

# ANALYZING TEXT SIMILARITIES

*A NLP approach on Metadecidim data*

Ana Valdivia  
22 May 2018



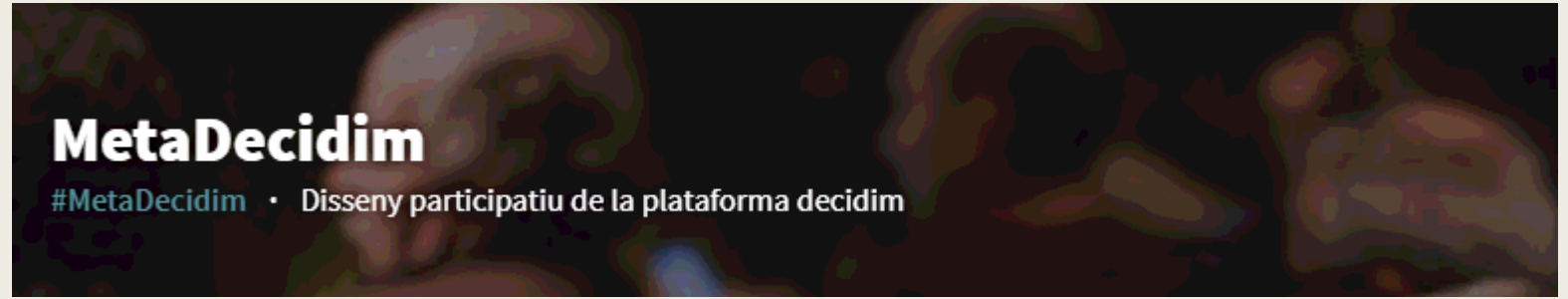
**Universitat  
Pompeu Fabra**  
*Barcelona*

# Questions

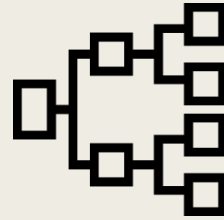
- Are proposals and actions similar?
- Do citizens write the same way as the city council?
- Can deep learning answer this questions?

# Concepts

## Metadecidim flow



1. Citizens, organizations and the city council write proposals.



2. Similar proposals are clustered into actions.



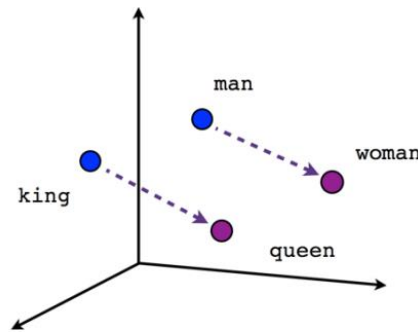
3. Some **actions** are approved and they go on to be executed.

# Methodology

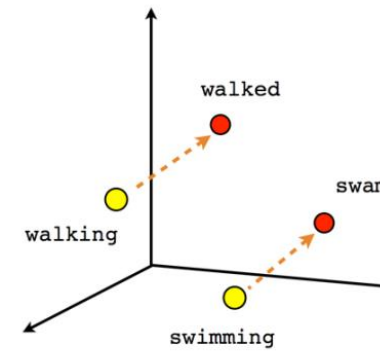
Apply NLP and deep learning techniques

## word2vec

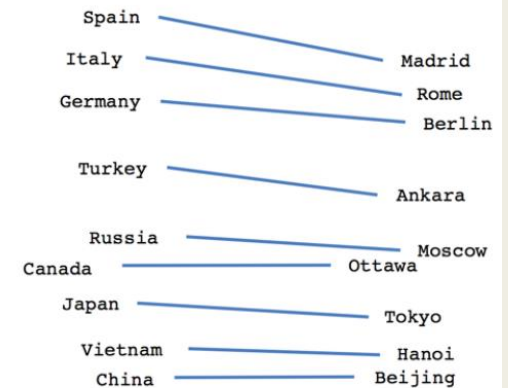
Deep learning model that represents words as vectors.



