

Multi-Prototype Models of Word Meaning

Joseph Reisinger and Raymond J. Mooney
The University of Texas at Austin

Vector Space Lexical Semantics

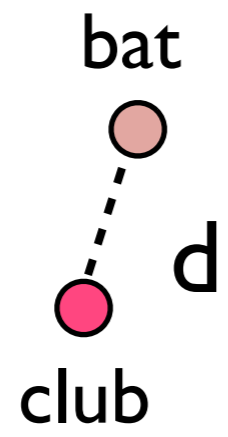
- Represent “meaning” as a point in some high-dimensional space
- Word relatedness correlates with some distance metric
- Attributional: Almuhareb and Poesio (2004), Bullinaria and Levy (2007), Erk (2007), Griffiths et al. (2007), Landauer and Dumais (1997), Padó and Lapata (2007), Sahlgren (2006), Schütze (1997)
- Relational: Moldovan (2006), Pantel and Pennacchiotti (2006), Turney (2006)

$\Omega =$

bat
club



$\Omega =$

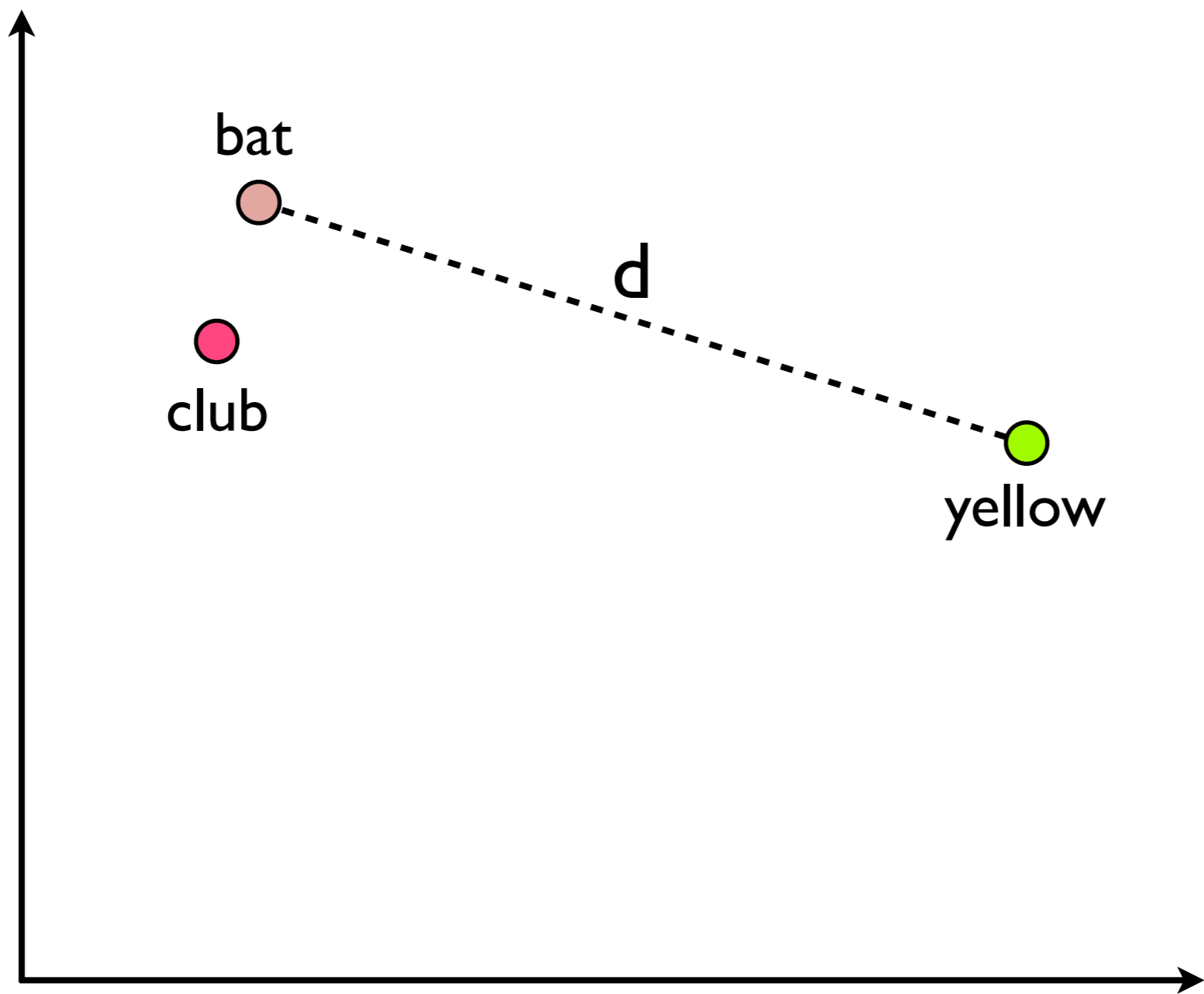


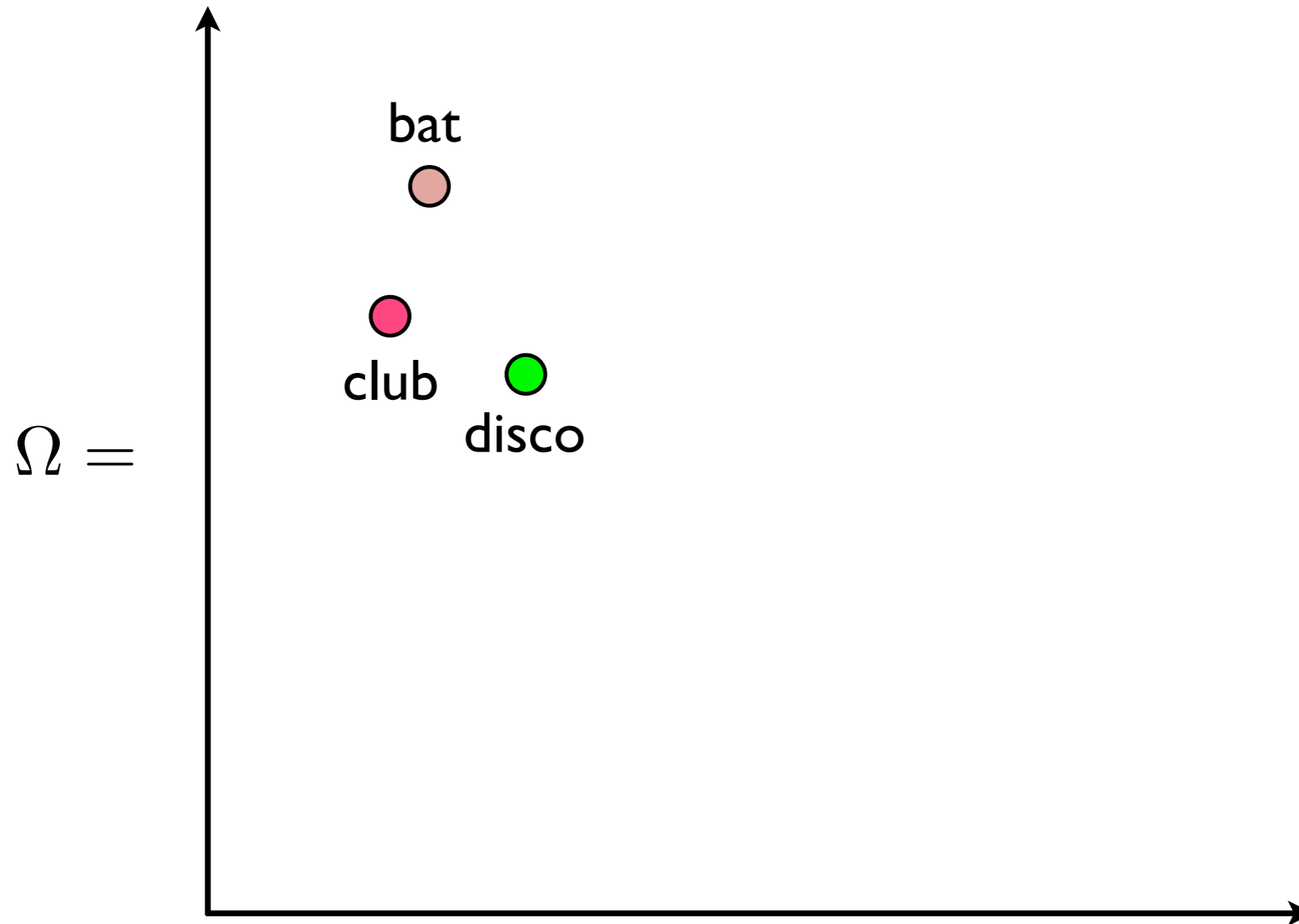
$\Omega =$

bat
club

yellow

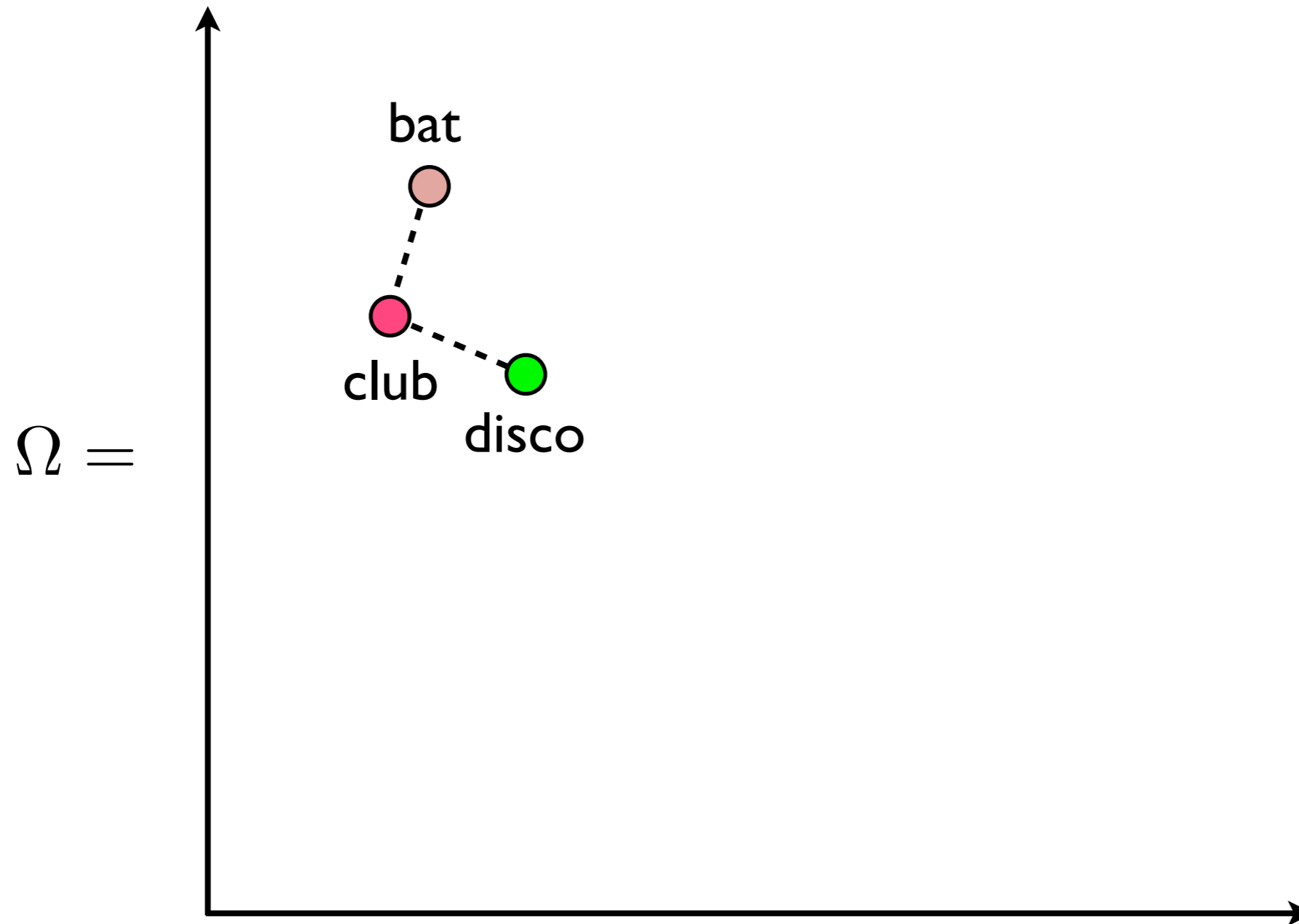
$\Omega =$





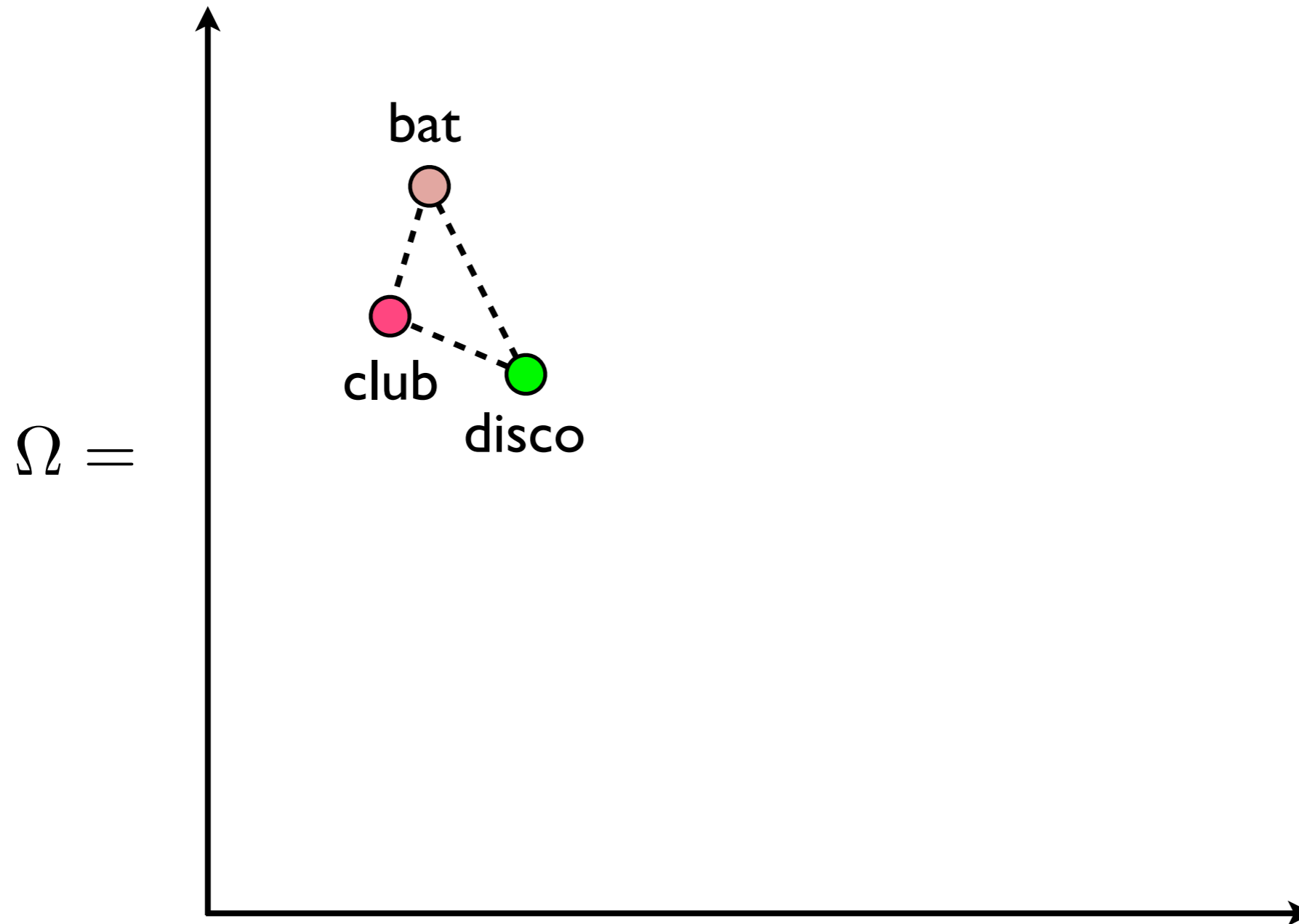
- Any inner product space; e.g. “dense” semantic spaces like LSA

Tversky and Gati (1982), Griffiths et al. (2007)



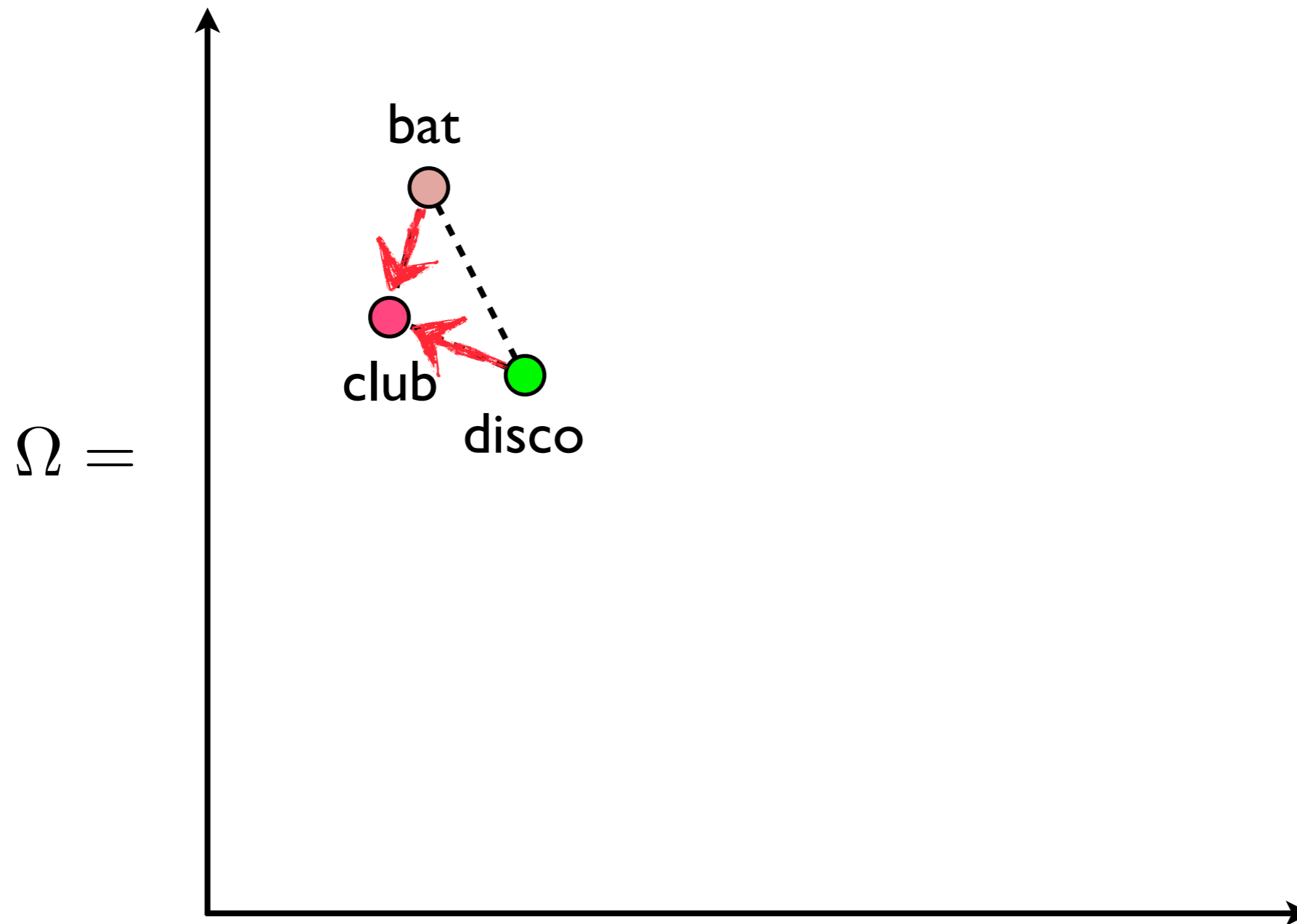
- Any inner product space; e.g. “dense” semantic spaces like LSA

Tversky and Gati (1982), Griffiths et al. (2007)



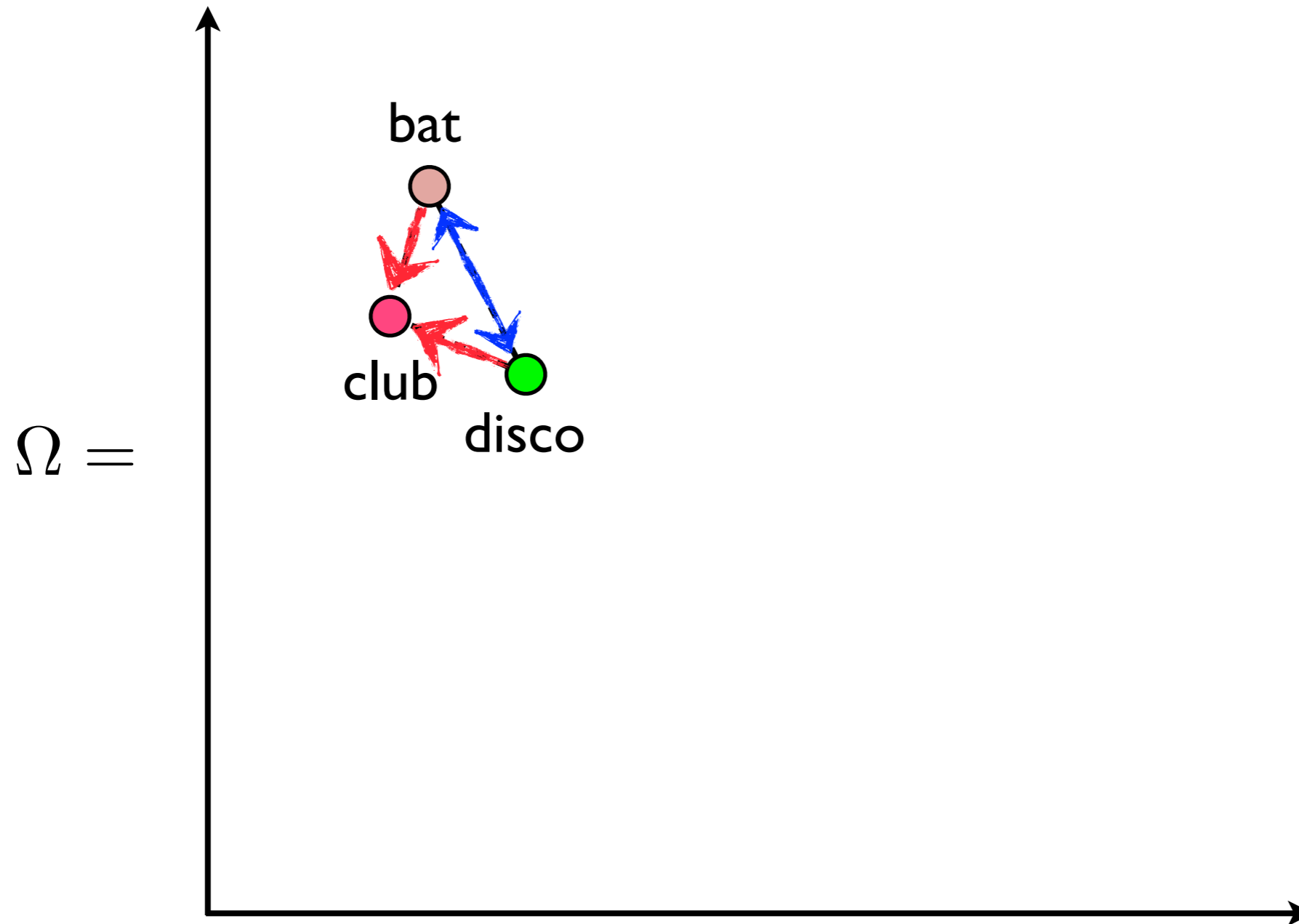
- Any inner product space; e.g. “dense” semantic spaces like LSA

