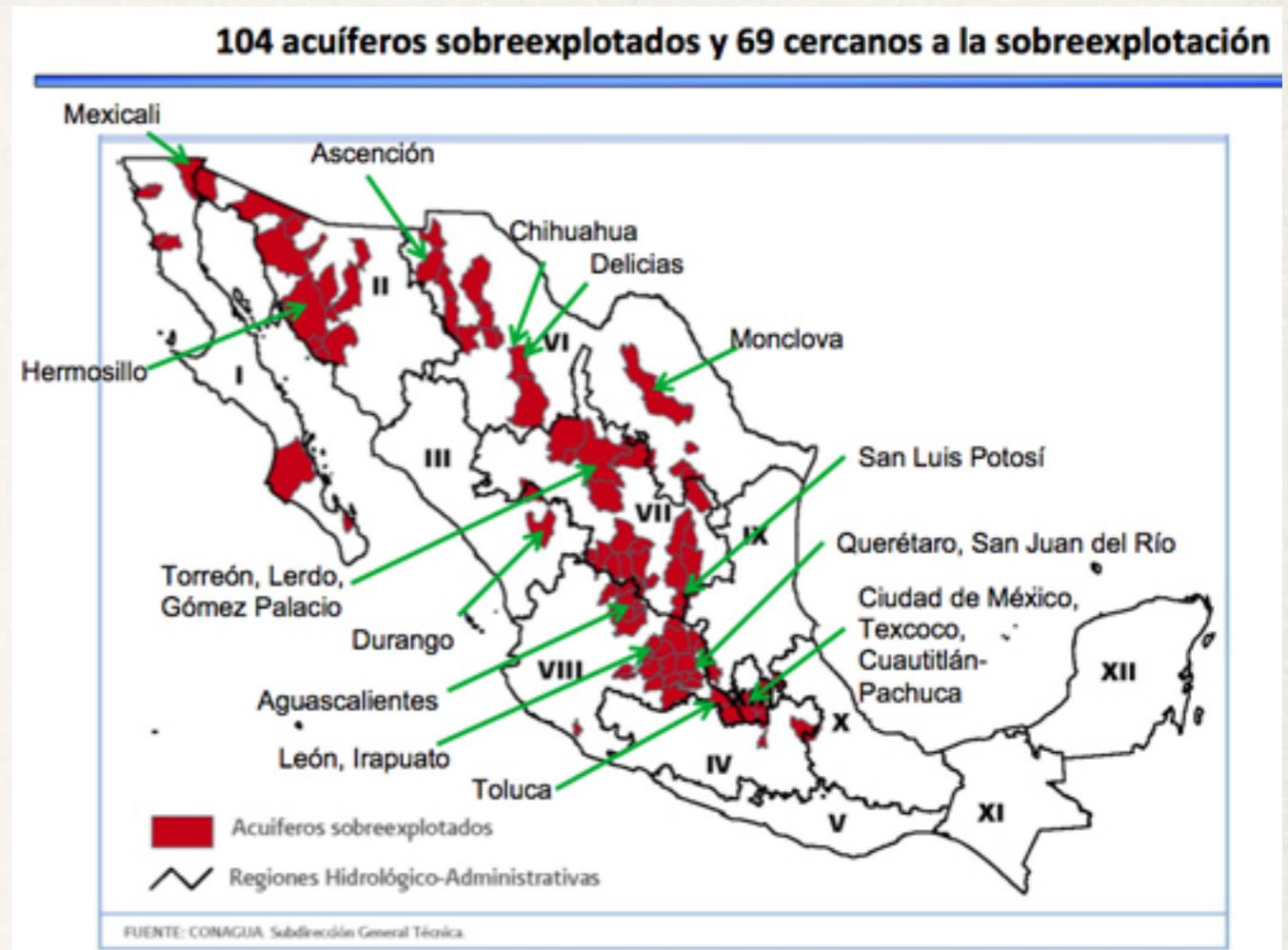