Male-Female



Verb tense



Country-Capital

Source: <https://towardsdatascience.com/deep-learning-4-embedding-layers-f9a02d55ac12>

# Methodology

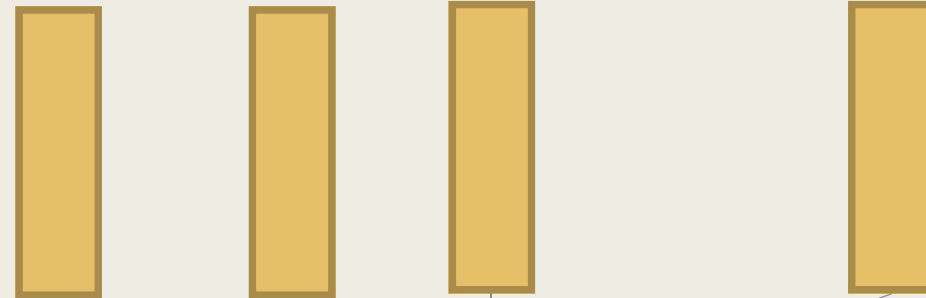
Apply NLP and deep learning techniques

## doc2vec

Mean of all  
word2vec of  
words within a  
text.

“Crear espais nous per a gossos.”

word2vecs



doc2vec of proposal x

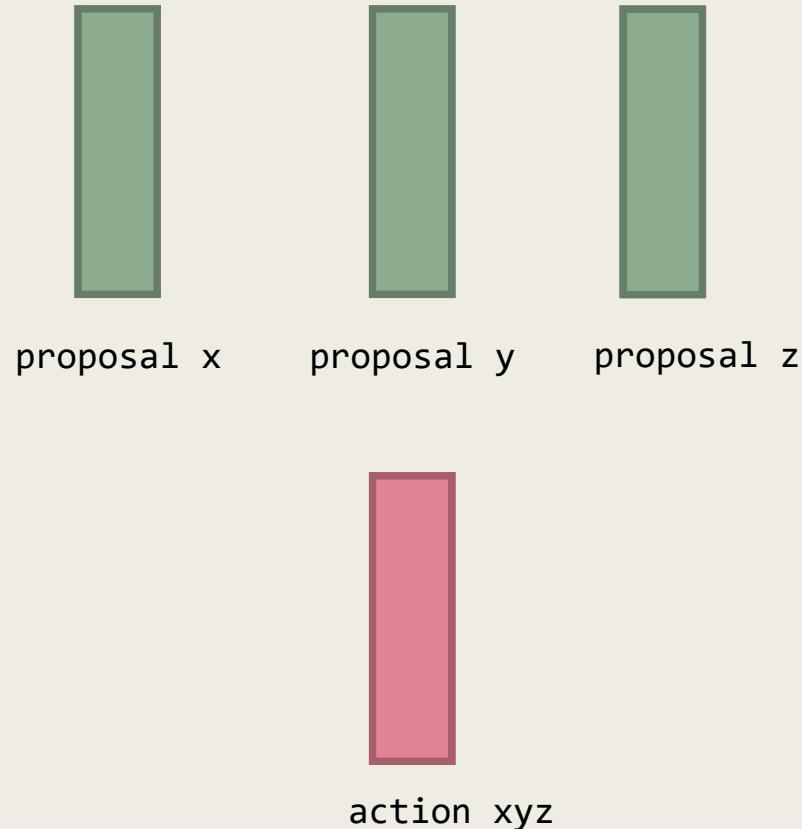


# Methodology

Apply NLP and deep learning techniques

## doc2vec

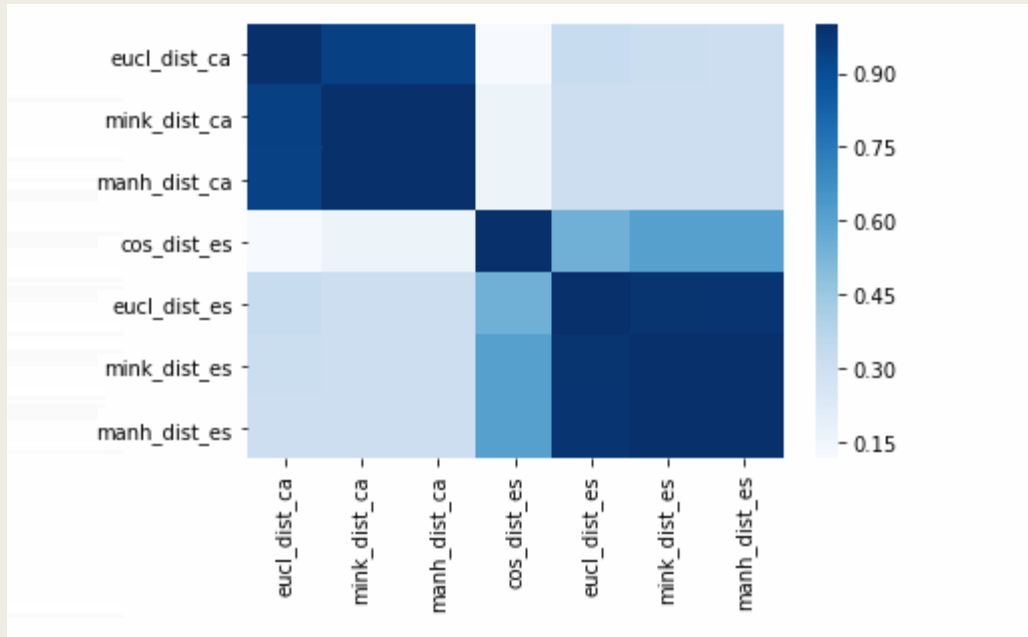
Mean of all  
word2vec of  
words within a  
text.



similarities  
=  
distance  
functions  
between  
vectors

# Experiments

Correlation between distance functions



1. Functions are strongly correlated.

2. Catalan and Spanish vectors are not correlated.

# Experiments

Ranking of the most modified proposals

cosinus

```
pd.merge(data_sort_cos_dist_ca, data_sort_cos_dist_es, on = 'id')
```

	id	origin_x	scope_x	district_x	category_x	subcategory_x	title_x	description_x	author_id_x	author_name_x	created_at_x	votes_x	comments_x
0	5748	meeting	district	Les Corts	Transició ecològica	Medi ambient i espai públic	Regular els objectes de la vorera	Estudiar i regular els objectes que hi ha a la...	9289	suport decidim.HG i Les Corts	3/17/2016	2	0