Tversky and Gati (1982), Griffiths et al. (2007)



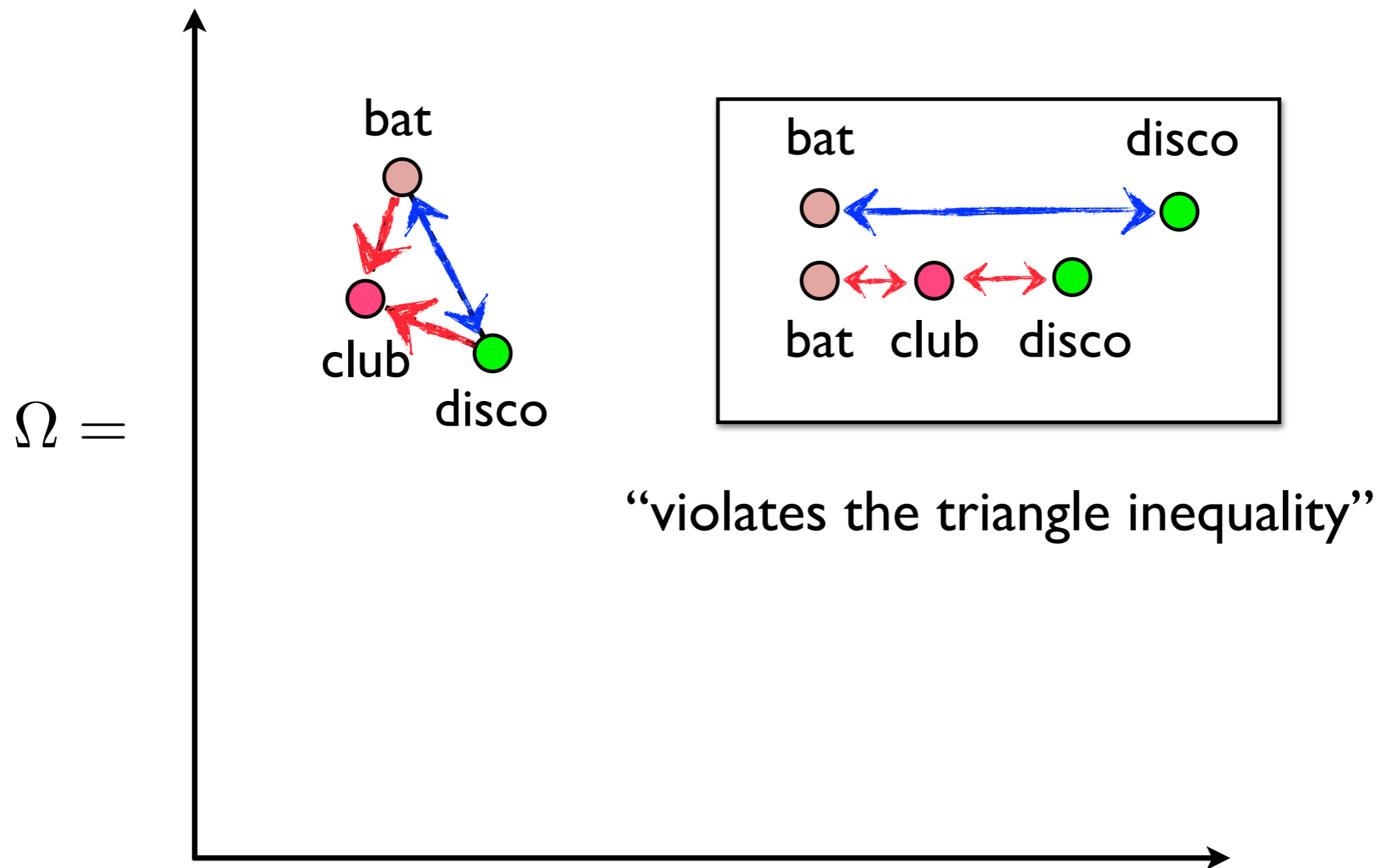
- Any inner product space; e.g. “dense” semantic spaces like LSA

Tversky and Gati (1982), Griffiths et al. (2007)



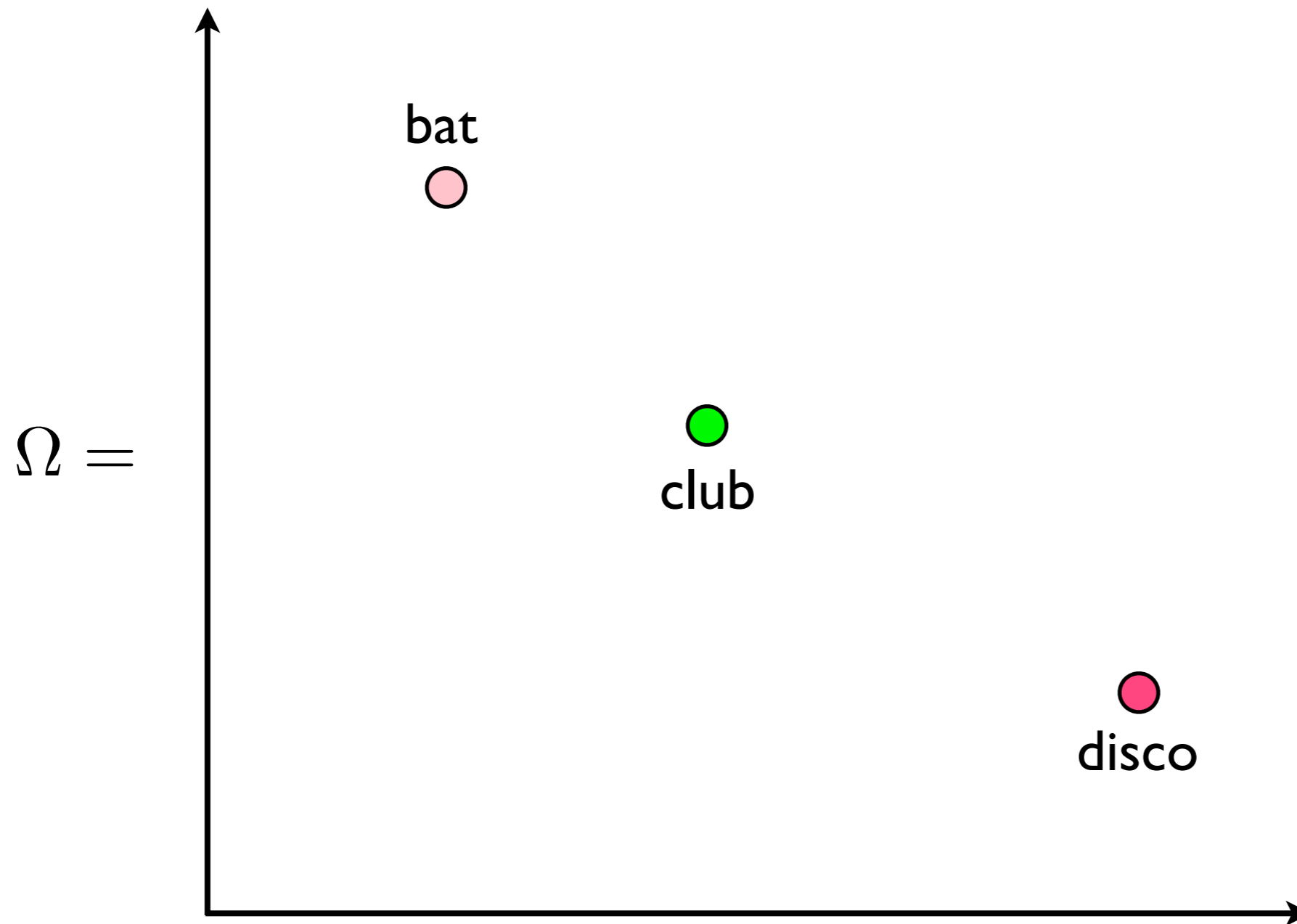
- Any inner product space; e.g. “dense” semantic spaces like LSA

Tversky and Gati (1982), Griffiths et al. (2007)



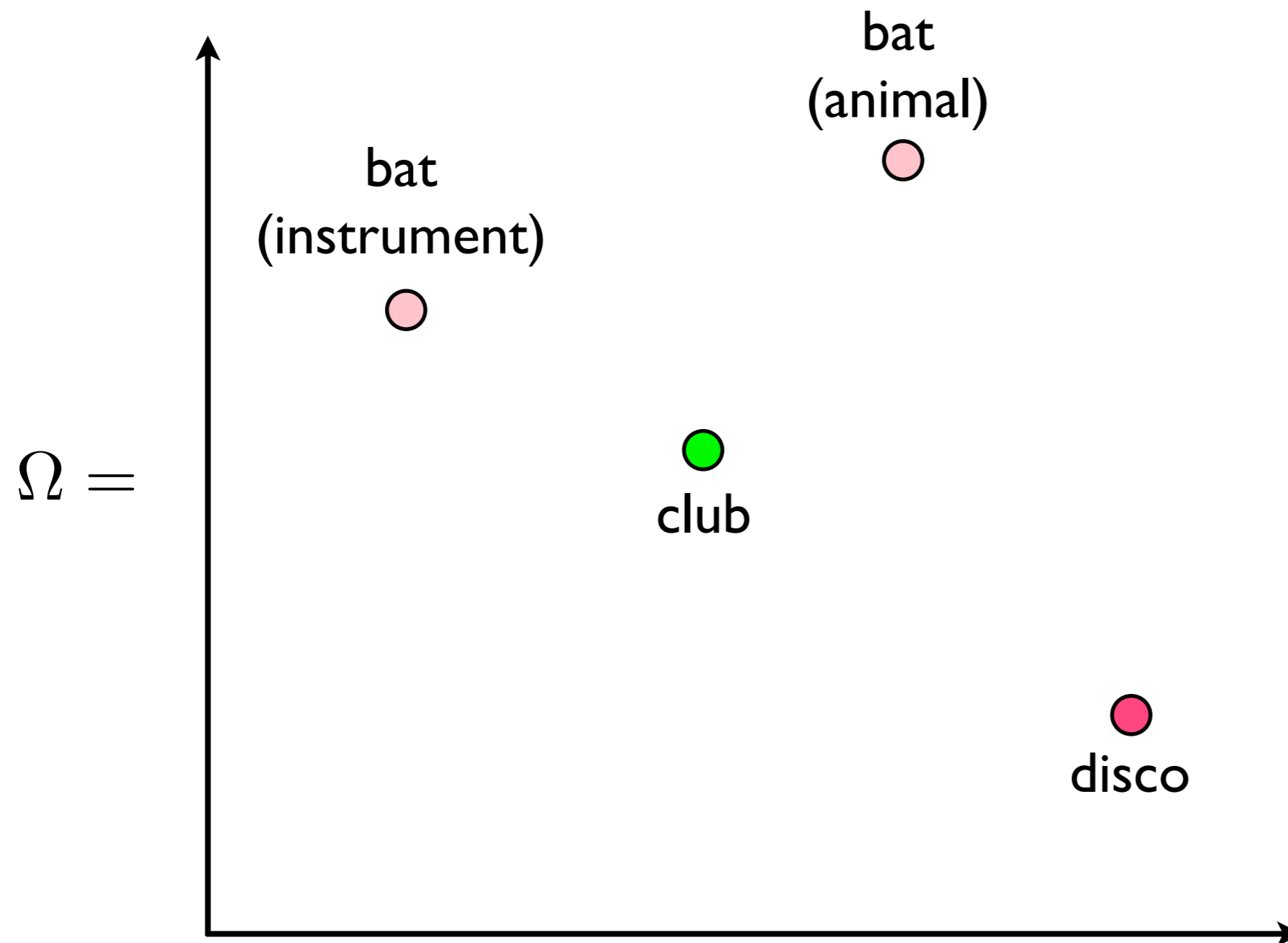
- Any inner product space; e.g. “dense” semantic spaces like LSA