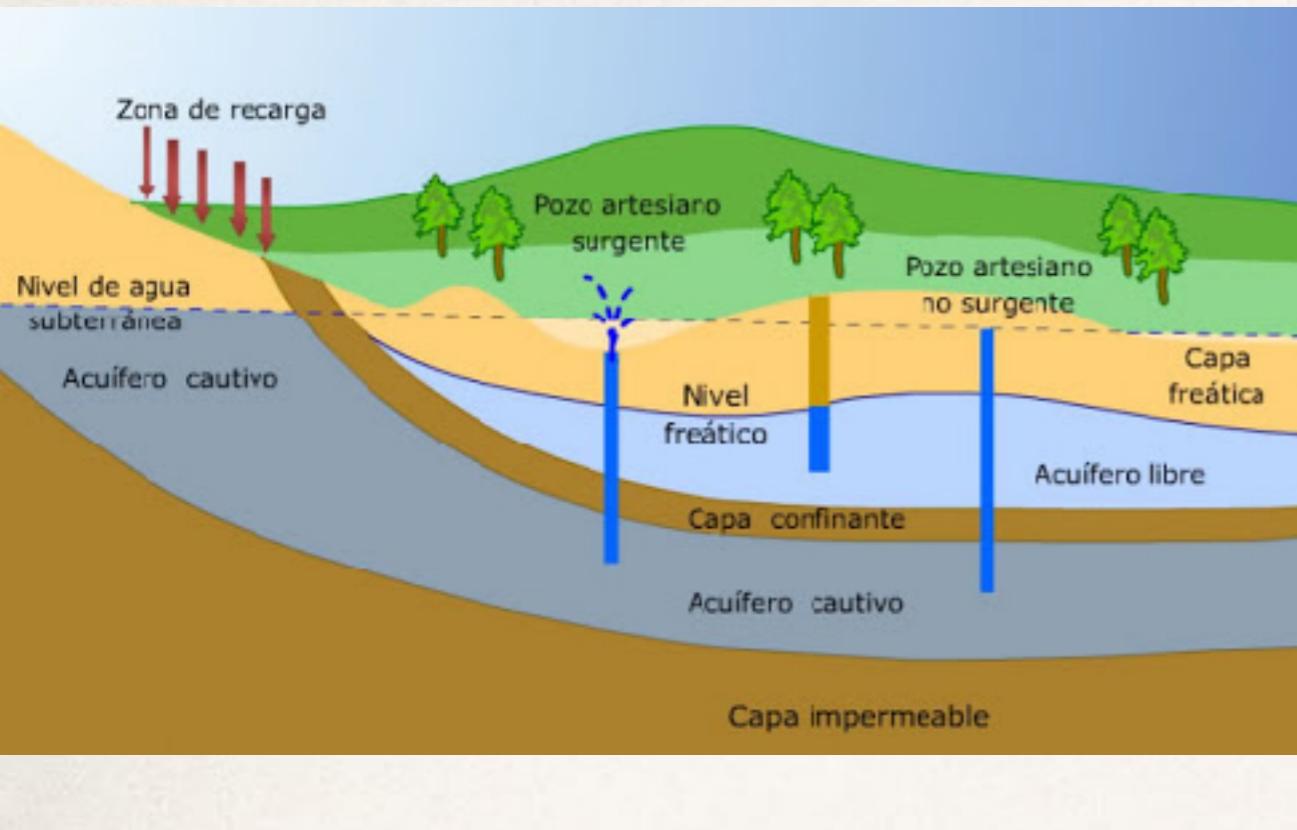