1	5751	meeting	district	Les Corts	Transició ecològica	Medi ambient i espai públic	Omplir amb resina els escocells	Omplir amb resina els escocells per evitar cai...	9289	suport decidim.HG i Les Corts	3/17/2016	6	2



# Experiments

Ranking of the most modified proposals

euclidean

	id	origin_x	scope_x	district_x	category_x	subcategory_x	title_x	description_x	author_id_x	author_name_x	created_at_x	votes_x
0	10173	citizenship	district	Sant Martí	Transició ecològica	Verd urbà i biodiversitat	Campanya de desratització	Campanya de desratització	21893	consellsdistrictesantmarti	04/08/2016	2

minkonswki

	id	origin_x	scope_x	district_x	category_x	subcategory_x	title_x	description_x	author_id_x	author_name_x	created_at_x	votes_x
0	10173	citizenship	district	Sant Martí	Transició ecològica	Verd urbà i biodiversitat	Campanya de desratització	Campanya de desratització	21893	consellsdistrictesantmarti	04/08/2016	2

# Experiments

Ranking of the most modified proposals

manhattan

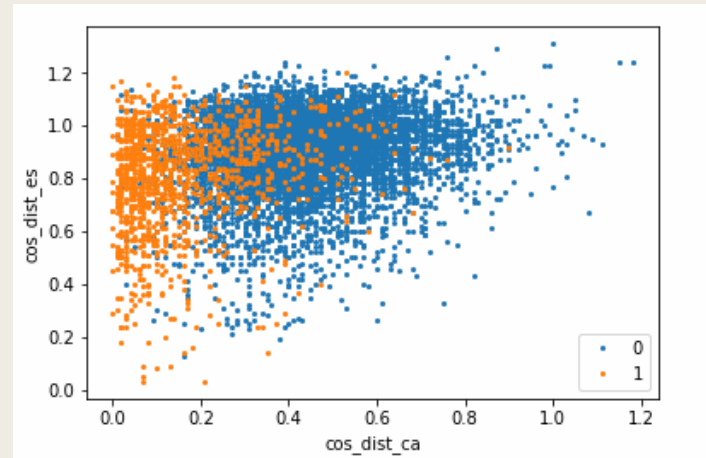
	id	origin_x	scope_x	district_x	category_x	subcategory_x	title_x	description_x	author_id_x	author_name_x	created_at_x	votes_x
0	10173	citizenship	district	Sant Martí	Transició ecològica	Verd urbà i biodiversitat	Campanya de desratització	Campanya de desratització	21893	consellsdistrictesantmarti	04/08/2016	2