Using multiple prototypes



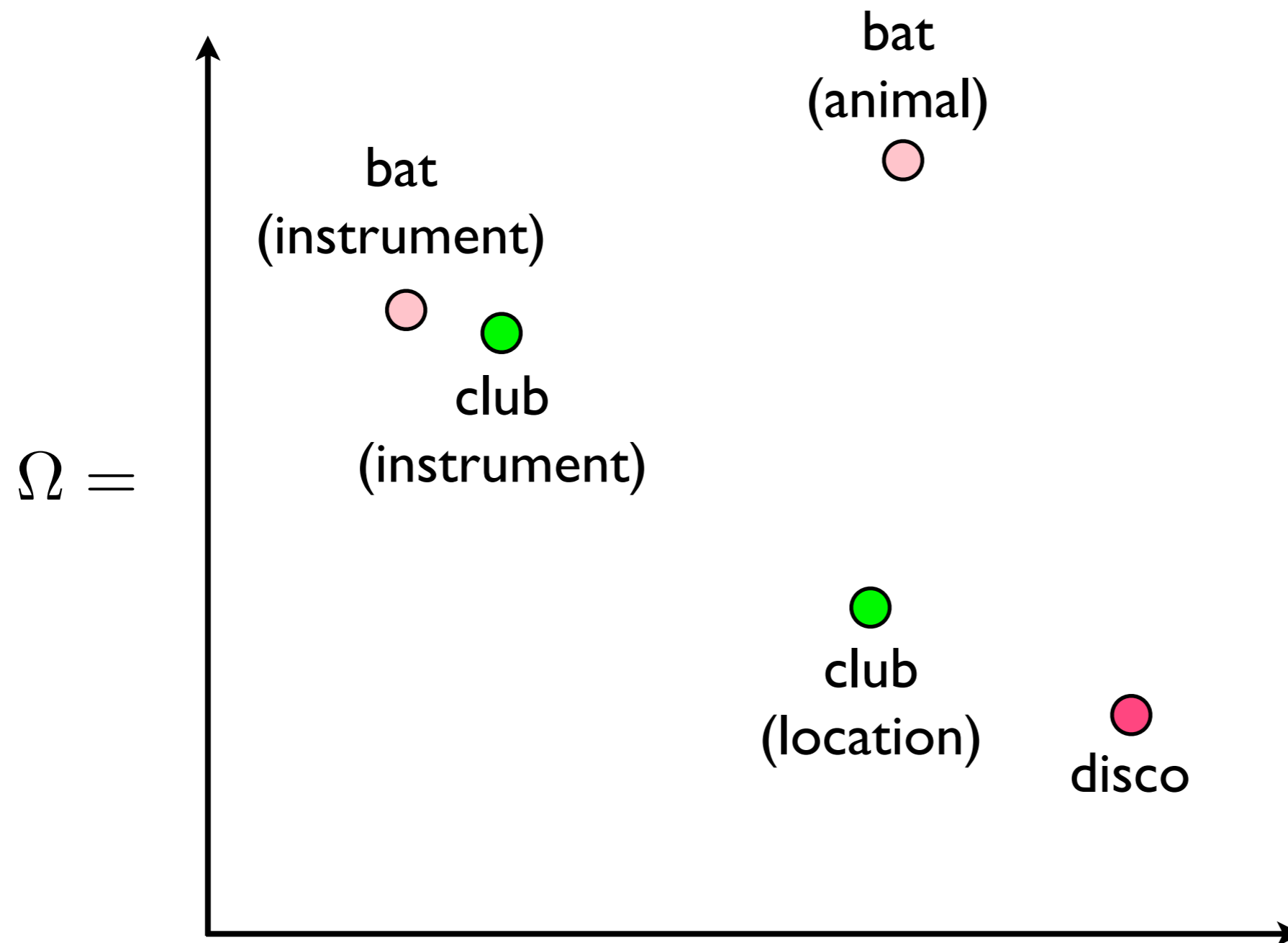
- Similar to unsupervised Word Sense Discovery, e.g. Pantel and Lin (2002), Schütze (1998), Yarowsky (1995)

Using multiple prototypes



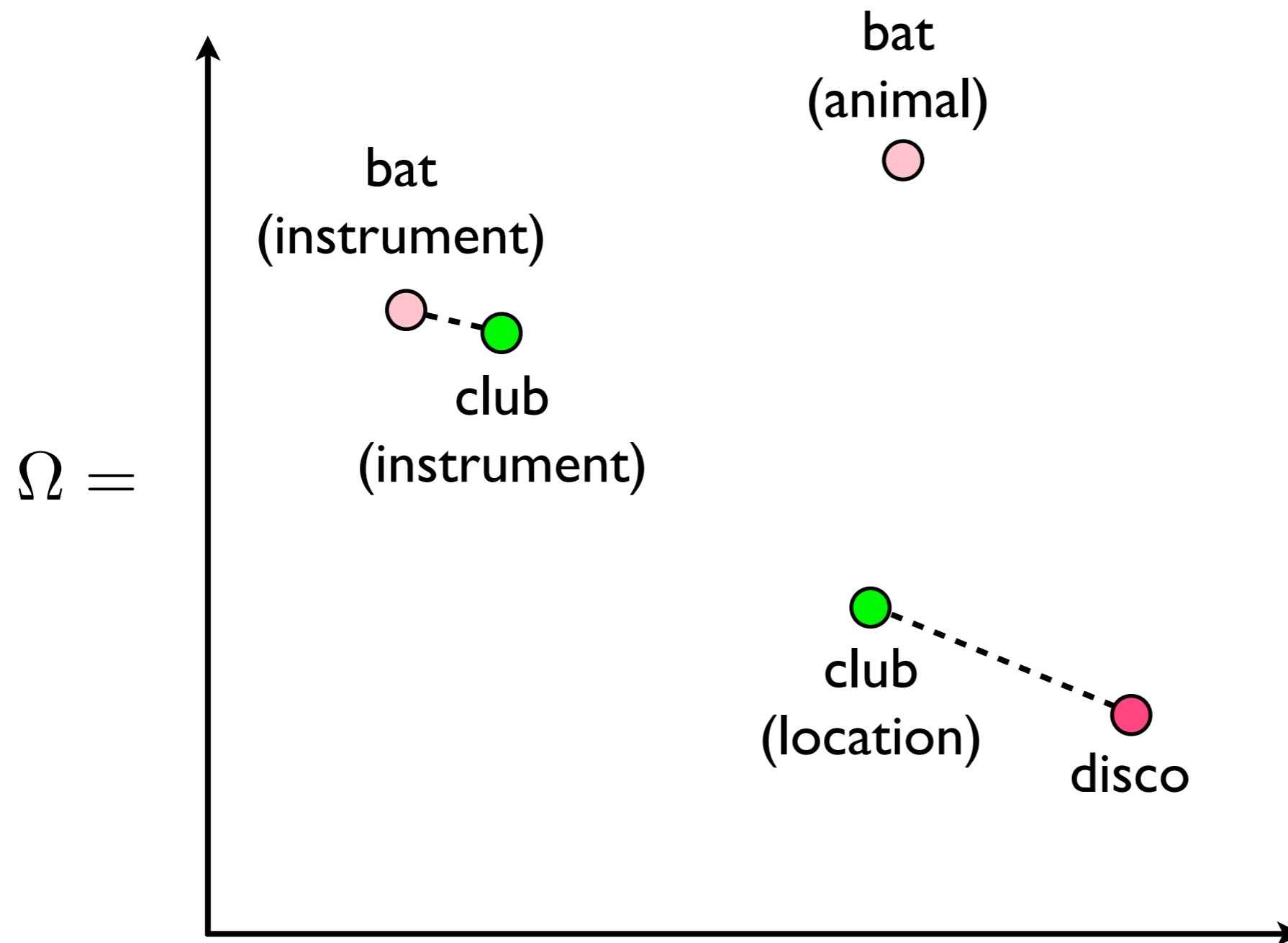
- Similar to unsupervised Word Sense Discovery, e.g. Pantel and Lin (2002), Schütze (1998), Yarowsky (1995)

Using multiple prototypes



- Similar to unsupervised Word Sense Discovery, e.g. Pantel and Lin (2002), Schütze (1998), Yarowsky (1995)

Using multiple prototypes

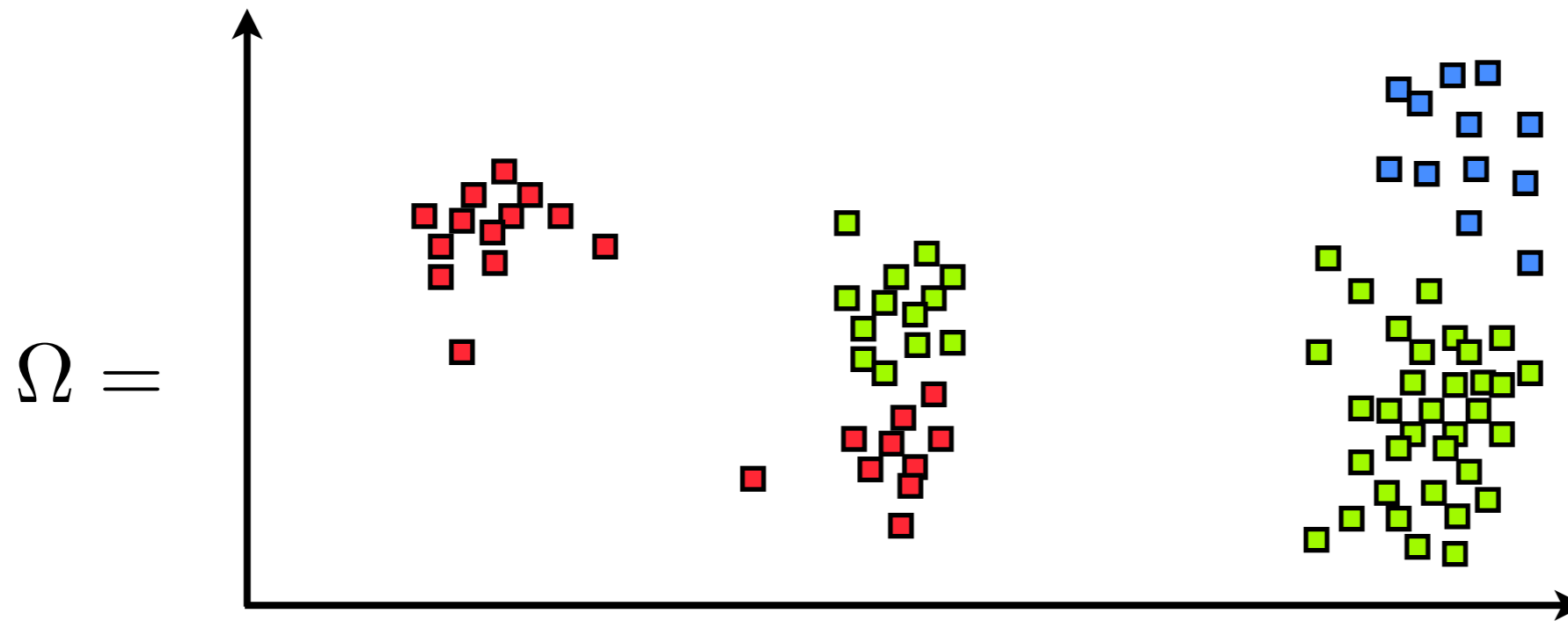


- Similar to unsupervised Word Sense Discovery, e.g. Pantel and Lin (2002), Schütze (1998), Yarowsky (1995)

