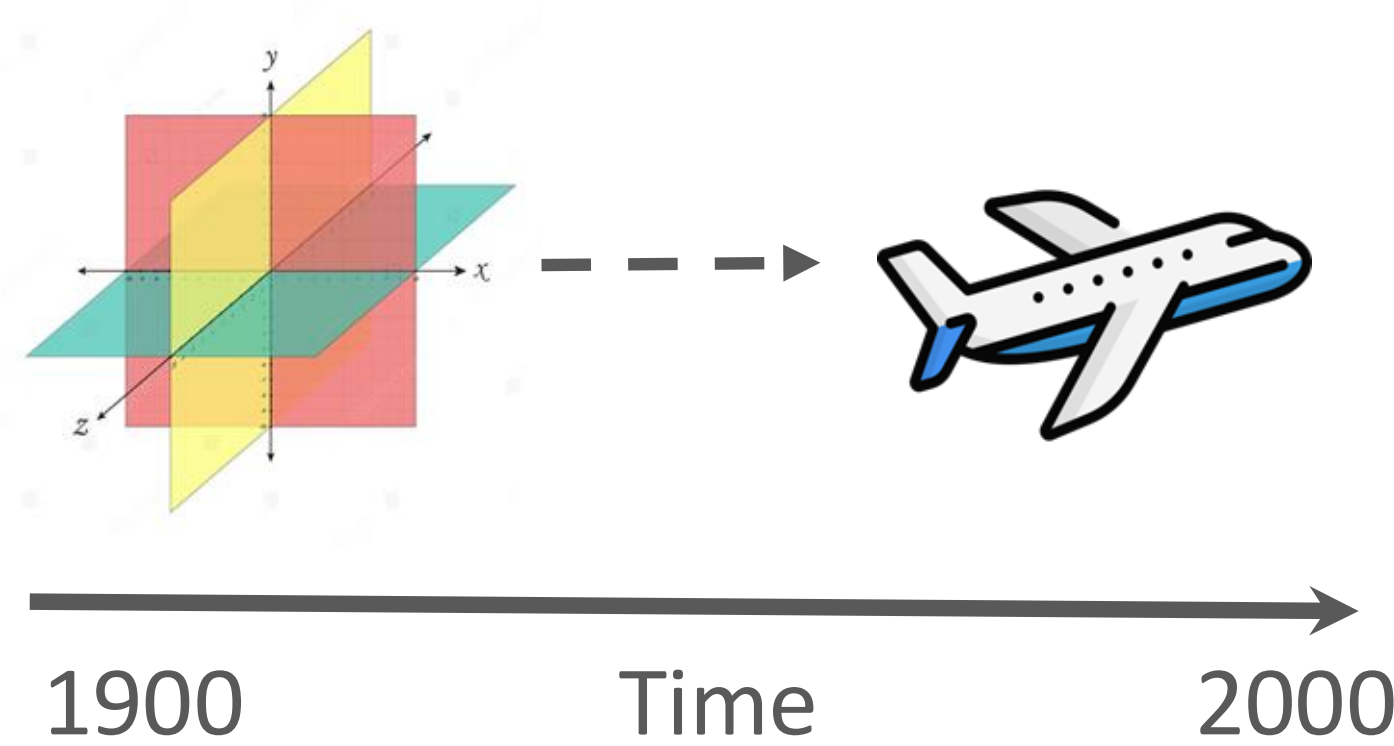


# Rethinking Metrics for Lexical Semantic Change Detection

Roksana Goworek Queen Mary University of London

Haim Dubossarsky QMUL · Alan Turing Institute · University of Cambridge



