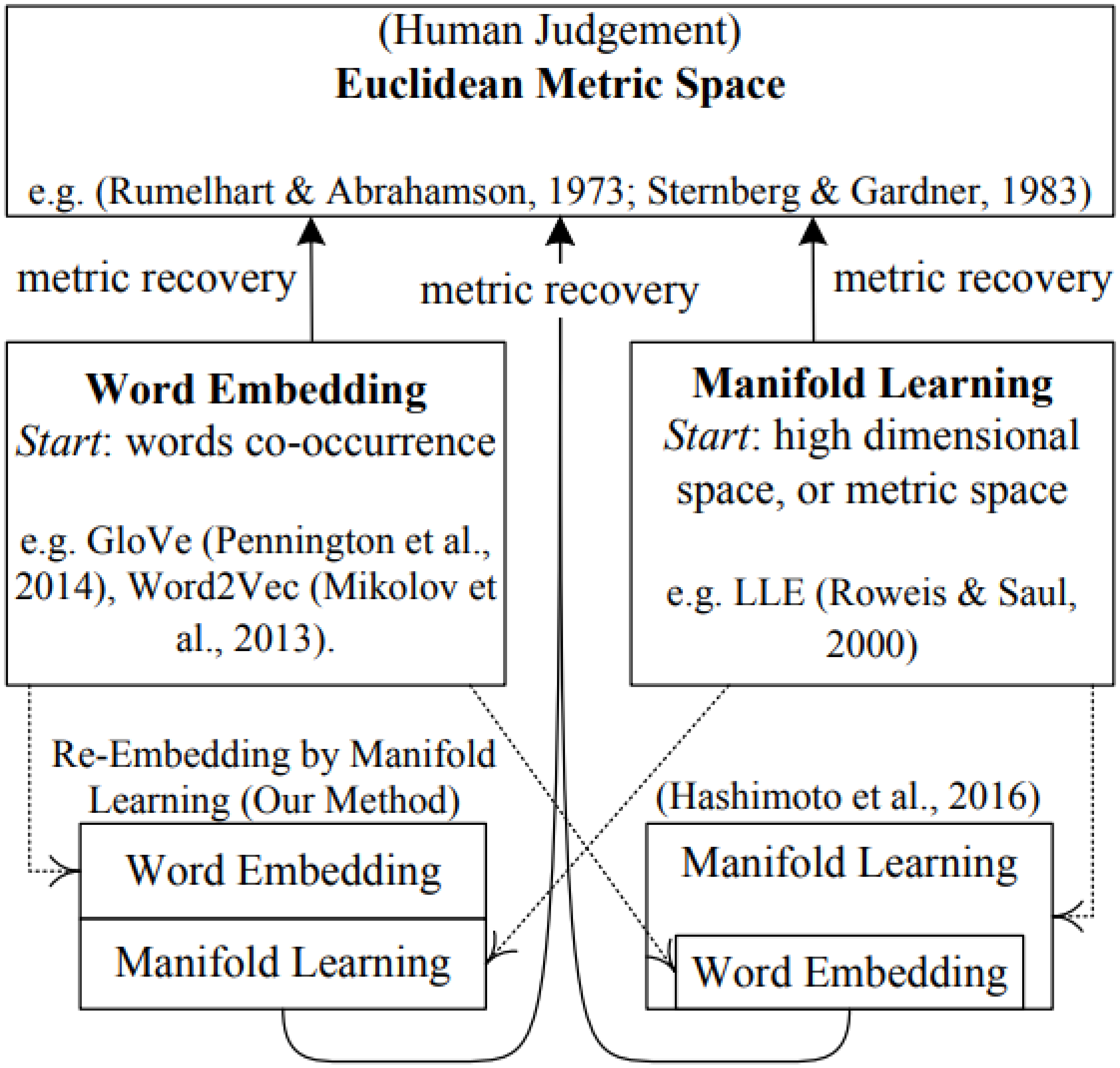


Word Re-Embedding: Problem



Can we improve an existing embedding space?