The proposal 10173 appears as the most modified in three functions.

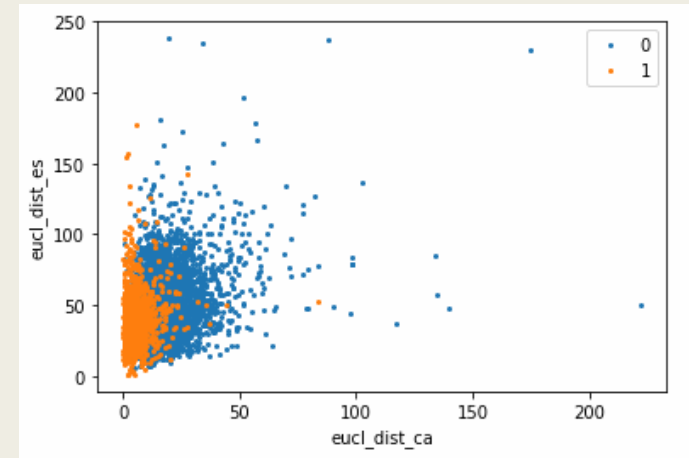
# Experiments

Distances between actions and proposals (0 = citizens, 1 = city council)

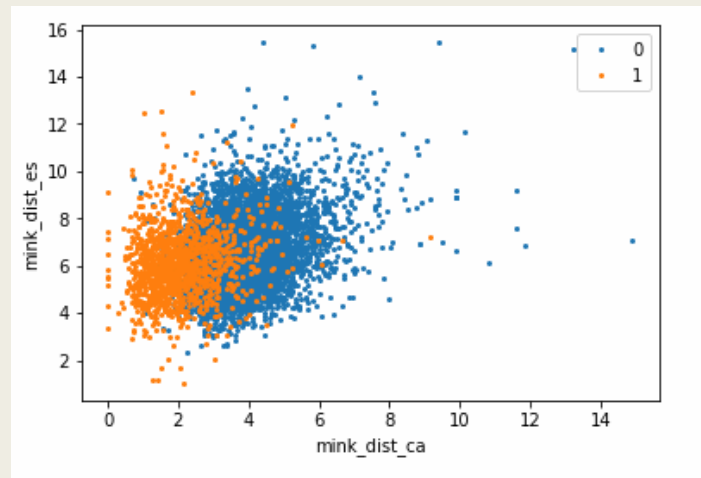
cosinus



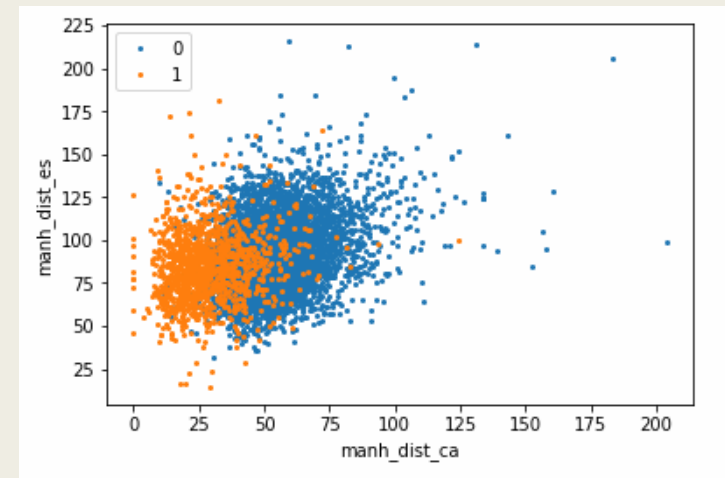
euclidean



minkowsky



manhattan



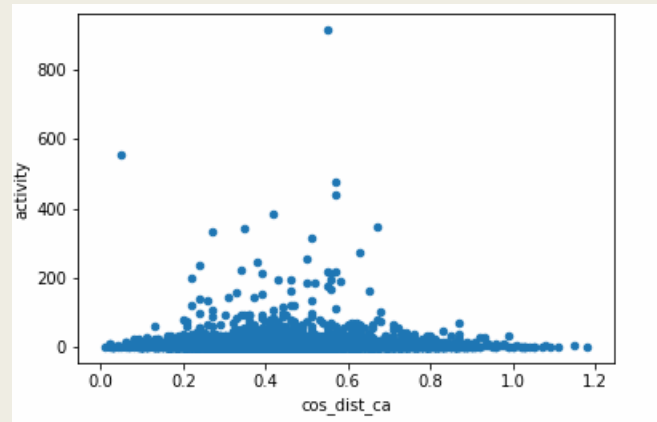