## LEXICAL SEMANTIC CHANGE DETECTION

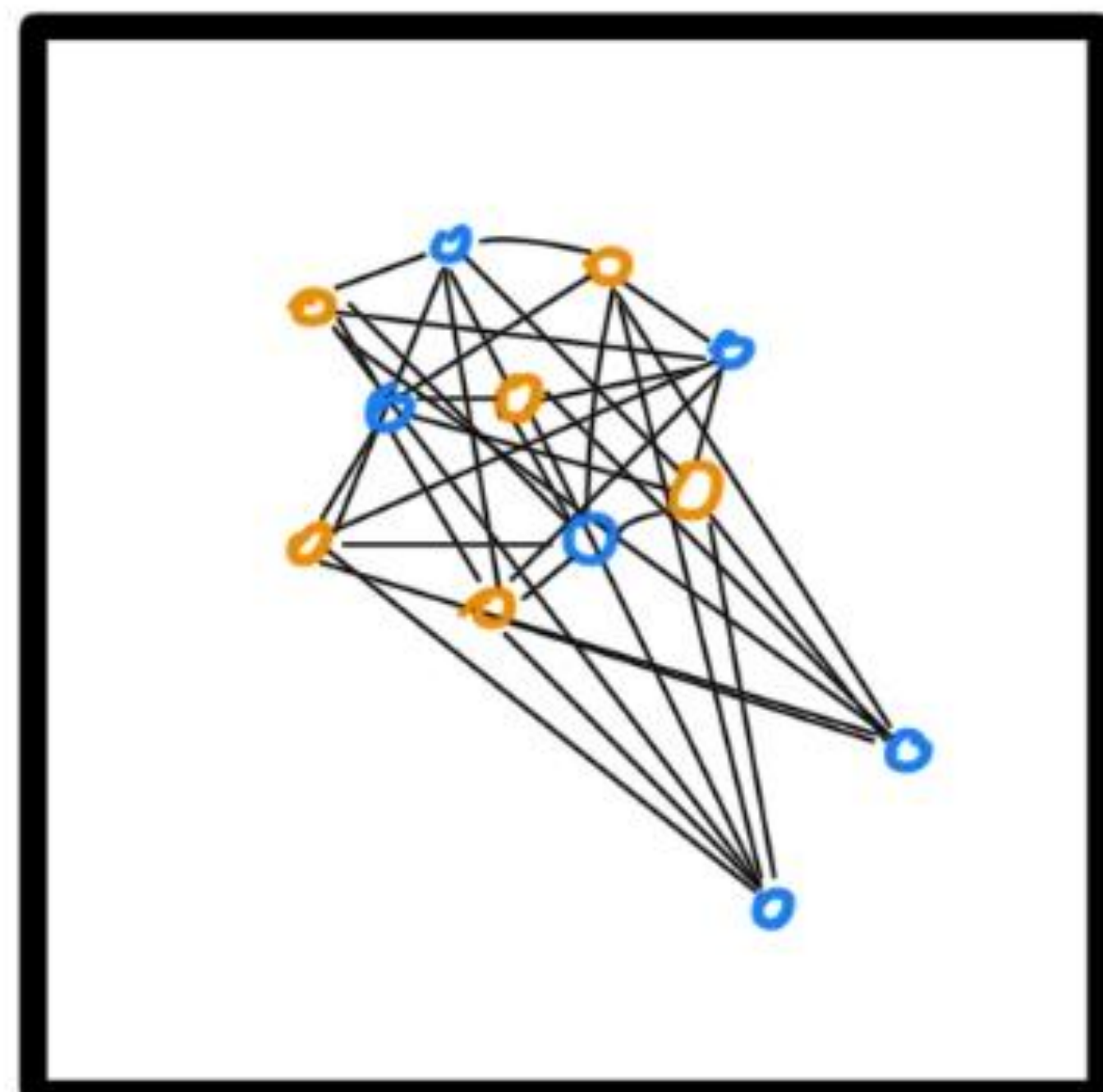
Words change meaning over time.  
A word's meaning = its contexts.

Language models encode each occurrence as a vector. Two time periods → two sets of embeddings.  
Measure the distance between the sets.