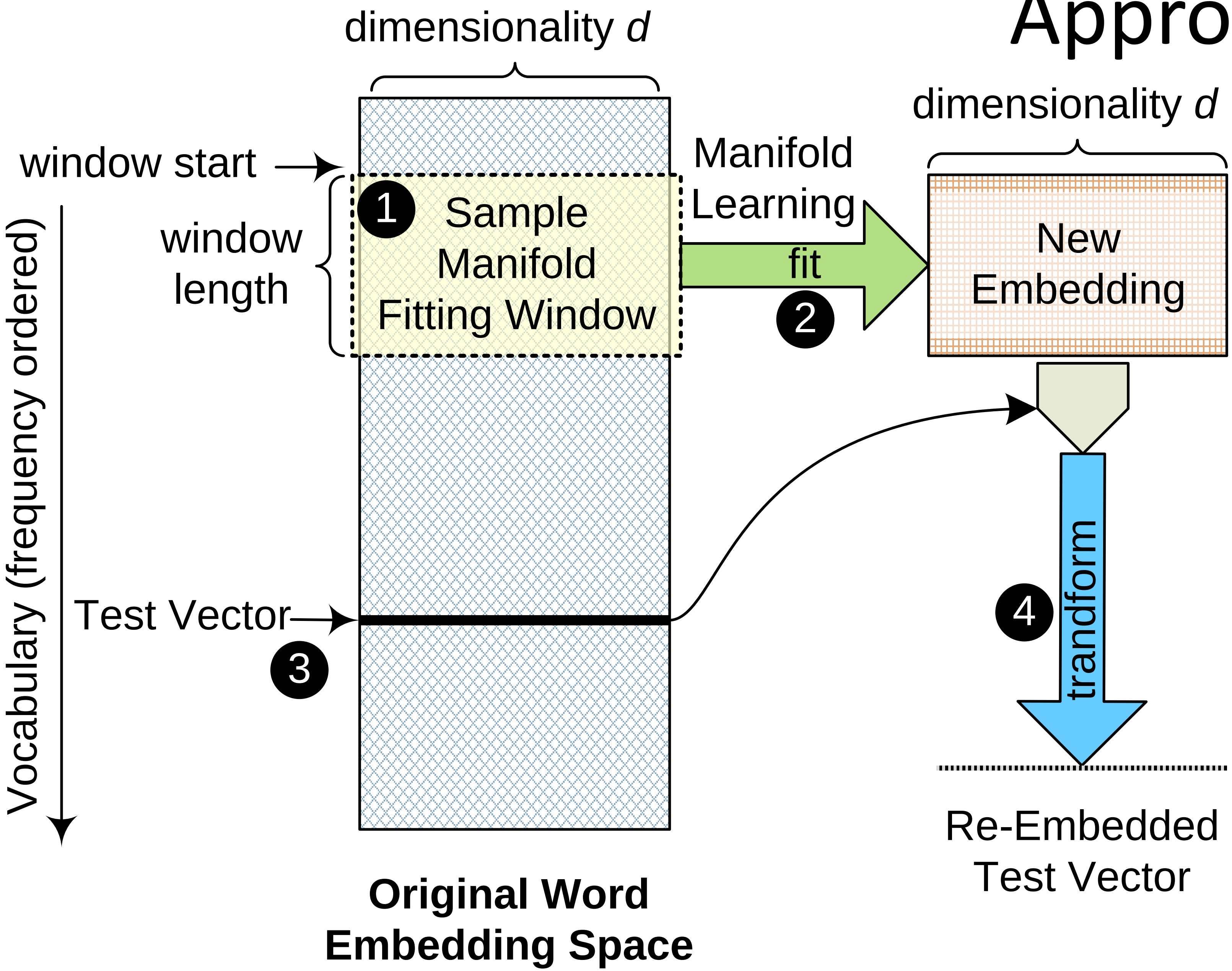
Word Re-Embedding: Methodology



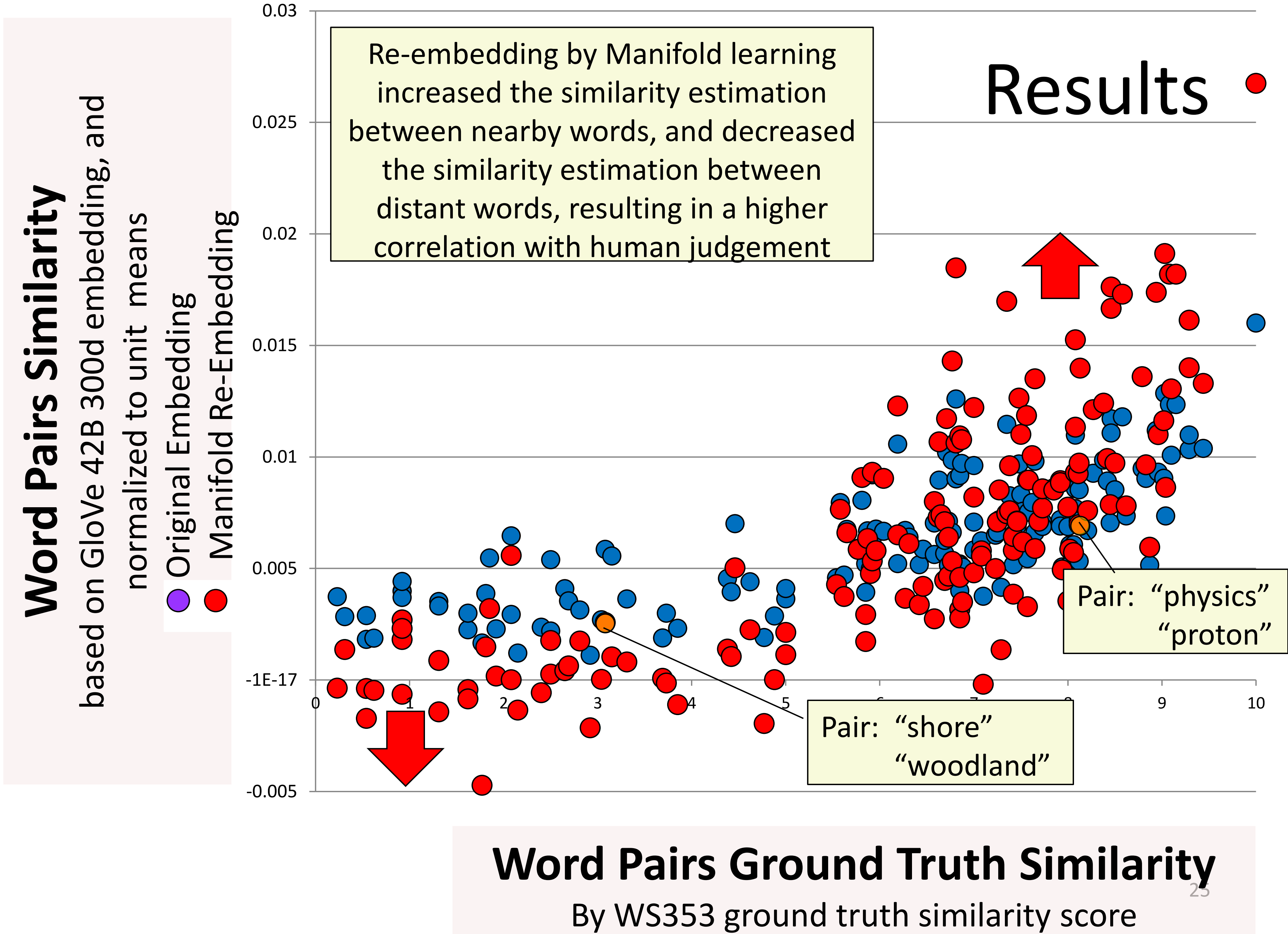
Word Re-Embedding: Related Work

- **Word embedding**: Word2Vec (Mikolov et al., 2013), GloVe (Pennington et al., 2014a)
- **Unified metric recovery** framework for word embedding and manifold learning (Hashimoto et al., 2016)
- **Manifold learning** for dimensionality reduction and embedding: Locally Linear Embedding (LLE) (Roweis and Saul, 2000), Isomap (Balasubramanian and Schwartz, 2002), t-SNE (Maaten and Hinton, 2008), etc.
- **Word embedding post-processing**: (Labutov and Lipson, 2013), (Lee et al., 2016), (Mu et al., 2017)
- **Need for generic, unsupervised, nonlinear, and theoretically-founded model for post-processing**

Approach



dimensionality *d*
window start
window length
Vocabulary (frequency ordered)
1
Sample Manifold Fitting Window
Manifold Learning
fit
2
dimensionality *d*
New Embedding
Test Vector
3
4
trandform
Re-Embedded Test Vector
Original Word Embedding Space