# Chihuahua, Mx, Septiembre 2020 — Agua Superficial



# Desarrollo de ciudades - Agua Subterránea



¿Cómo administrar mejor este recurso?



# Towards Smart Water Management via Machine Learning

Luis Carlos González Gurrola



UNIVERSIDAD AUTÓNOMA DE  
**CHIHUAHUA**



UNIVERSIDAD AUTÓNOMA DE  
**CHIHUAHUA**

# Team

Martín Mauricio Alarcón Barraza

Luis Carlos González Gurrola



Sergio Alberto Valdés Rabelo (DataZone)

Miguel Ángel Medina Pérez (ITESM-CEM)



# Problemas en el suministro de agua

---

No hay agua!



¿y toda esta agua?



# ¿Qué se hace actualmente?

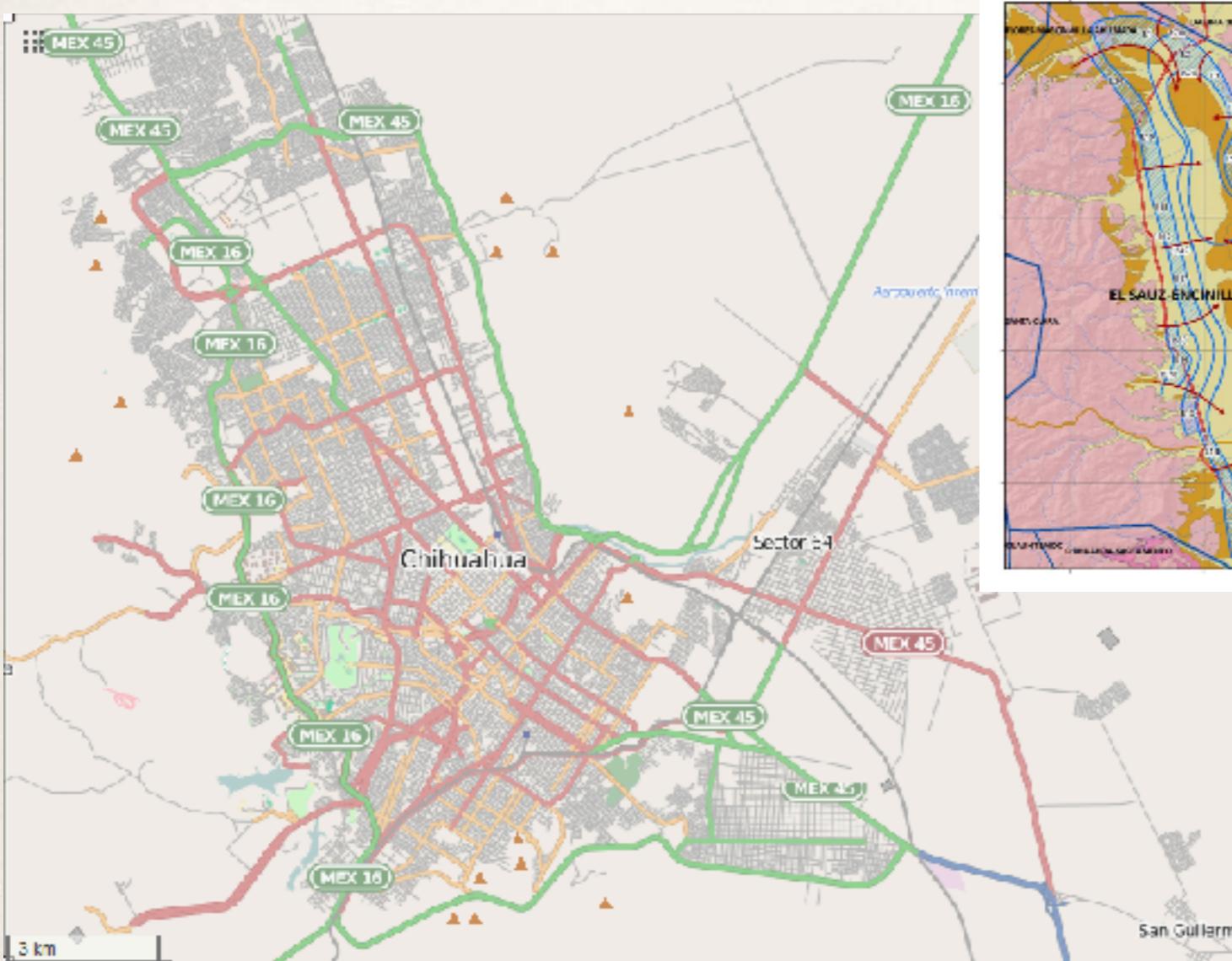


# ¿Qué se podría hacer?



# ¡Qué bonito es Chihuahua!

*Estríbillo corrido de Chihuahua*



~1,000,000 hab  
~250,000 viviendas  
Acuífero Sauz-Encinillas

## Red Hidráulica

38 Tanques  
97 Pozos  
29 Rebombeos  
37 Válvulas



**JUNTA MUNICIPAL**  
DE AGUA Y SANEAMIENTO  
DE CHIHUAHUA

## SCADA (Supervisory Control and Data Acquisition)

Pozo