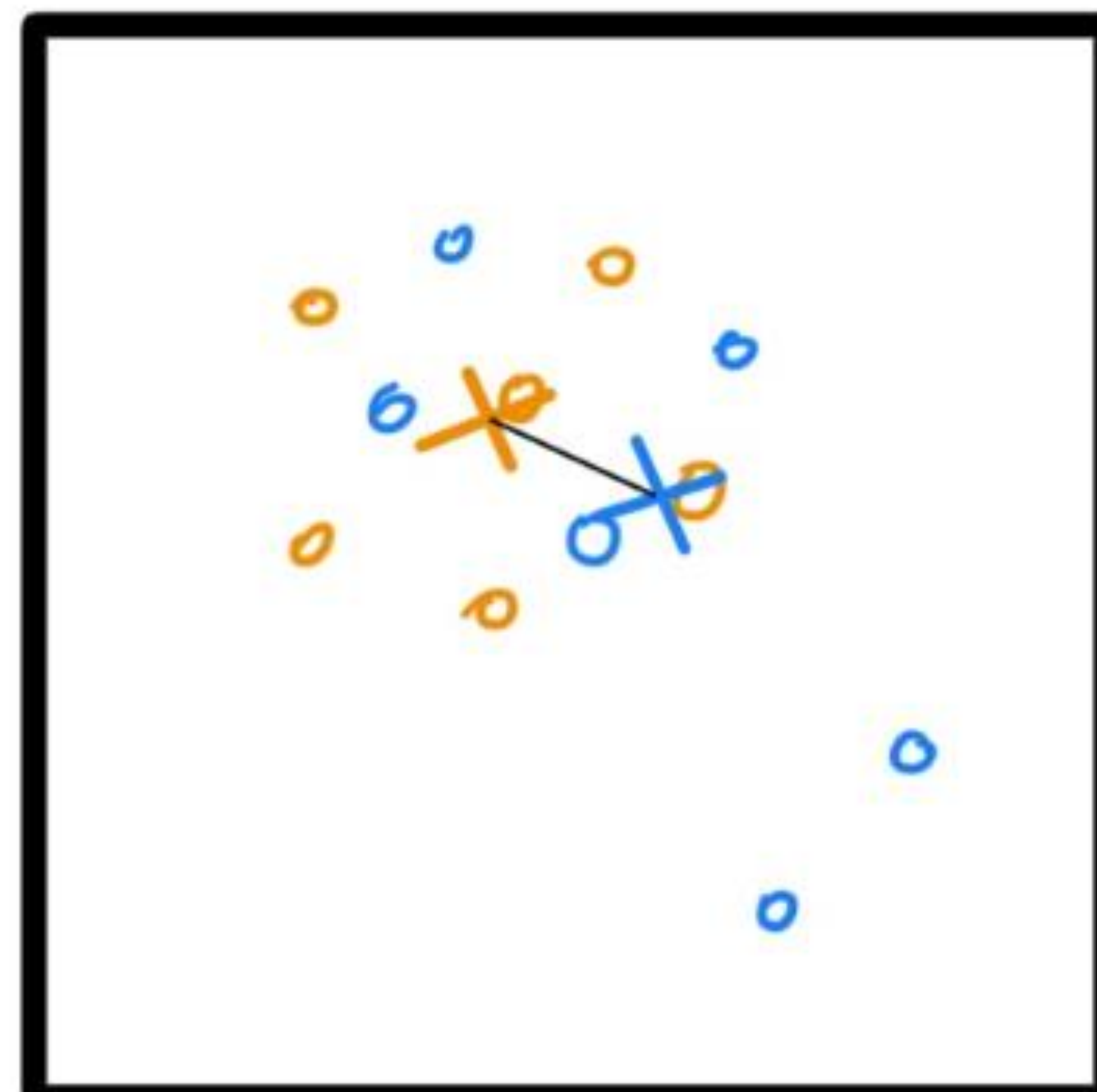
We introduce two new metrics that outperform existing ones and stay robust when embeddings are noisy or compressed.