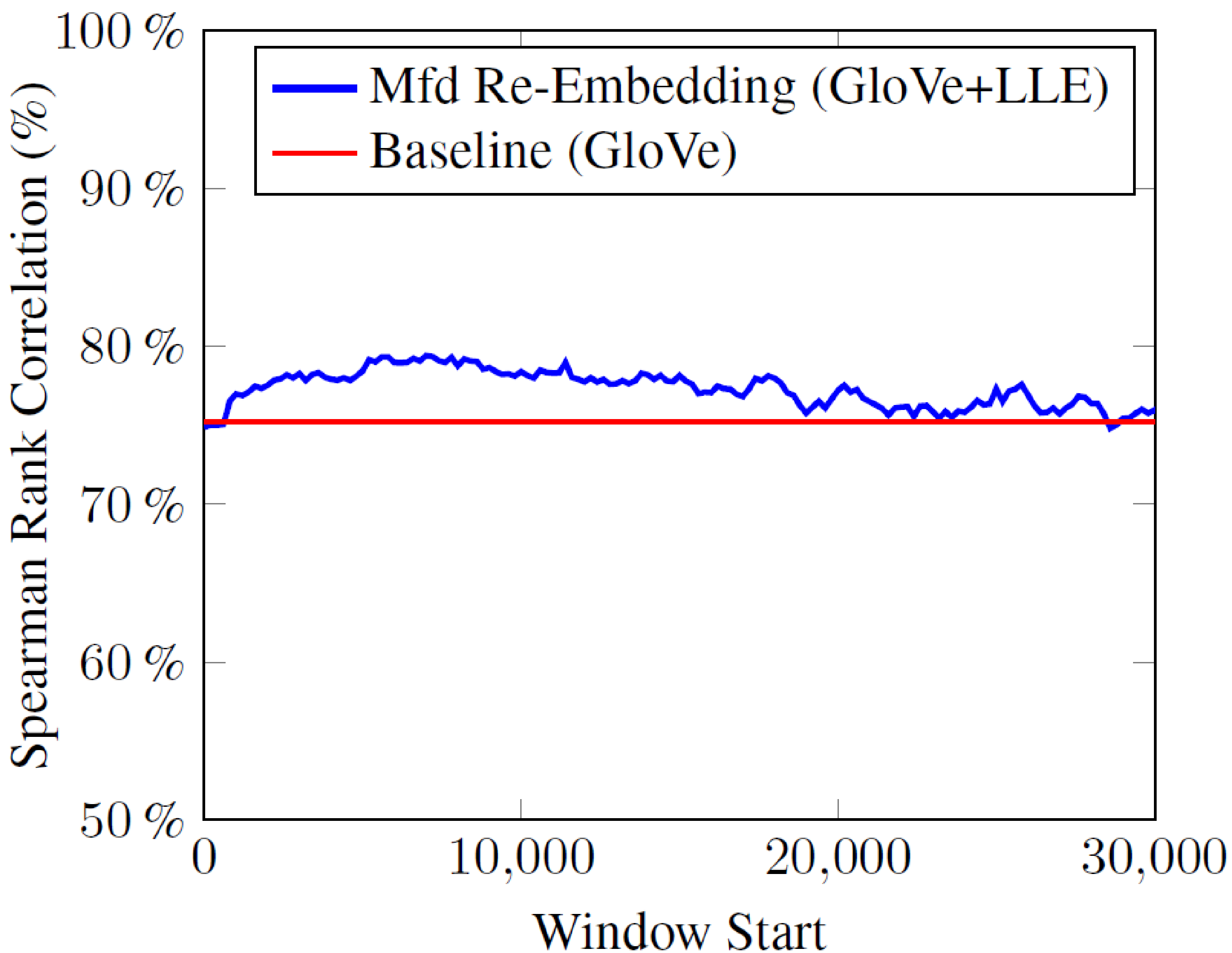

Results
Re-embedding by Manifold learning increased the similarity estimation between nearby words, and decreased the similarity estimation between distant words, resulting in a higher correlation with human judgement
Word Pairs Similarity
based on GloVe 42B 300d embedding, and normalized to unit means
Original Embedding
Manifold Re-Embedding
0.03
0.025
0.02
0.015
0.01
0.005
-1E-17
-0.005
0
1
2
3
4
5
9
10
Pair: “physics” “proton”
Pair: “shore” “woodland”
Word Pairs Ground Truth Similarity
By WS353 ground truth similarity score