Some practical benefits

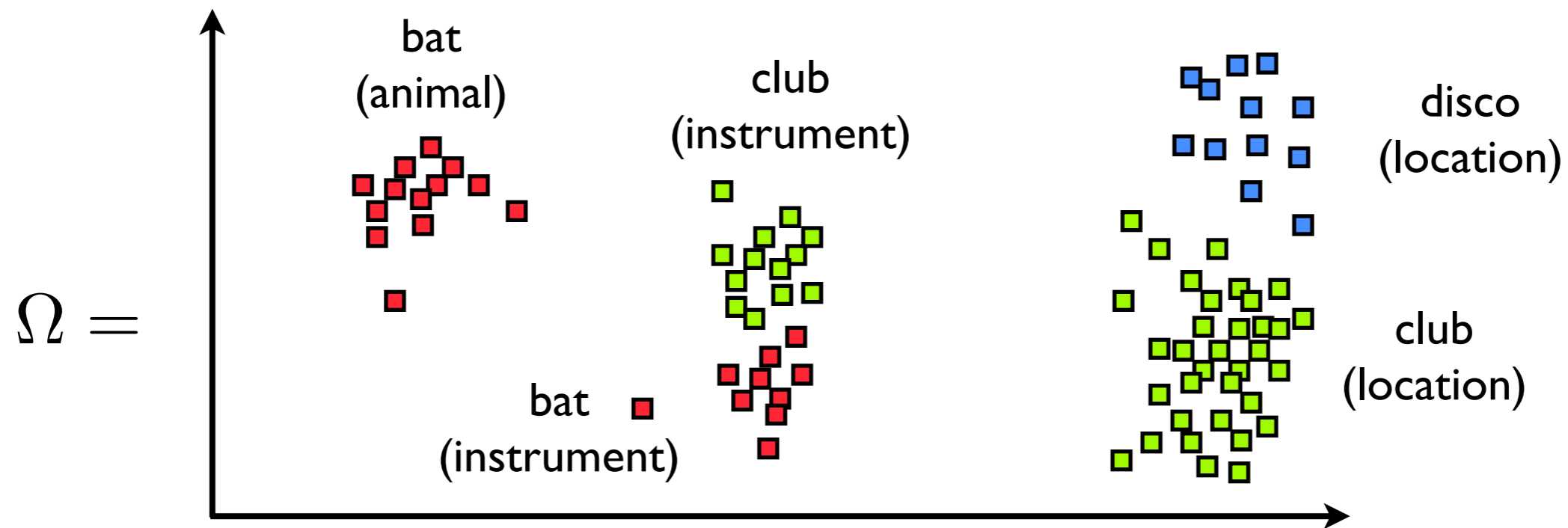
- “Meaning” is a mixture over prototypes, capturing polysemy and thematic variation.
- Can exploit contextual information to refine word similarity computations:
 - e.g., is “the *bat* flew out of the cave” similar to “the girls left the *club*” ?
- “Senses” are thematic and very fine-grained
 - e.g., the *hurricane* sense of *position*

Single Prototype \leftrightarrow Multi-Prototype \leftrightarrow Exemplar



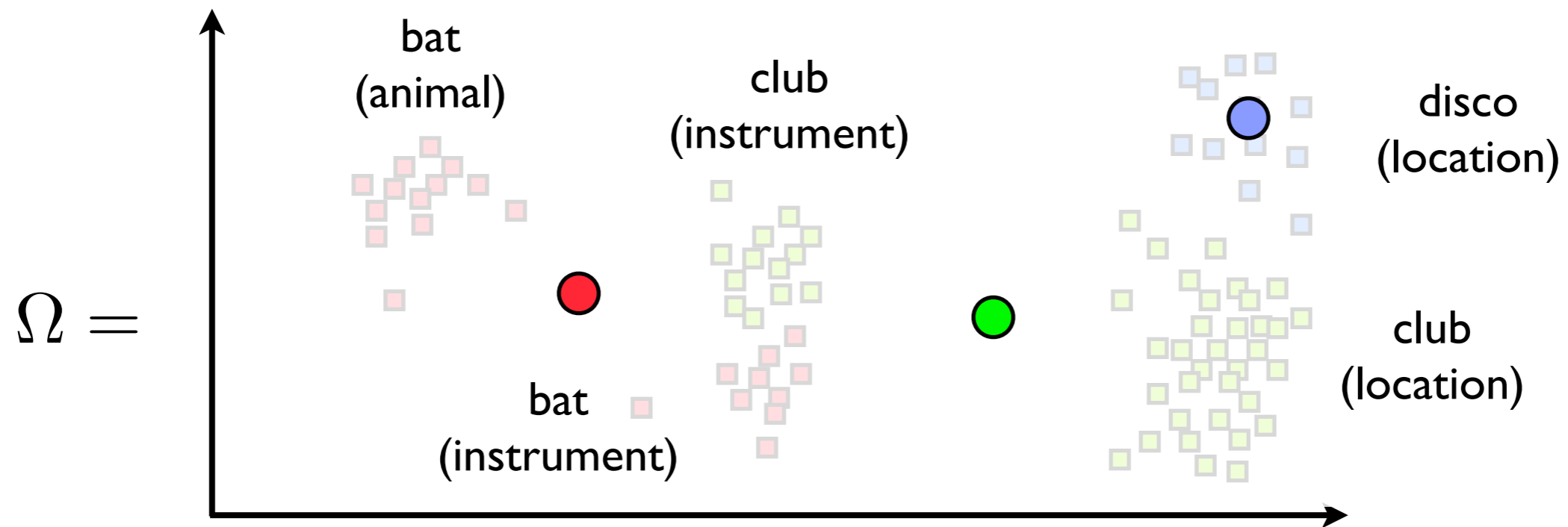
- Find the centroid of the individual word occurrences
- Conflates senses

Single Prototype \leftrightarrow Multi-Prototype \leftrightarrow Exemplar



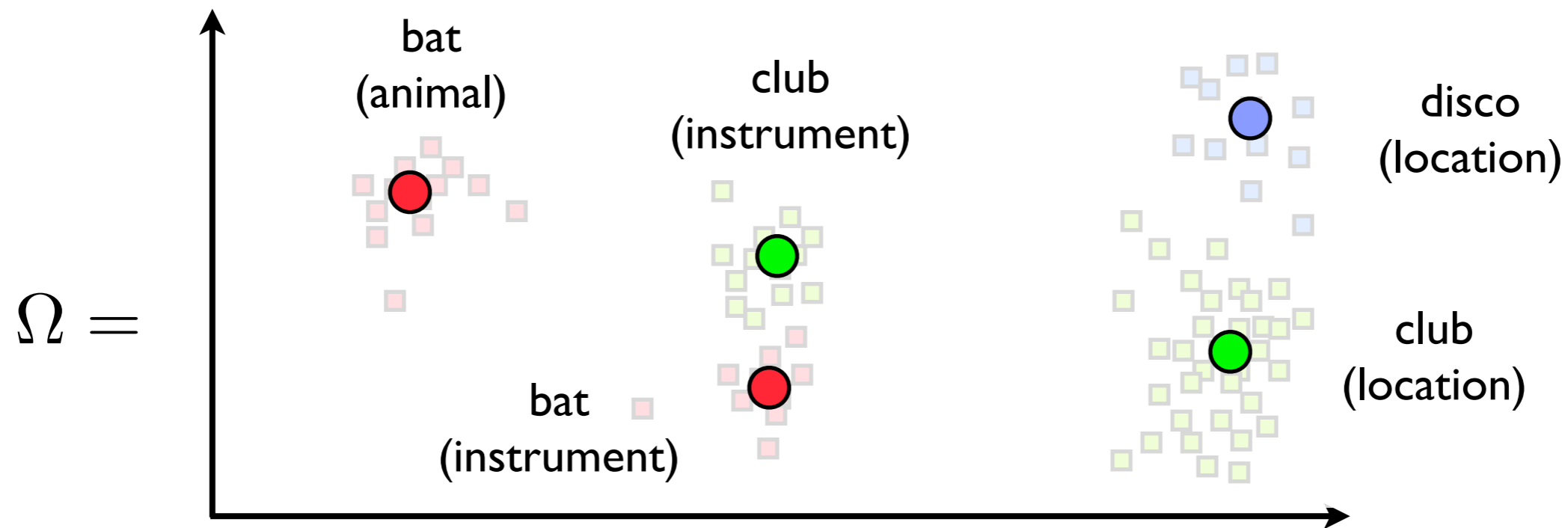
- Find the centroid of the individual word occurrences
- Conflates senses

Single Prototype \leftrightarrow Multi-Prototype \leftrightarrow Exemplar



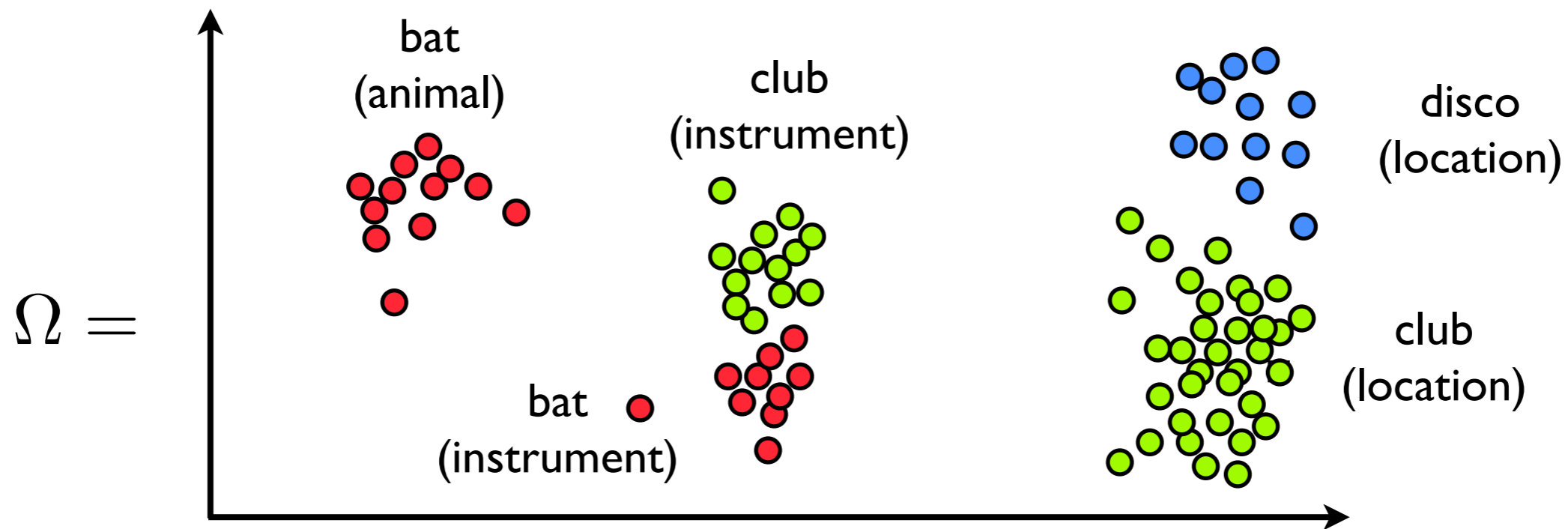
- Find the centroid of the individual word occurrences
- Conflates senses

Single Prototype \leftrightarrow Multi-Prototype \leftrightarrow Exemplar



- Essentially just clustering word occurrences
- Doesn't find lexicographic senses; captures contextual variance directly.