## PROBLEM

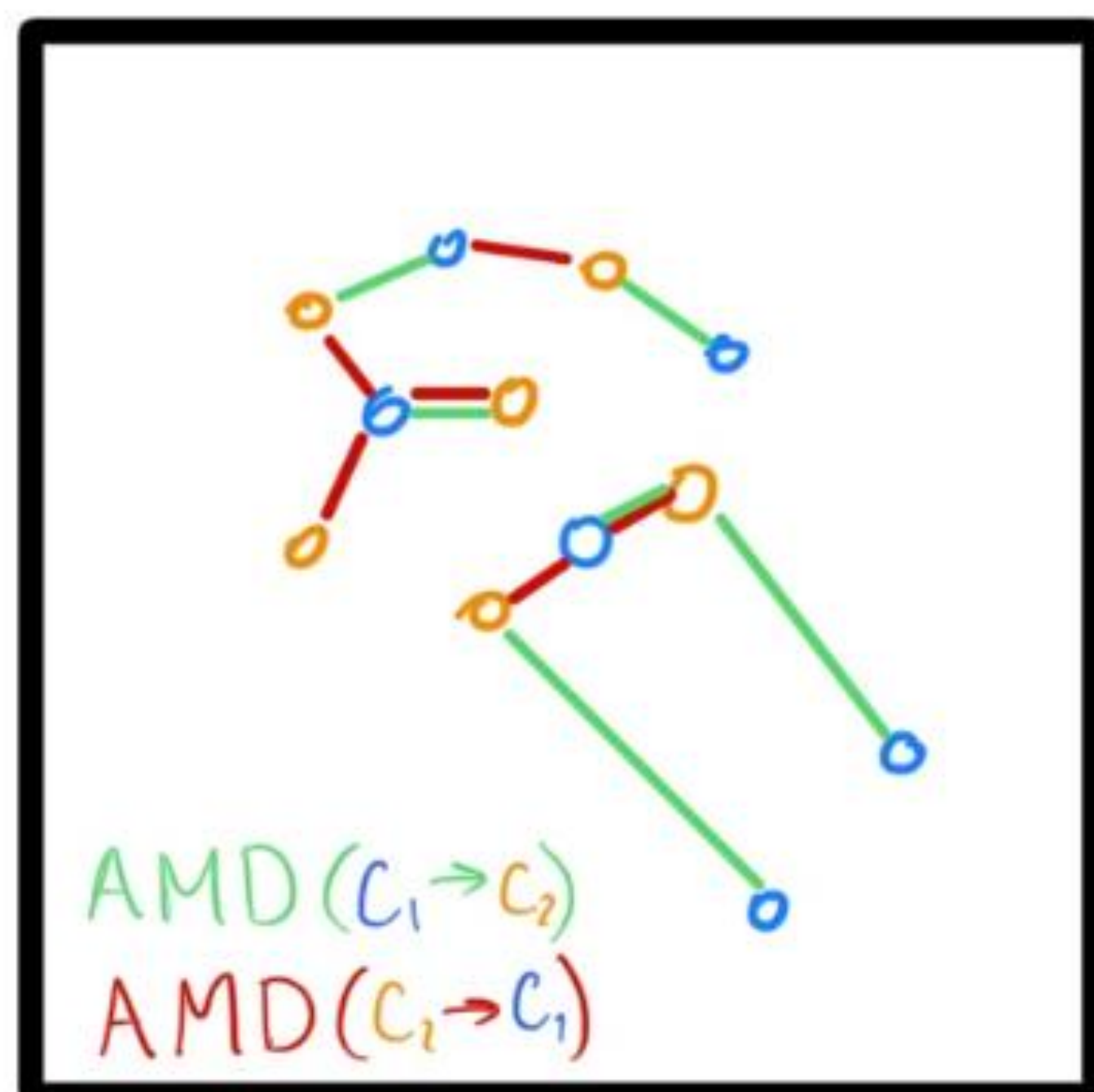
Global aggregation, such as APD or PRT, measures the average shift of embeddings, often losing signals of local correspondence between the two corpora, or drowning out outlier usages, which is where semantic change actually shows up.