Sensado

Tanque



Telemetría



Válvula



Rebombeo

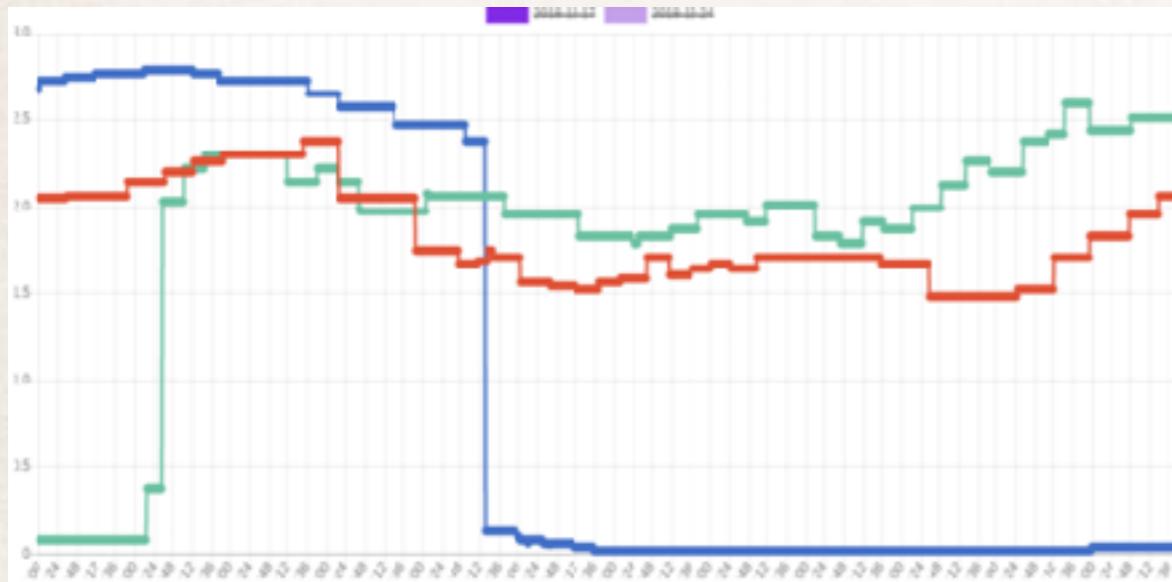


# Datos, datos, datos....

Pozos



Presión  
Gasto  
Falla de fase



Tanques



Gasto Salida  
Gasto Entrada  
Nivel

Rebombeo



Gasto Salida  
Gasto Entrada  
Presión Succión  
Presión Descarga

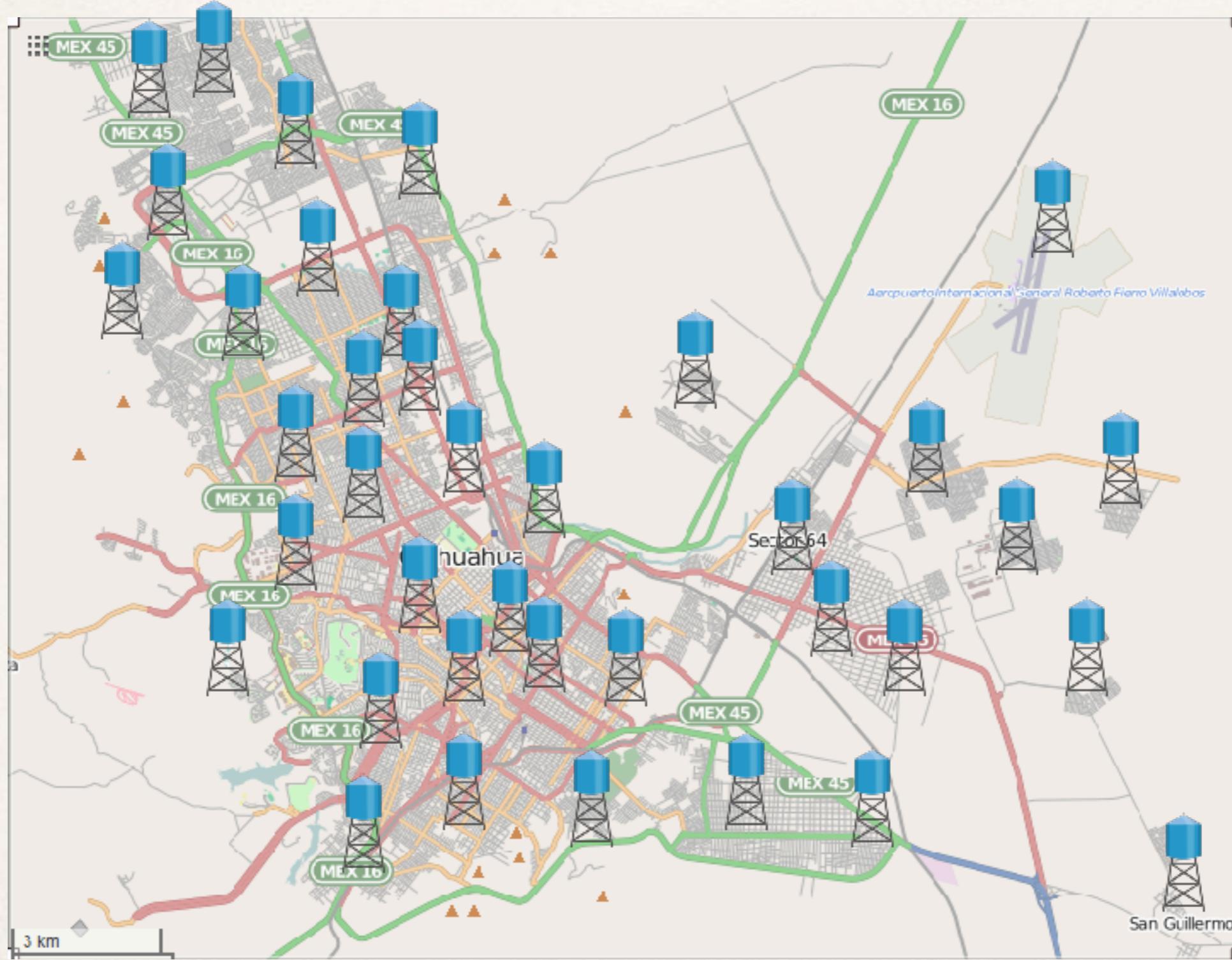
Válvula



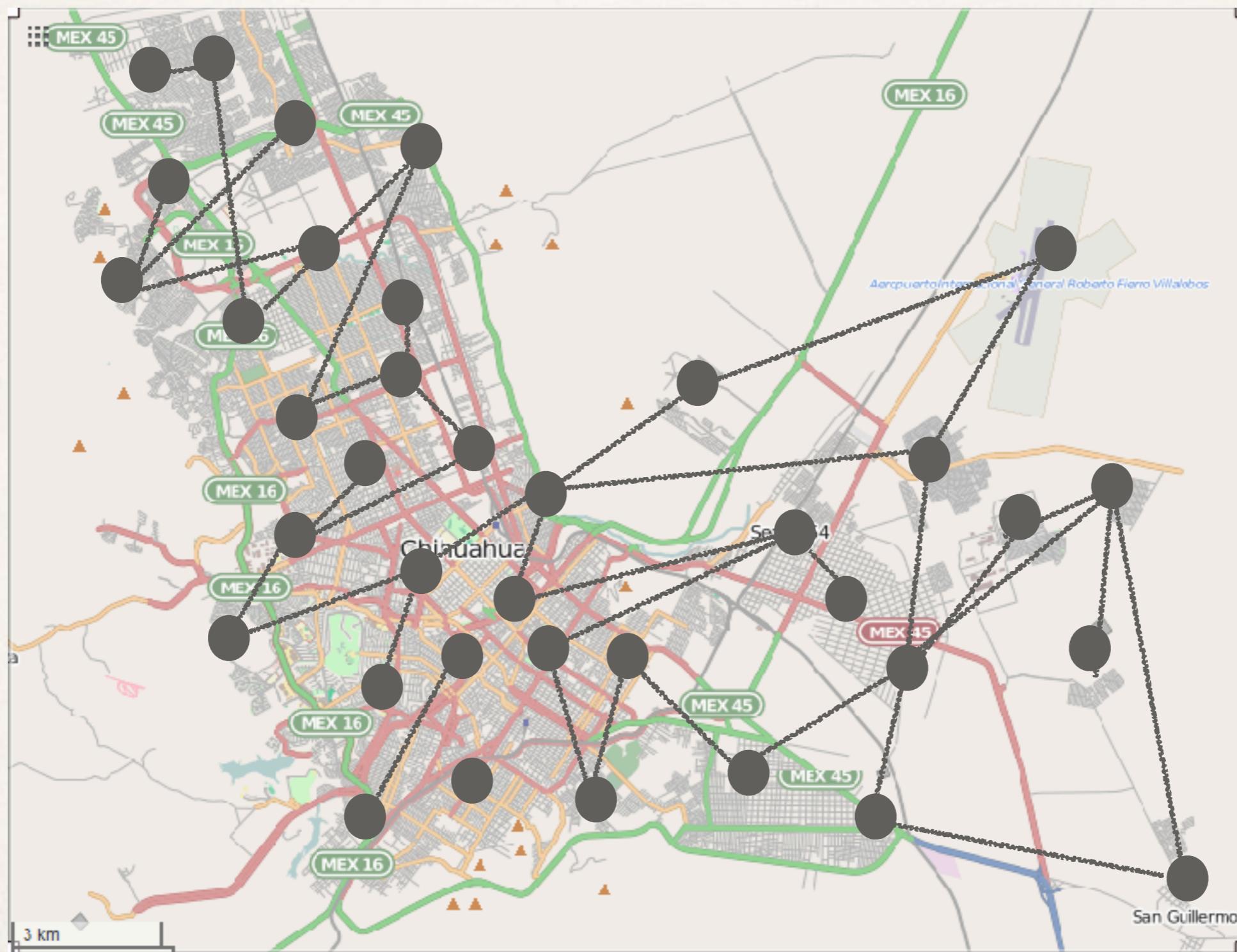
Presión Descarga  
Presión Salida

¿Cómo se ven estos datos?