# Experiments

Analyzing citizens proposals

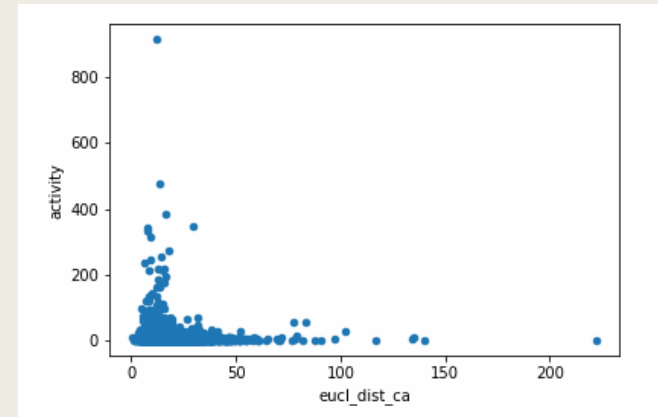
Catalan

$$\text{activity} = 0.5 \times (\text{votes} + \text{comments})$$

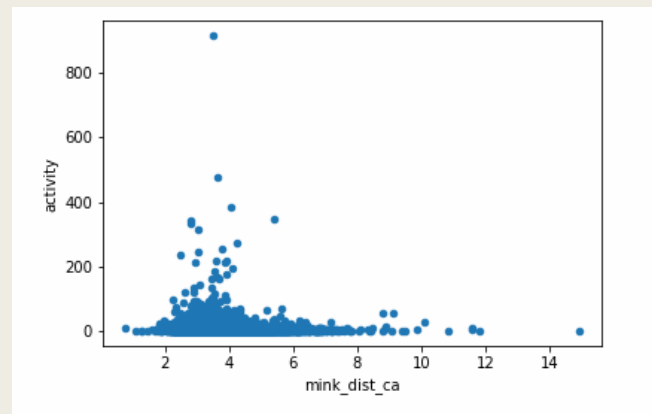
cosinus



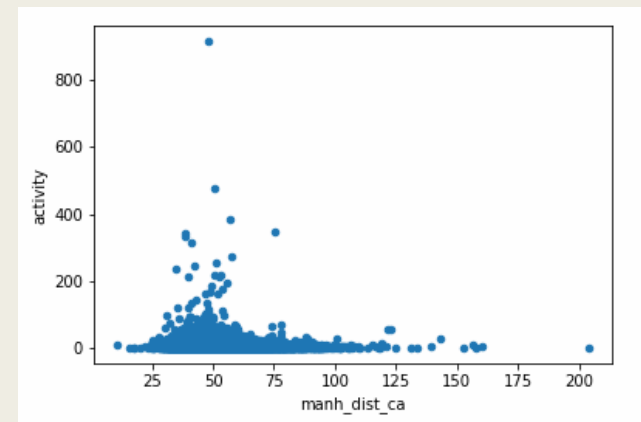
euclidean



minkowsky



manhattan



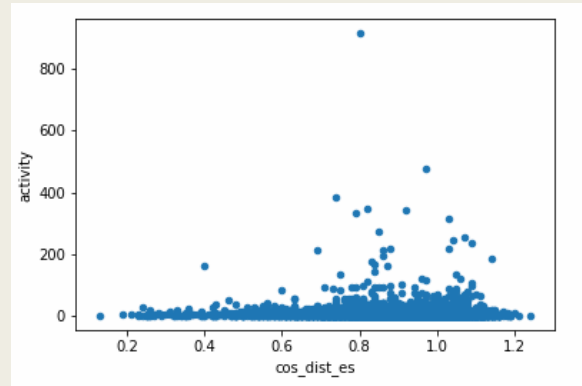
# Experiments

Analyzing citizens proposals

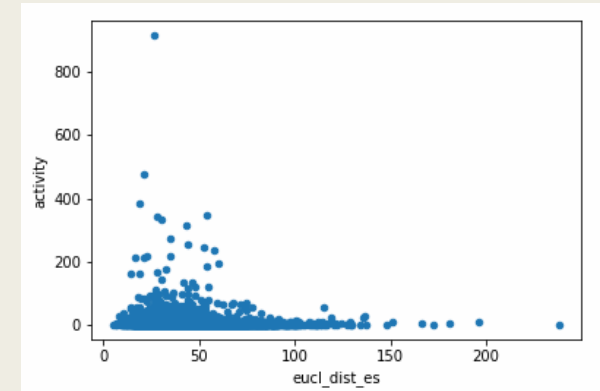