Word Re-Embedding: Results

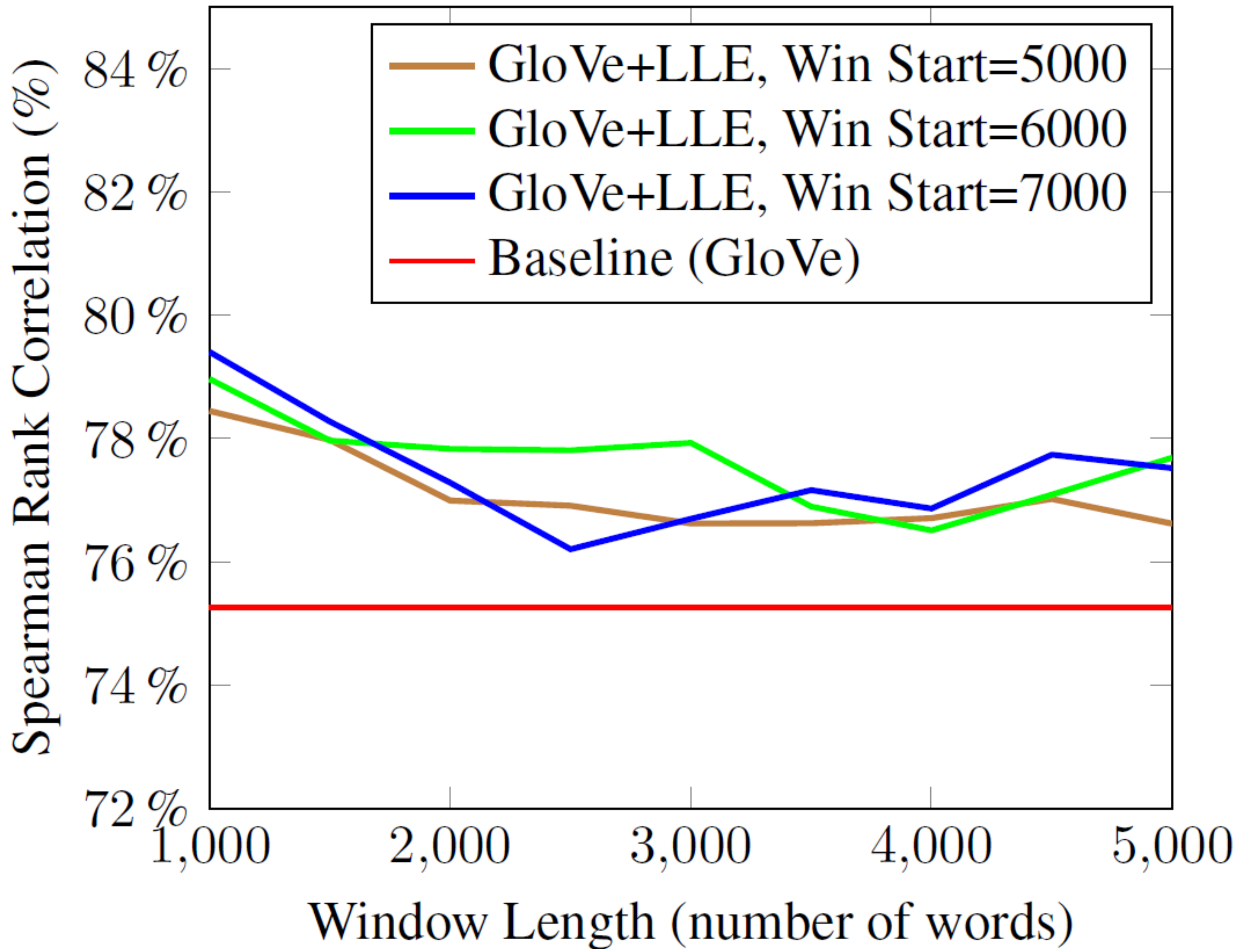
Space	Task	GloVe	Re-Embedding
6B 50d	WS353	**61.2**	56.6
6B 50d	RG65	**60.2**	53.0
6B 100d	WS353	**64.5**	64.3
6B 100d	RG65	65.3	**67.3**
6B 200d	WS353	68.5	**69.7**
6B 200d	RG65	75.5	**76.0**
6B 300d	WS353	65.8	**70.3**
6B 300d	RG65	75.5	**80.5**
42B 300d	WS353	75.2	**78.4**
42B 300d	RG65	80.0	**83.4**

Table 1: Average performance on similarity tasks. (Window start $\in [5000, 15000]$, Number of LLE local neighbours =1000, Window length = 1001, Manifold dimensionality = Space dimensionality.)