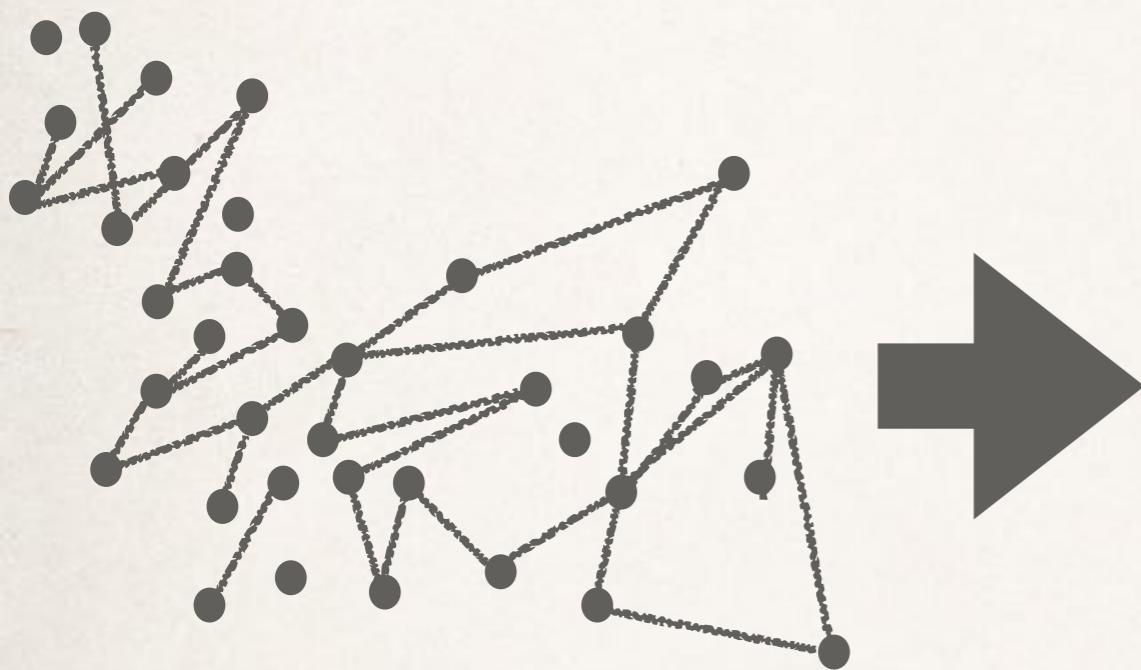
# Red Hidráulica



# Red Hidráulica



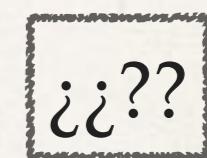