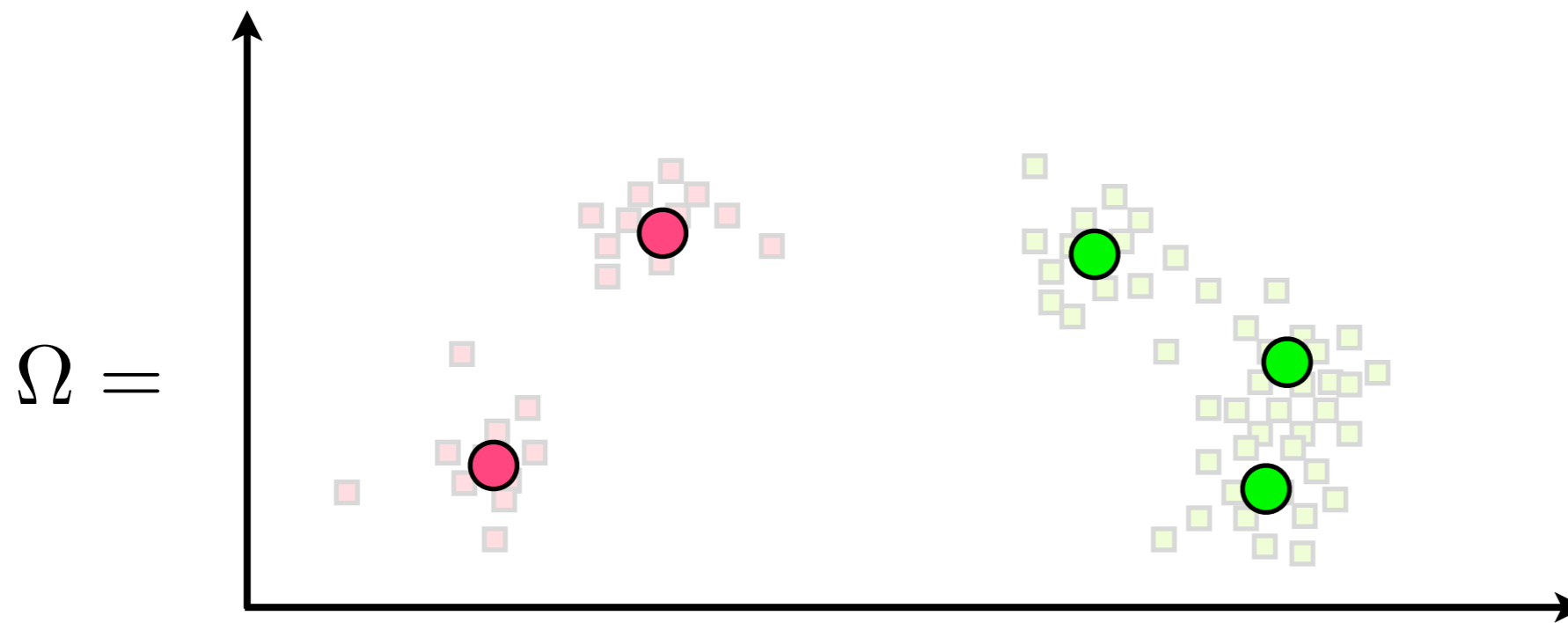
Single Prototype \leftrightarrow Multi-Prototype \leftrightarrow Exemplar



- Just treat all occurrences as an ensemble representing meaning.
- Compute similarity as the average of the K most similar pairs.
- Heavily influenced by noise, but captures more structure

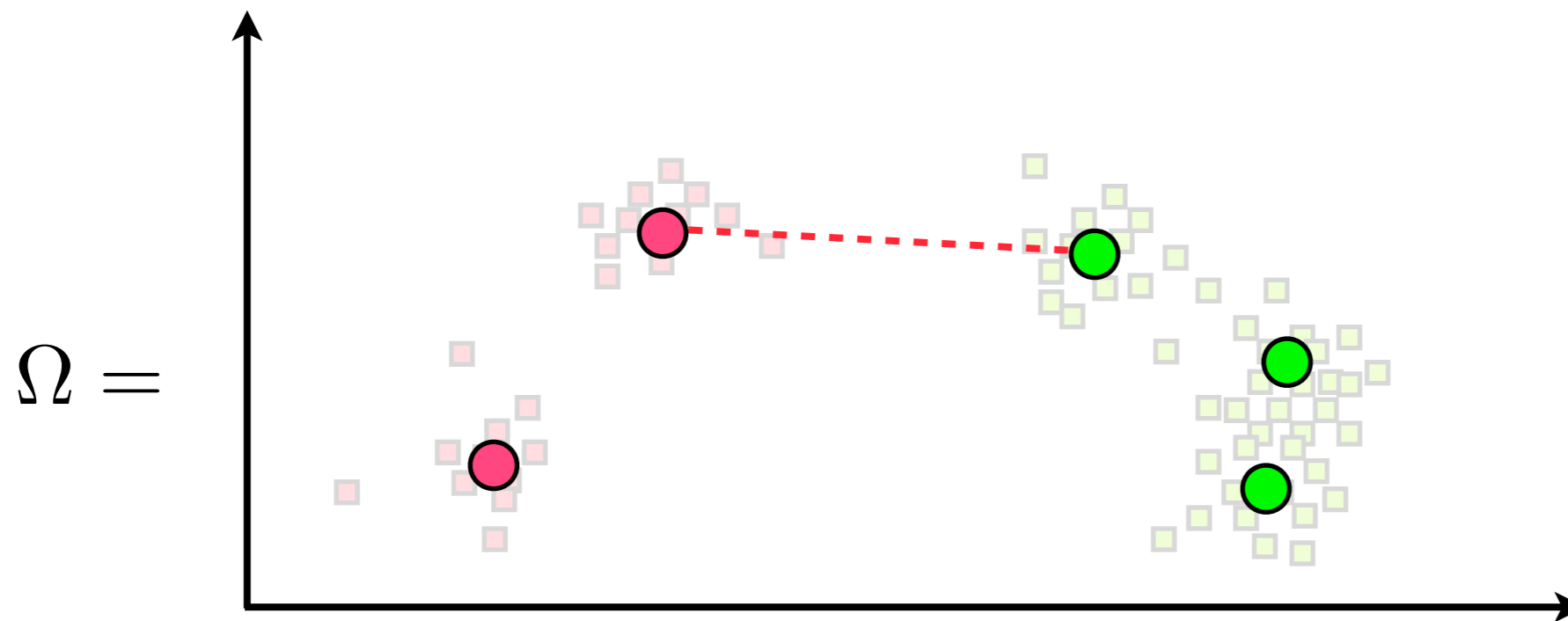
Erk (2007), Vandekerckhove et al. (2009)

Multi-Prototype Similarity Metrics



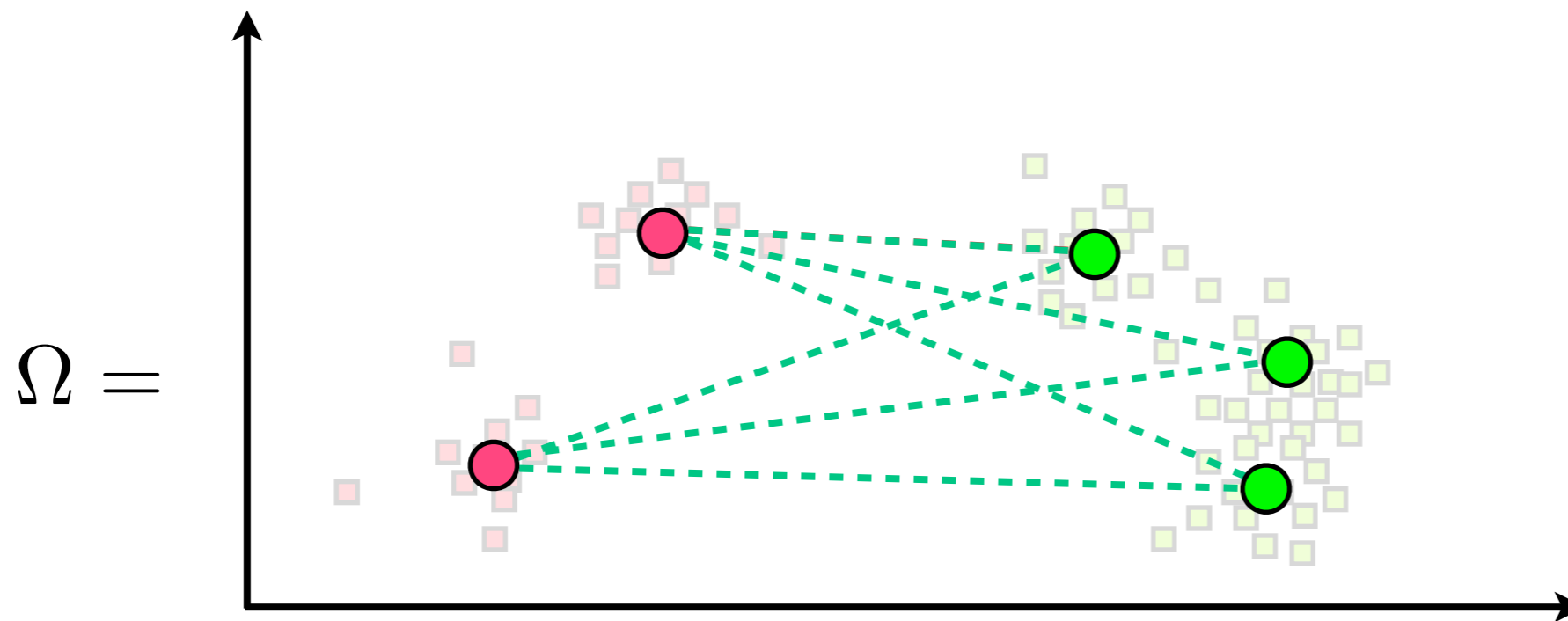
- MaxSim — Maximum pairwise similarity between any two prototypes.
- AvgSim — Average pairwise similarity over all prototypes.

Multi-Prototype Similarity Metrics



- MaxSim — Maximum pairwise similarity between any two prototypes.
- AvgSim — Average pairwise similarity over all prototypes.

Multi-Prototype Similarity Metrics



- MaxSim — Maximum pairwise similarity between any two prototypes.
- AvgSim — Average pairwise similarity over all prototypes.