Spanish

$$\text{activity} = 0.5 \times (\text{votes} + \text{comments})$$

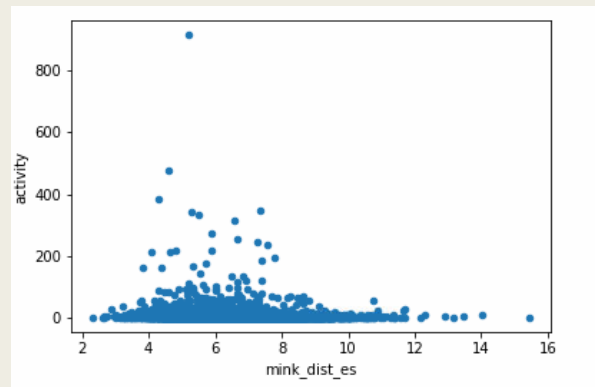
cosinus



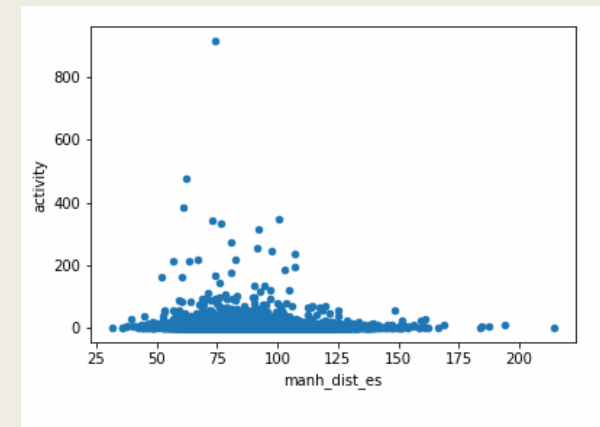
euclidean



minkowsky



manhattan

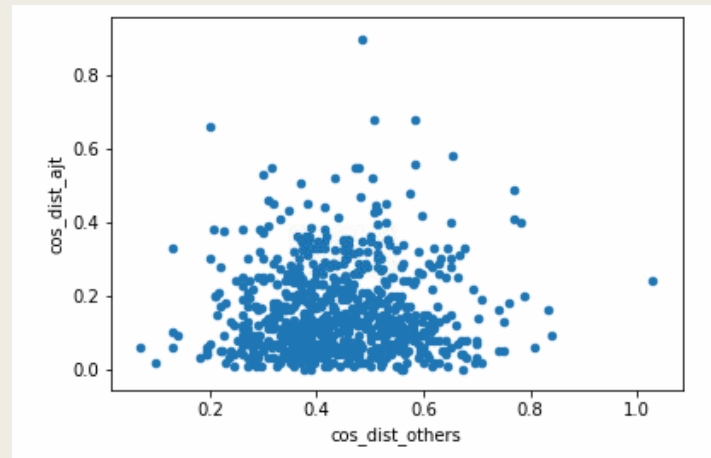


# Experiments

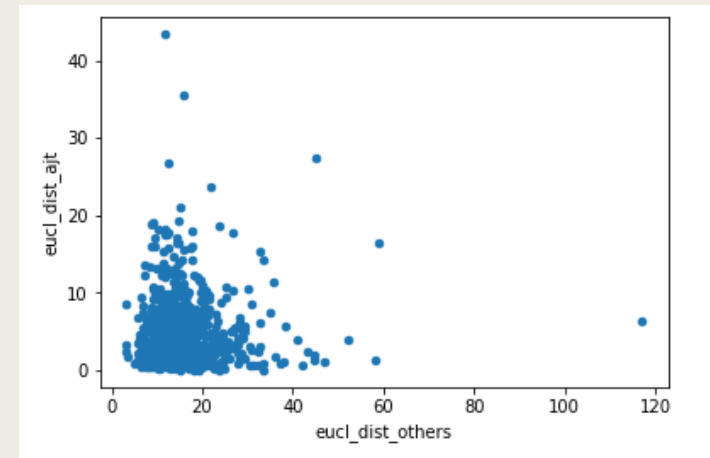
Catalan

Similarities between the proposals of the city council and the citizens