# Machine Learning



Método  
Computacional

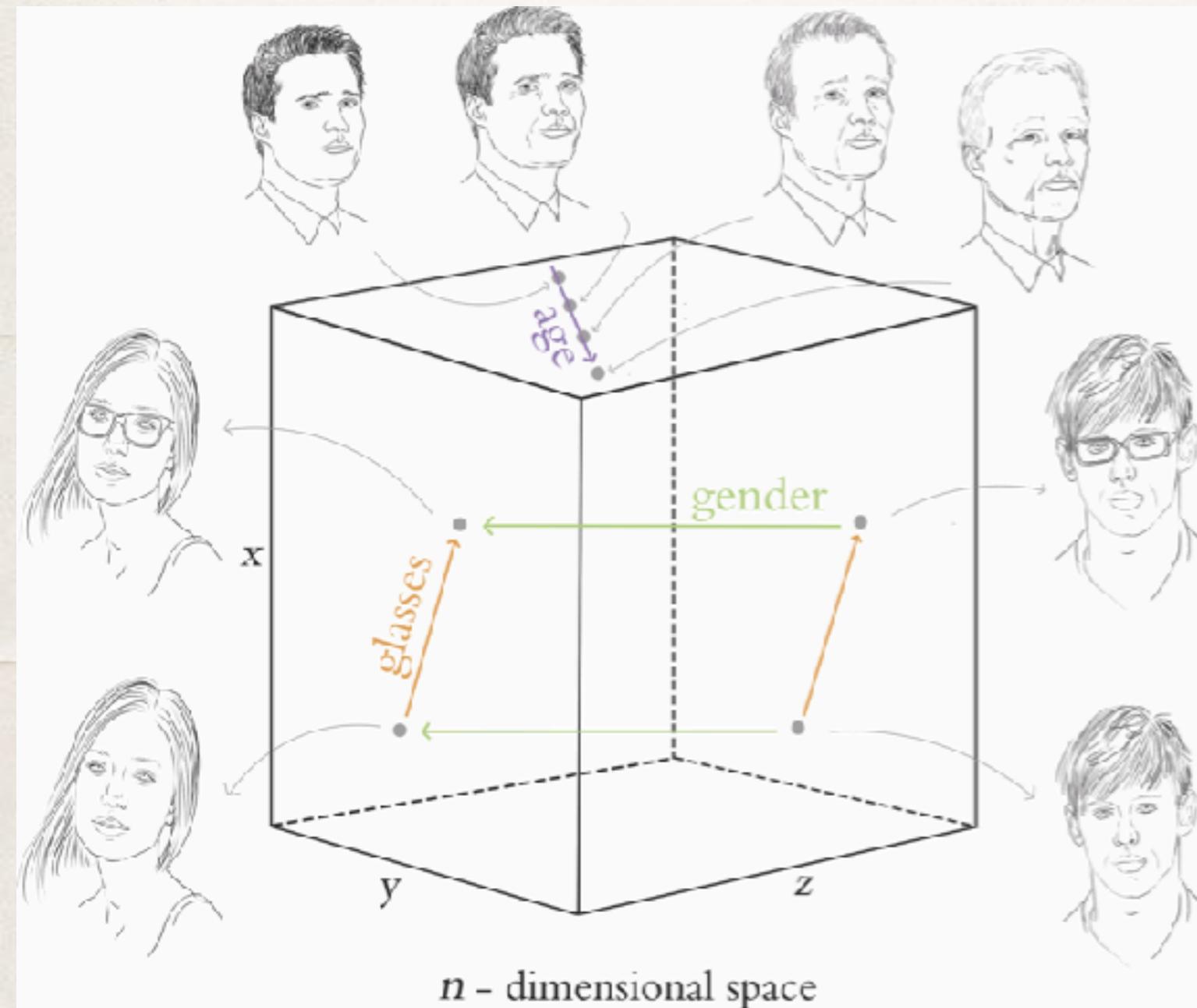


No  
Problema

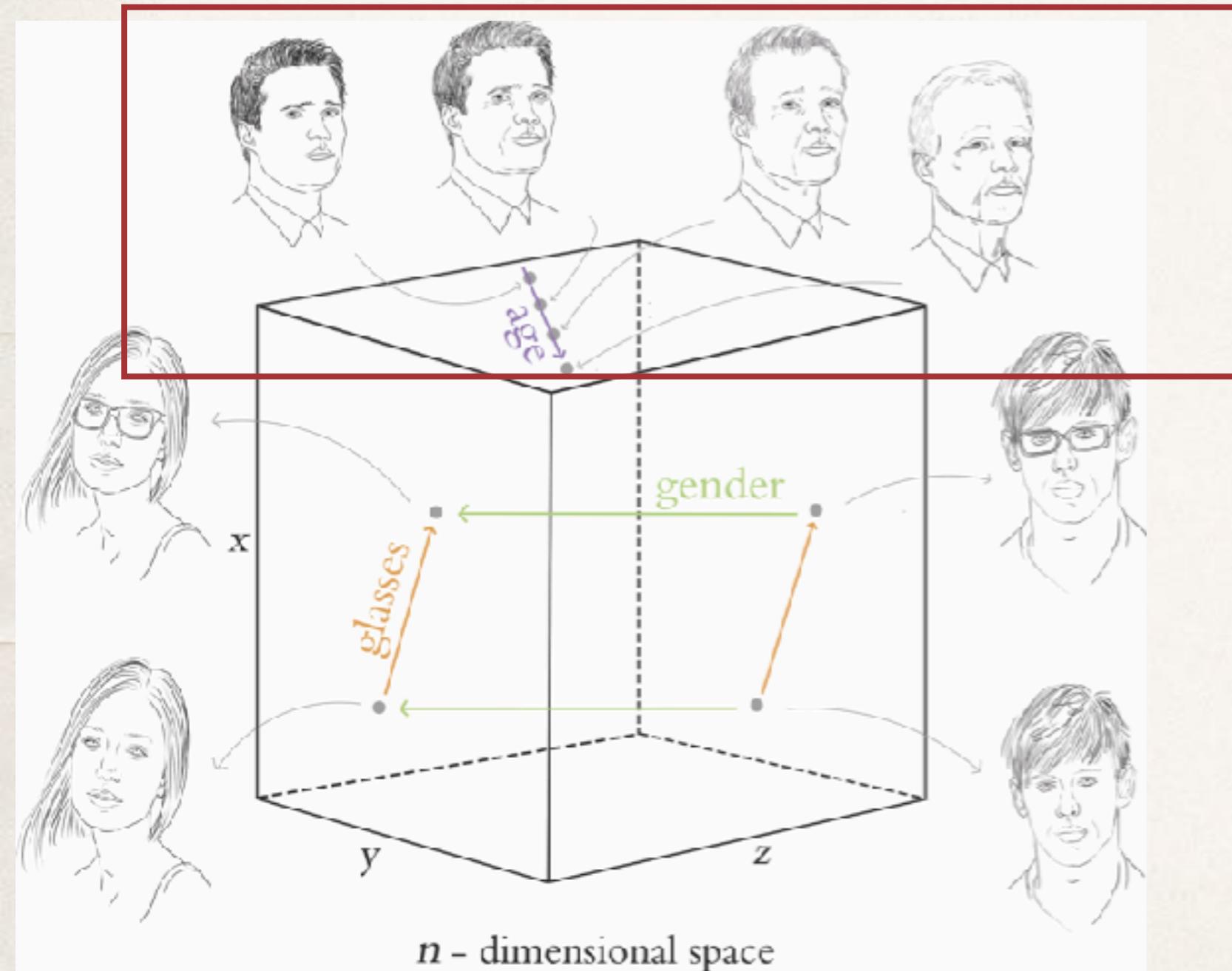