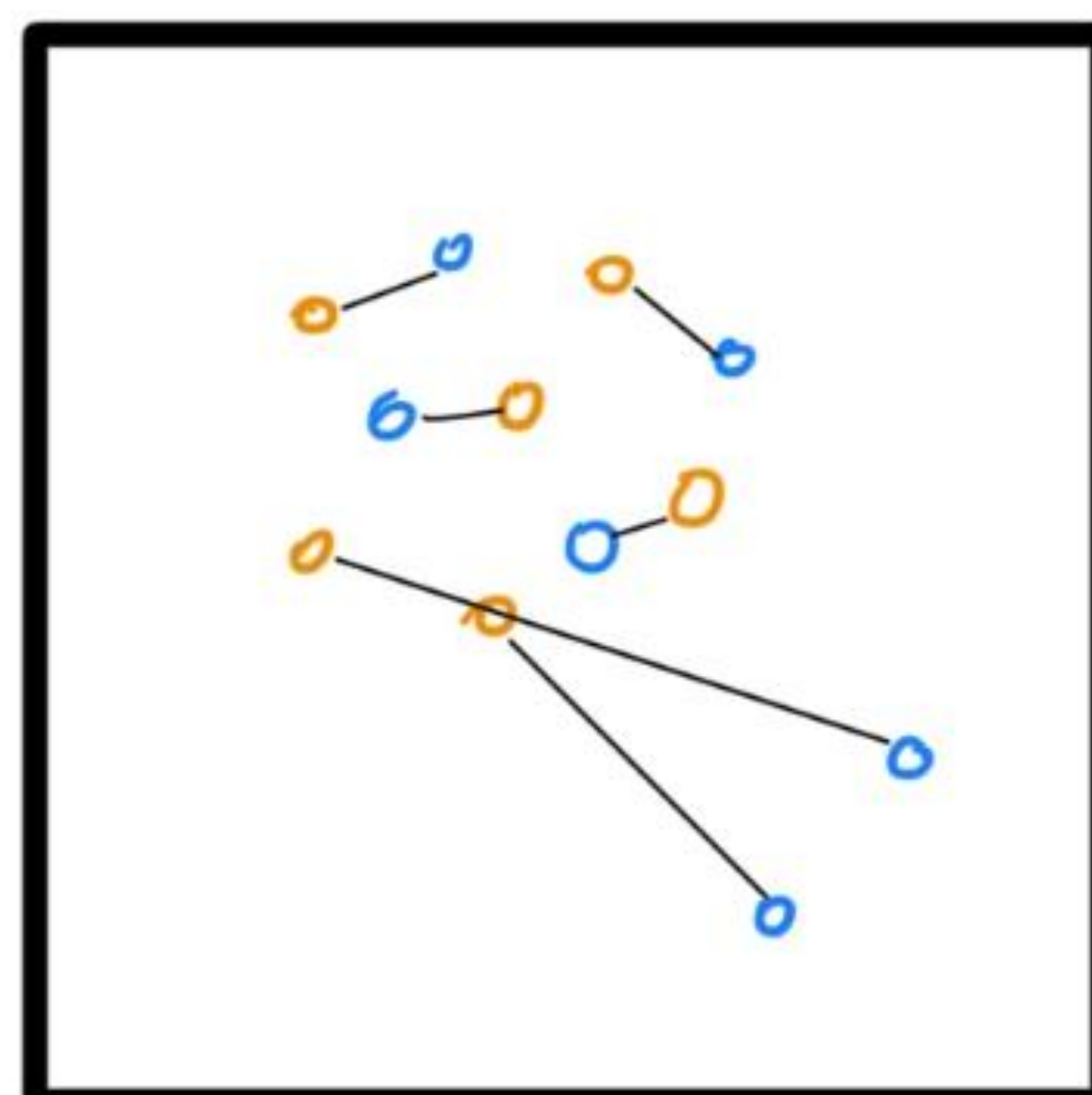
(a) APD: average distance over all cross-corpus usage pairs.



(b) PRT: cosine distance between corpus centroids.



(c) AMD: nearest cross-corpus links ( $C_1 \rightarrow C_2$  in green,  $C_2 \rightarrow C_1$  in red).



(d) SAMD: greedy one-to-one matching between cross-corpus usages.