Feature Engineering / Weighting

- Choosing an embedding vector space:
 - **features** (unigram, bigram, collocation, dependency, ...)
 - **feature weighting** (t-test, tf-idf, χ^2 , MI, ...)
 - **metric / inner product** (cosine, Jaccard, KL, ...)
- The multi-prototype method is essentially agnostic to these implementation details

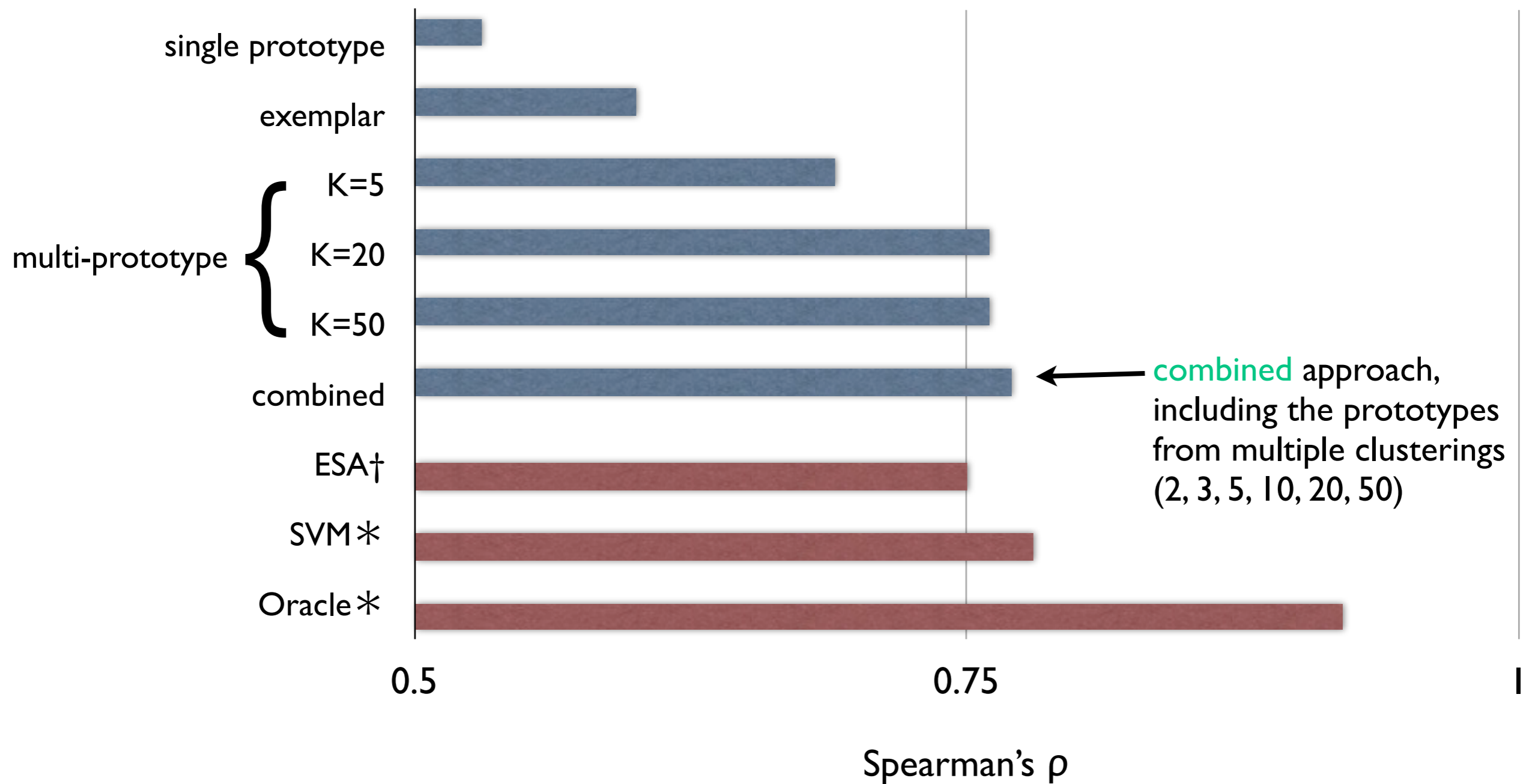
Feature Engineering / Weighting

- Choosing an embedding vector space:
 - **features** (unigram, bigram, collocation, dependency, ...)
 - **feature weighting** (t-test, tf-idf, χ^2 , MI, ...)
 - **metric / inner product** (cosine, Jaccard, KL, ...)
- The multi-prototype method is essentially agnostic to these implementation details

Experimental setup

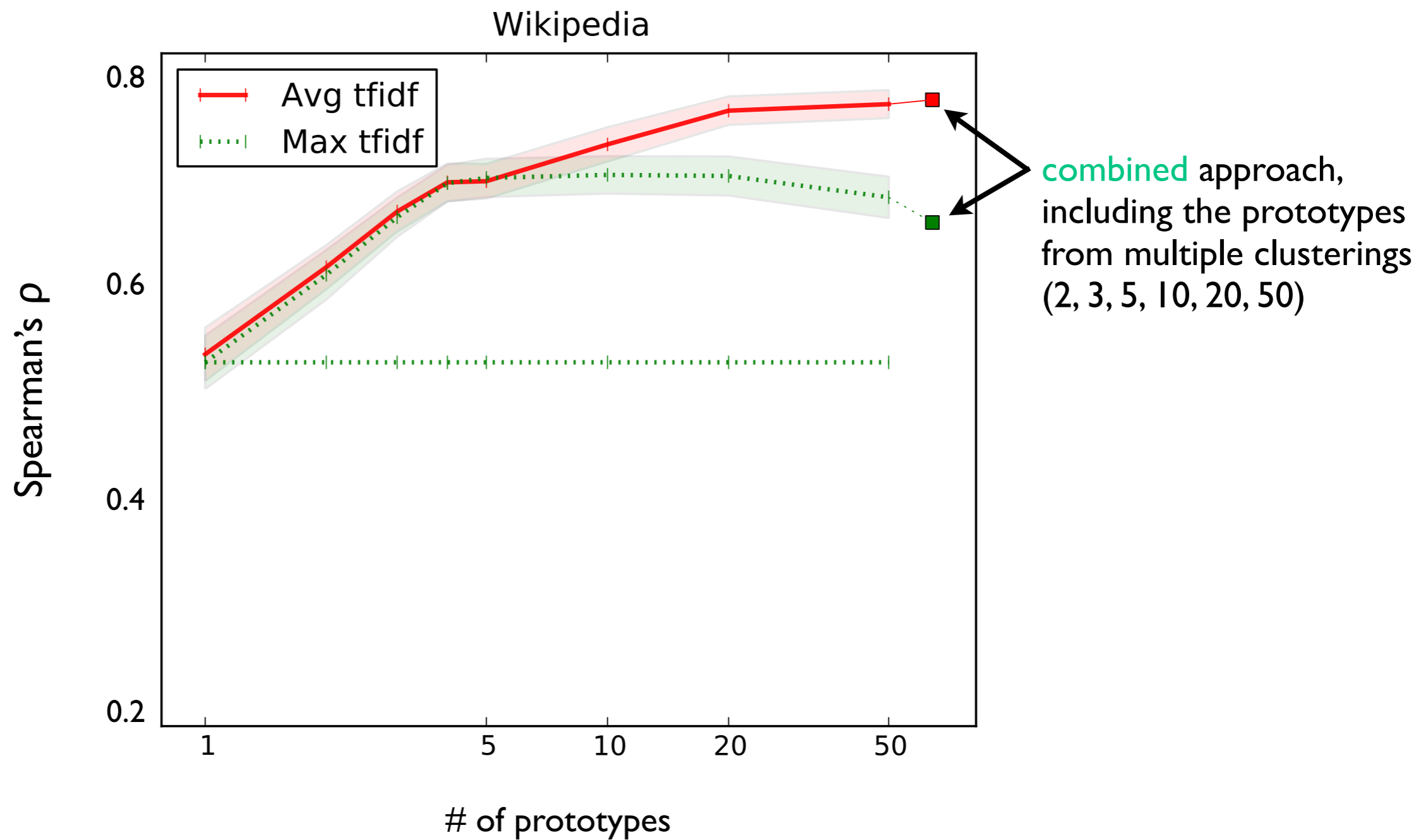
- Wikipedia as the base textual corpus (2.8M articles, 2B words)
- Evaluation:
 1. WordSim-353 collection (353 word pairs with ~15 human similarity judgements each) [Finkelstein et al. \(2002\)](#); using Spearman's rank correlation [Agirre et al. \(2009\)](#)
 2. Predicting related words; human raters from Amazon Mechanical Turk

Results: WordSim-353 Correlation



†Gabrilovich and Markovitch (2007), * Agirre et al. (2009)


Results: WordSim-353 Correlation



Predicting related words

Predicting related words

top-word:



beta Artificial Intelligence

party

Which word is more related to **party**?

☐ **government**

☐ **political**

reservation


Which word is more related to **reservation**?

☐ **settlers**

☐ **tribal**

Predicting related words

top-word:



party

Which word is more related to **party**?

☐ **government**

☐ **political**


reservation

Which word is more related to **reservation**?

☐ **settlers**

☐ **tribal**

top-set:



journal

Which set of words is more related to **journal**?

☐ **research, study, published**

☐ **publication, paper, study**

train

Which set of words is more related to **train**?

☐ **station, line, services**

☐ **passenger, rail, freight**

Predicting related words

top-word:



The screenshot shows two side-by-side task cards from Amazon Mechanical Turk. Each card has the Amazon Mechanical Turk logo at the top. The left card is for the word 'party' and asks 'Which word is more related to party?' with two radio button options: 'government' and 'political'. The right card is for the word 'reservation' and asks 'Which word is more related to reservation?' with two radio button options: 'settlers' and 'tribal'.

amazonmechanical turk
beta Artificial Artificial Intelligence

party

Which word is more related to **party**?

☐ **government**

☐ **political**

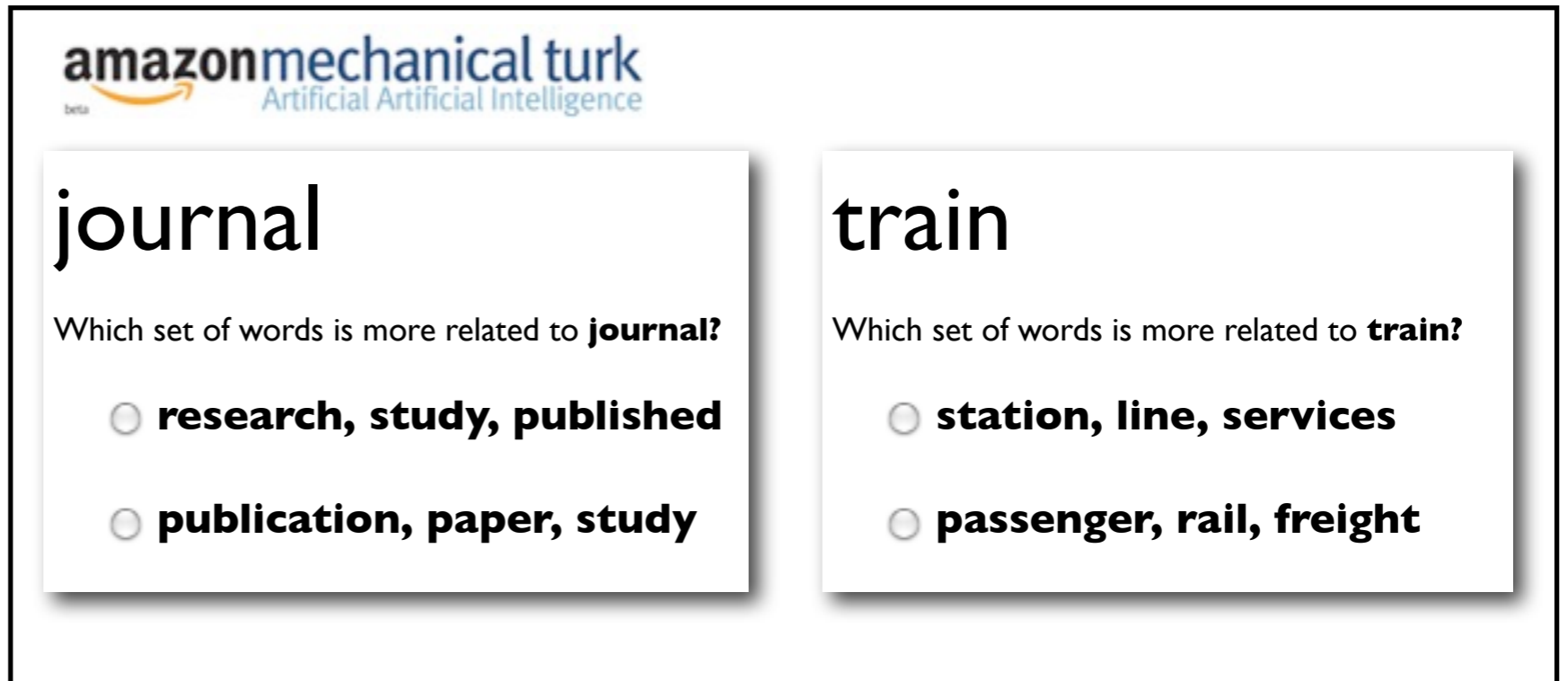
reservation

Which word is more related to **reservation**?