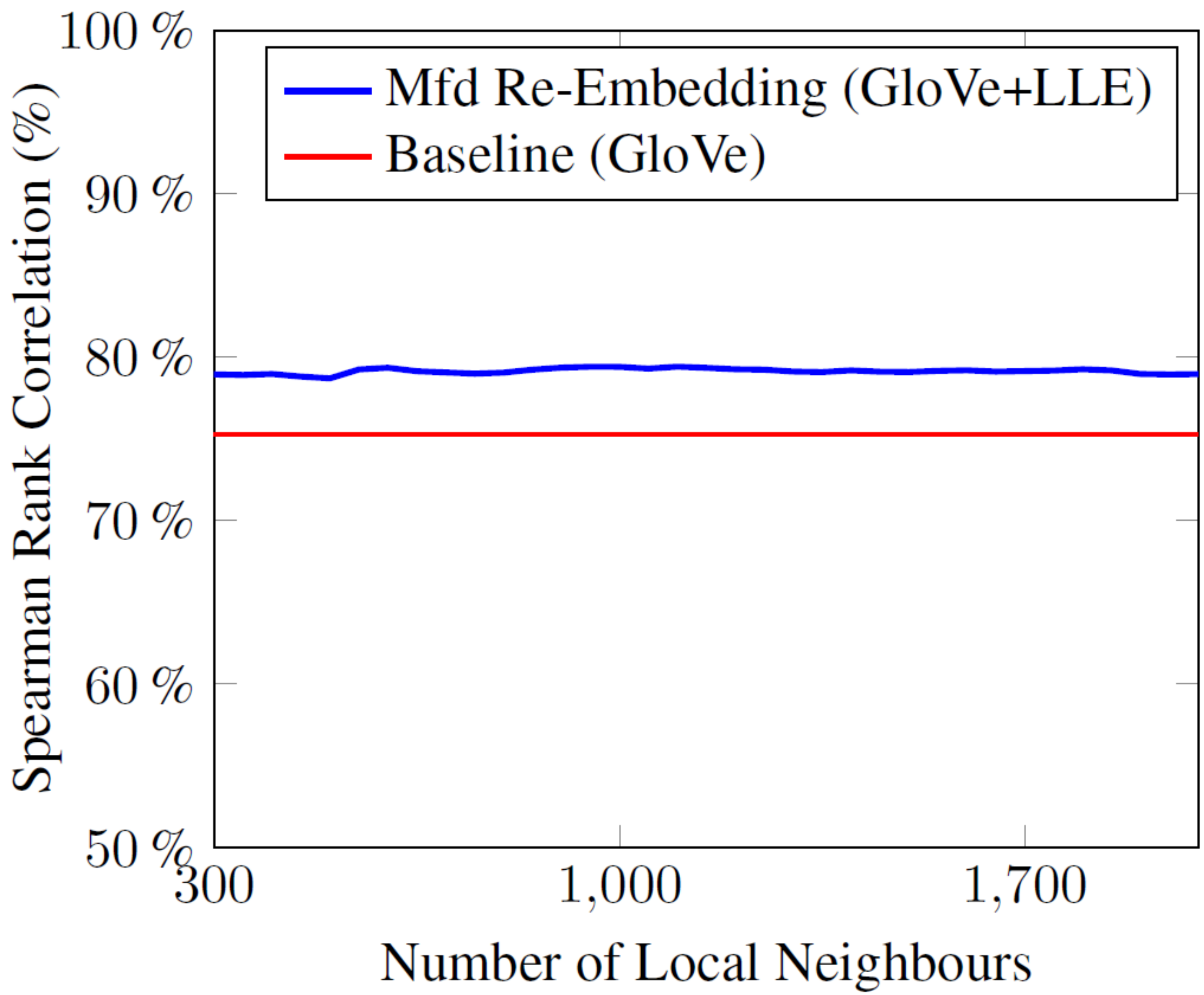
Word Re-Embedding: Results



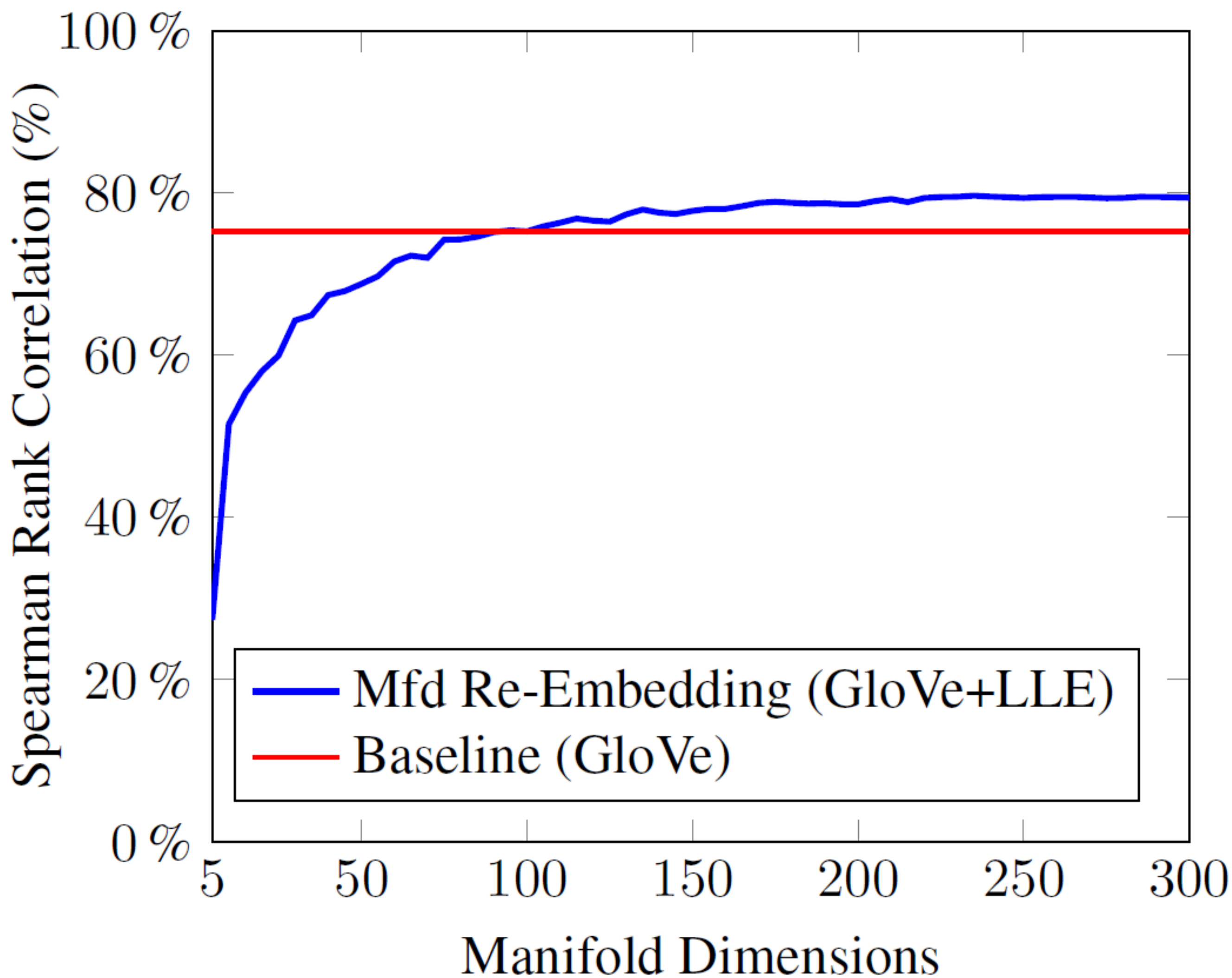
Word Re-Embedding: Results



Word Re-Embedding: Results



Word Re-Embedding: Results



Conclusions

- Word re-embedding improves performance on word similarity tasks
- The sample window start should be chosen just after the stop words
- The sample length should be close or equal to the number of local neighbours, which in turn can be chosen from a wide range
- The dimensionality of the original embedding space should be retained