## AMD • Average Minimum Distance

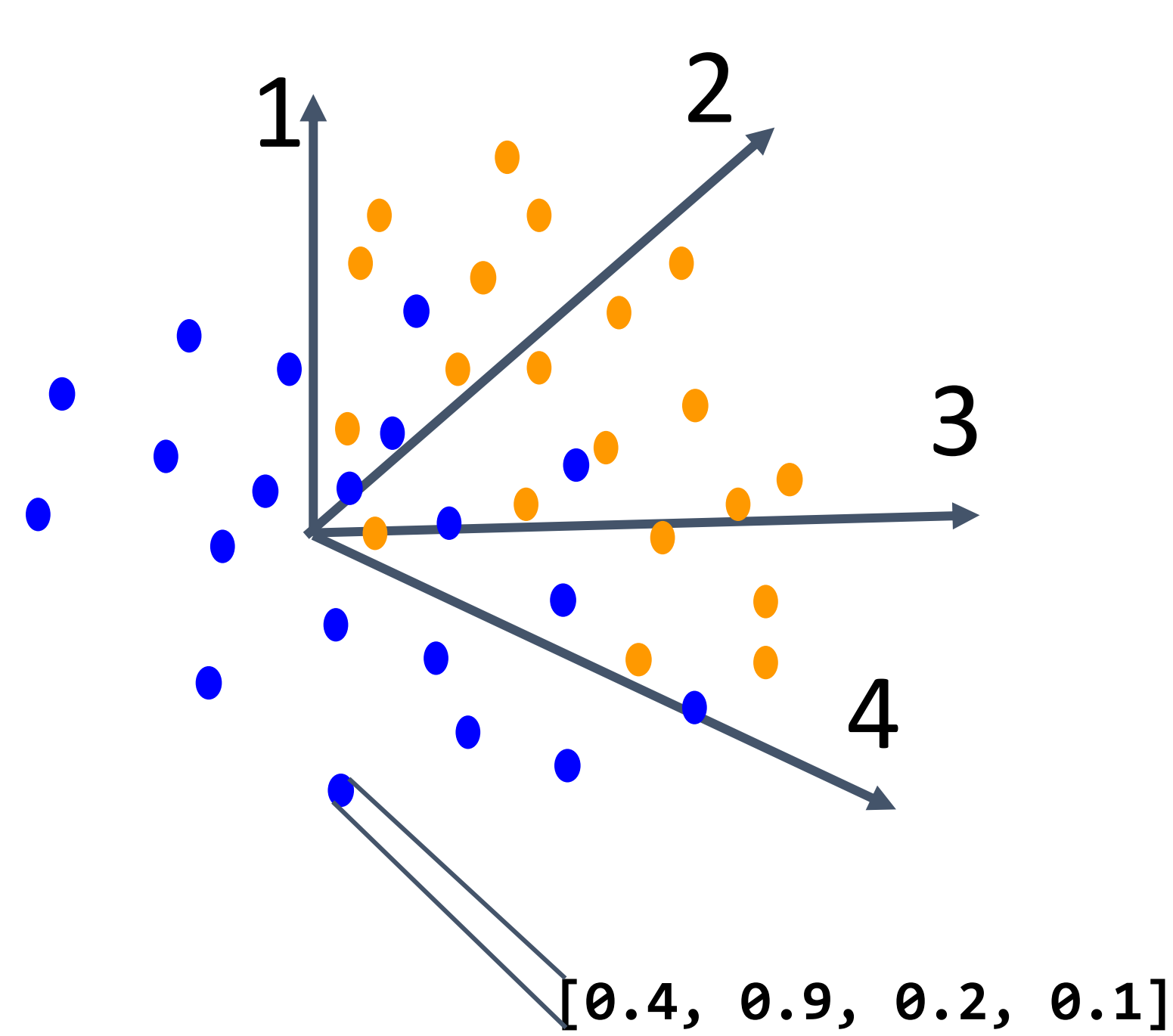
For each usage in one corpus, AMD finds its single closest counterpart in the other corpus and averages these minimum distances in both directions.

If meaning is stable, usages should find good matches across time; if it has shifted, usages will not have similar usages in the other corpus. AMD is sensitive to changes affecting only a subset of usages, where global metrics like APD or PRT, dominated by the bulk of the distribution, could miss them entirely.

## SAMD • Symmetric AMD

AMD with one extra constraint: each usage can only be matched once. Pairs are claimed greedily, closest first, until all usages are assigned. This prevents hubness, where a few embeddings monopolise nearest-neighbour slots, deflating the measured distance. The result is a single, symmetrical metric for one-to-one usage correspondence across time periods.

## DEFINITION SPACE



Plane:

- Plane: A flat surface on which a straight line joining any two points would wholly lie.
- Plane: A powered flying vehicle with fixed wings and one or more engines; an airplane.
- Plane: A level of existence, thought, or development.
- Plane: A tool consisting of a block with a projecting blade, for shaping and smoothing a wooden surface.