☐ **settlers**

☐ **tribal**

top-set:



The screenshot shows two side-by-side task cards from Amazon Mechanical Turk. Each card has the Amazon Mechanical Turk logo at the top. The left card is for the word 'journal' and asks 'Which set of words is more related to journal?' with two radio button options: 'research, study, published' and 'publication, paper, study'. The right card is for the word 'train' and asks 'Which set of words is more related to train?' with two radio button options: 'station, line, services' and 'passenger, rail, freight'.

amazonmechanical turk
beta Artificial Artificial Intelligence

journal

Which set of words is more related to **journal**?

☐ **research, study, published**

☐ **publication, paper, study**

train

Which set of words is more related to **train**?

☐ **station, line, services**

☐ **passenger, rail, freight**

- 79 raters, 7.6K comparisons

Predicting related words

top-word:

amazonmechanical turk
beta Artificial Intelligence

party

Which word is more related to **party**?

☐ government

☒ political

reservation

Which word is more related to **reservation**?

☒ settlers

☐ tribal

top-set:

amazonmechanical turk
beta Artificial Intelligence

journal

Which set of words is more related to **journal**?

☐ research, study, published

☒ publication, paper, study

train

Which set of words is more related to **train**?

☐ station, line, services

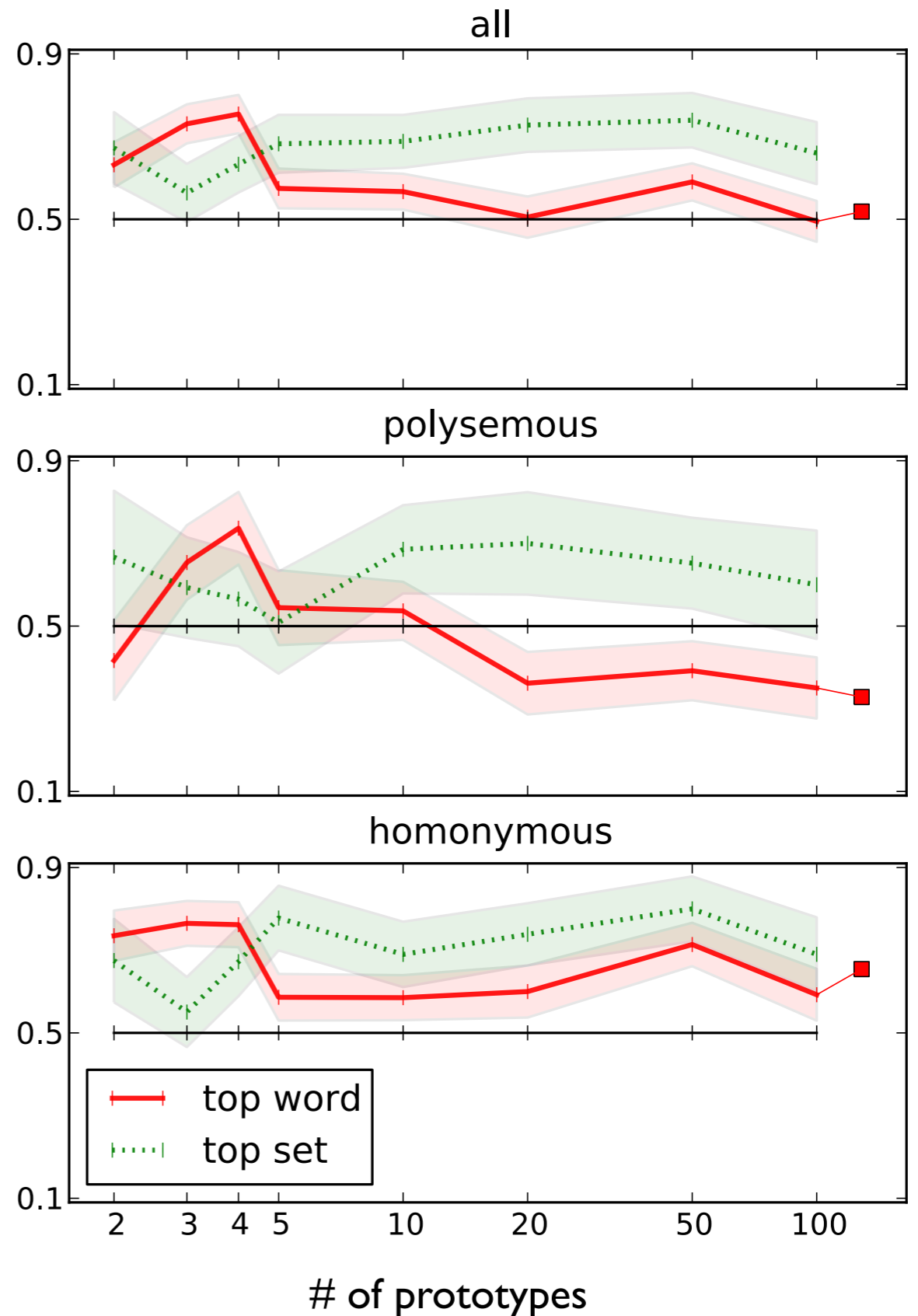
☒ passenger, rail, freight

- 79 raters, 7.6K comparisons

Results: Non-contextual Prediction

homonymous	carrier, crane, cell, company, issue, interest, match, media, nature, party, practice, plant, racket, recess, reservation, rock, space, value
polysemous	cause, chance, journal, market, network, policy, power, production, series, trading, train

% Multi-prototype favored



Contextual Prediction

I have some reservation due to the high potential for violations.

Which word is more related to **reservation** as used in the sentence above?

- ☐ **tribal**
- ☐ **thoughtful**

When there is more variation in wage offers, the searcher may want to wait longer (that is, set a higher reservation wage) in hopes of receiving an exceptionally high wage offer.

Which word is more related to **reservation** as used in the sentence above?

- ☐ **tribal**
- ☐ **minimum**

Contextual Prediction

I have some reservation due to the high potential for violations.

Which word is more related to **reservation** as used in the sentence above?

- ☐ **tribal**
- ☐ **thoughtful**

When there is more variation in wage offers, the searcher may want to wait longer (that is, set a higher reservation wage) in hopes of receiving an exceptionally high wage offer.

Which word is more related to **reservation** as used in the sentence above?

- ☐ **tribal**
- ☐ **minimum**

- 127 raters, ~10K comparisons

Contextual Prediction

I have some reservation due to the high potential for violations.

Which word is more related to **reservation** as used in the sentence above?

☐ **tribal**

☒ **thoughtful**

When there is more variation in wage offers, the searcher may want to wait longer (that is, set a higher reservation wage) in hopes of receiving an exceptionally high wage offer.

Which word is more related to **reservation** as used in the sentence above?

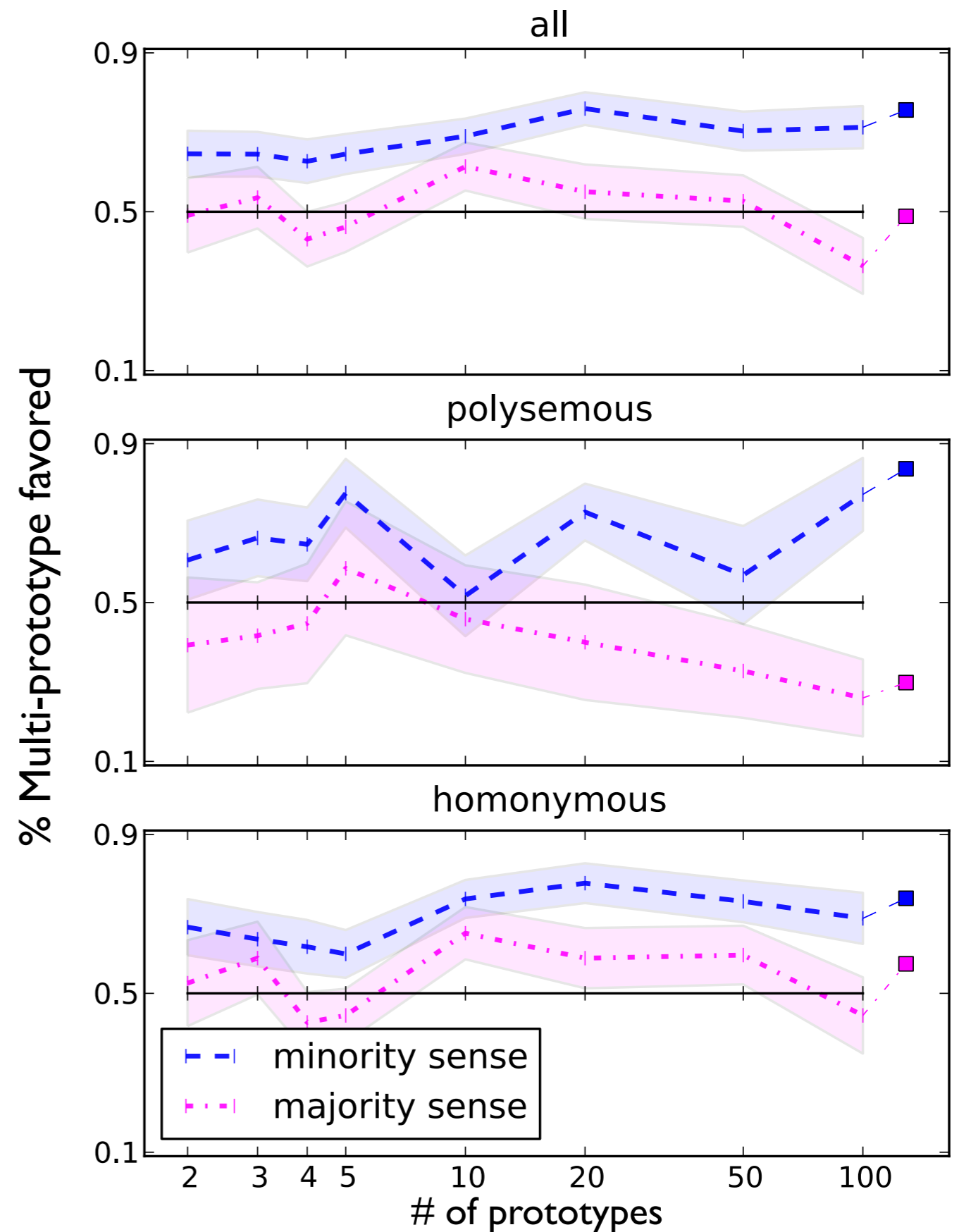
☐ **tribal**

☒ **minimum**

- 127 raters, ~10K comparisons

Results: Contextual Prediction

homonymous	carrier, crane, cell, company, issue, interest, match, media, nature, party, practice, plant, racket, recess, reservation, rock, space, value
polysemous	cause, chance, journal, market, network, policy, power, production, series, trading, train



Conclusion

- Represent word meaning as a collection of prototype vectors.
- Outperforms single-prototype, but introduces more noise (like exemplar).
- Trade-off for doing clustering step.
- Can we define better distance metrics? KL?
 - account for asymmetry?

Questions?

Pruning