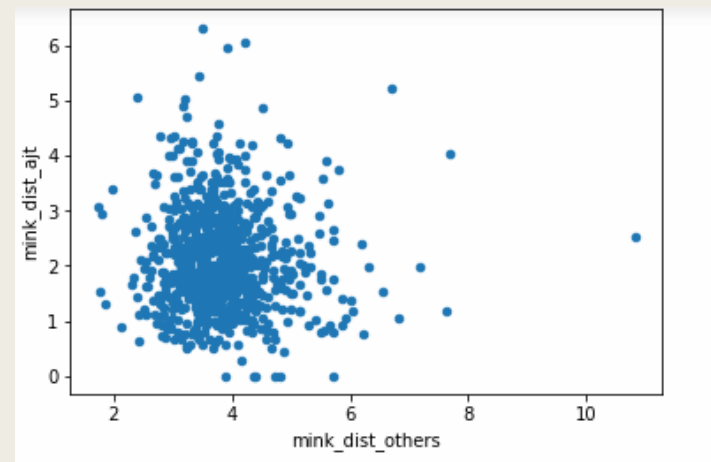
cosinus



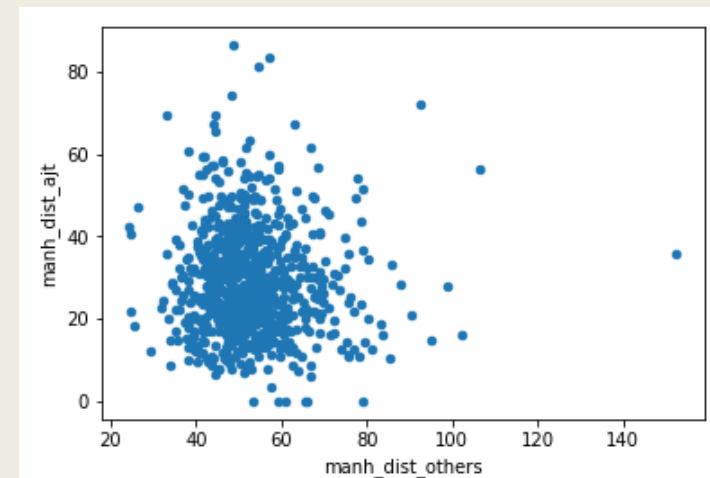
euclidean



minkowsky



manhattan

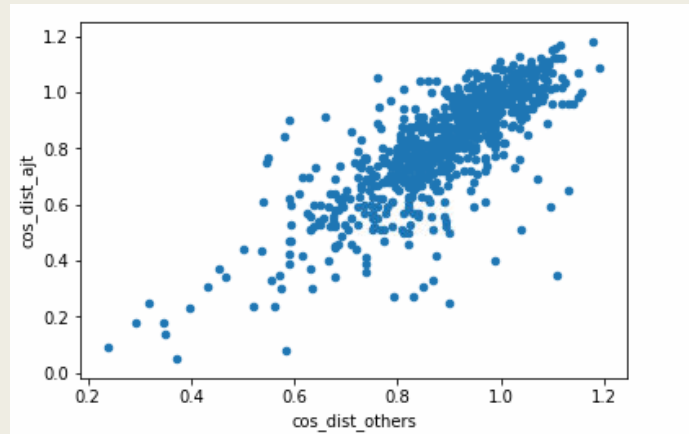


# Experiments

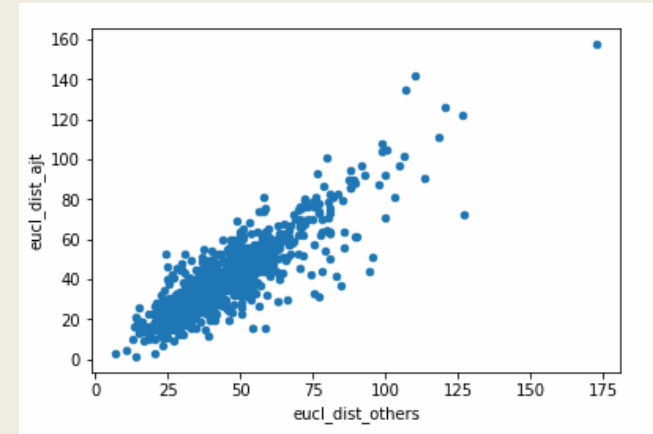
Spanish

Similarities between the proposals of the city council and the citizens

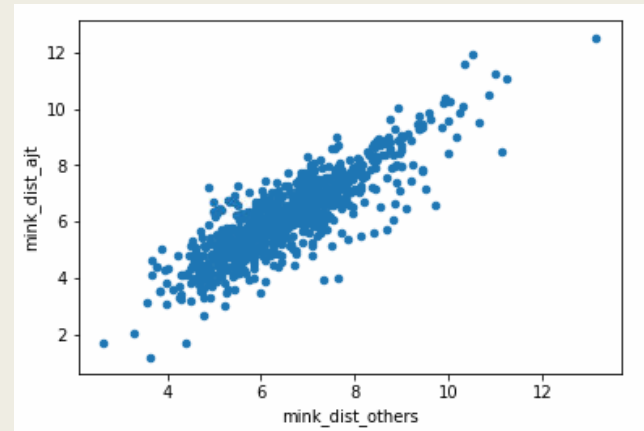
cosinus



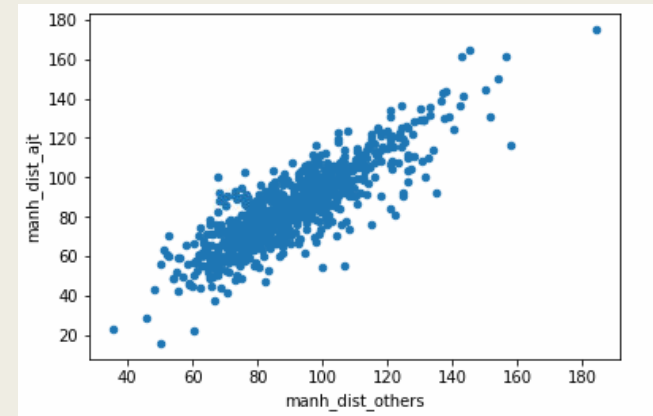
euclidean



minkowsky



manhattan



# Conclusions

- City council proposals are closer to action texts than citizen proposals.
- There exists differences on the distributions of Catalan and Spanish word2vec.



# Future work

- Analyze “Ronda de Dalt” as a paradigmatic case.
- There exists differences on the distributions of Catalan and Spanish word2vec.
- What else...?

**Thanks!**