## DEFINITION SPACE

An interpretable low-dimensional projection



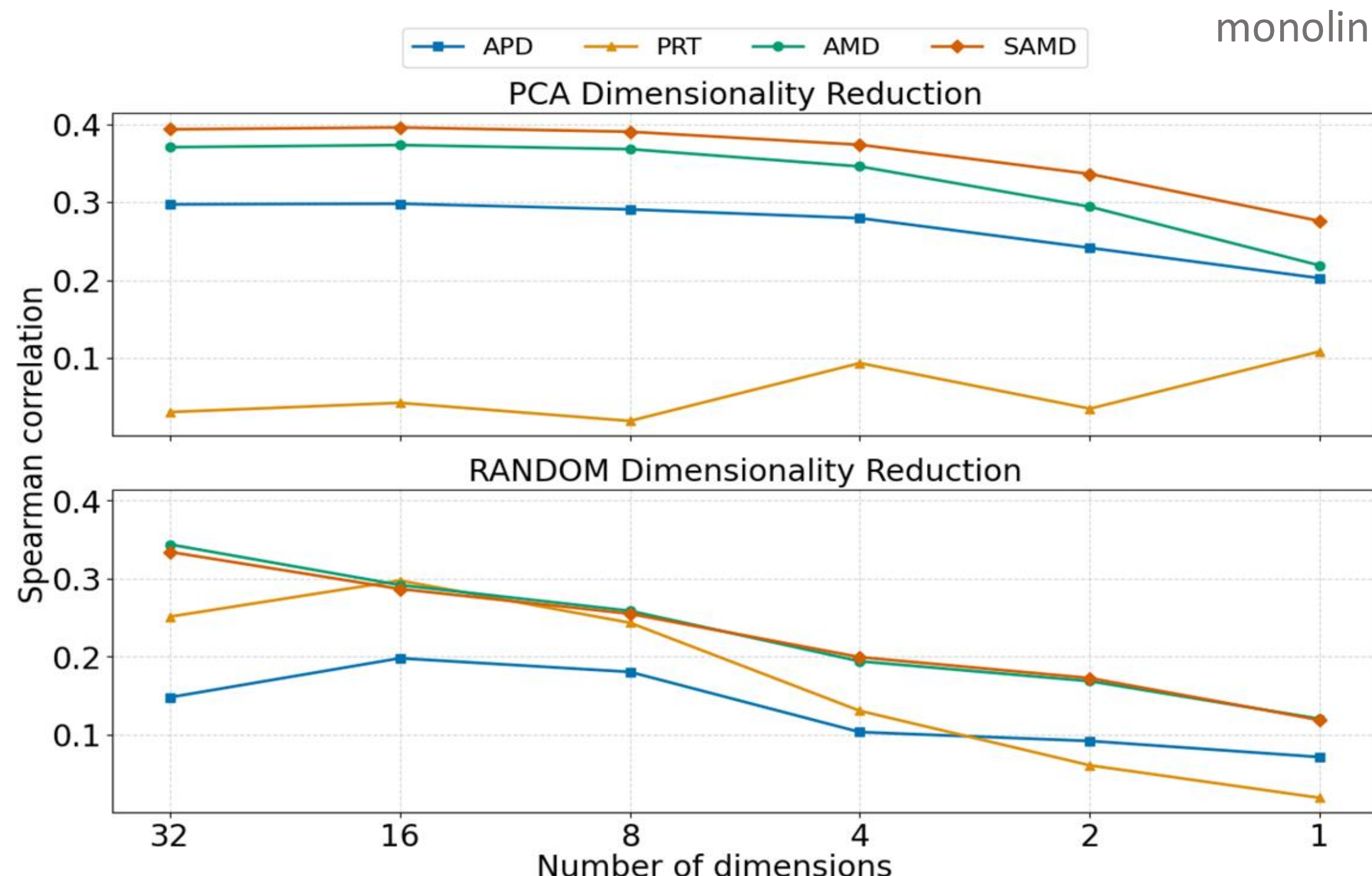
Lexicographer provides definitions for words of interest.

## RESULTS

Metric	FULL	DEF	PCA	RAND
APD	.405 (.228)	.353 (.246)	.190 (.265)	.118 (.225)
PRT	.395 (.196)	.346 (.236)	.093 (.197)	.257 (.190)
AMD	.442 (.176)	.519 (.172)	.542 (.119)	.480 (.126)
SAMD	.444 (.184)	.443 (.167)	.550 (.141)	.464 (.129)

AMD and SAMD outperform existing LSCD metrics and are enhanced by projections to Definition space (DEF) or PCA. Standard metrics collapse under dimensionality reduction. Results broken down by language and model type (specialised XL-LEXEME, monolingual, and multilingual encoders) are available in the paper and code.

## DIMENSIONALITY REDUCTION STRESS TEST



## Key References

- Kutuzov & Giulianelli (2020). UiO-UvA at SemEval-2020 Task 1: Contextualised embeddings for lexical semantic change detection. *SemEval @ COLING 2020*
- Schlechtweg et al. (2020). SemEval-2020 Task 1: Unsupervised lexical semantic change detection. *SemEval 2020*.
- Cassotti et al. (2023). XL-LEXEME: WiC pretrained model for cross-lingual lexical semantic change. *ACL 2023*.



Paper



Code