Falta  
de Agua | Fuga  
de Agua

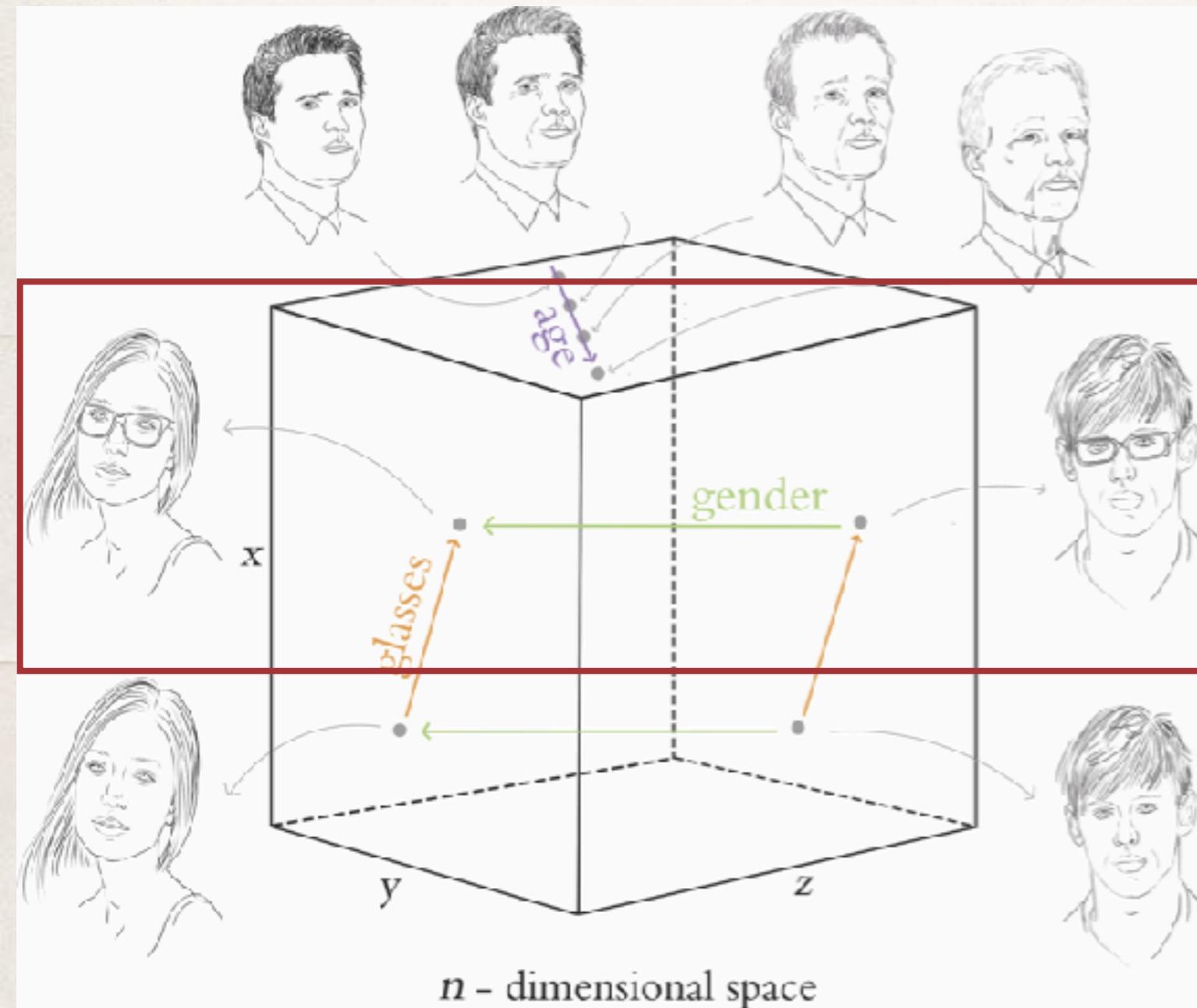
# Representaciones distribuidas



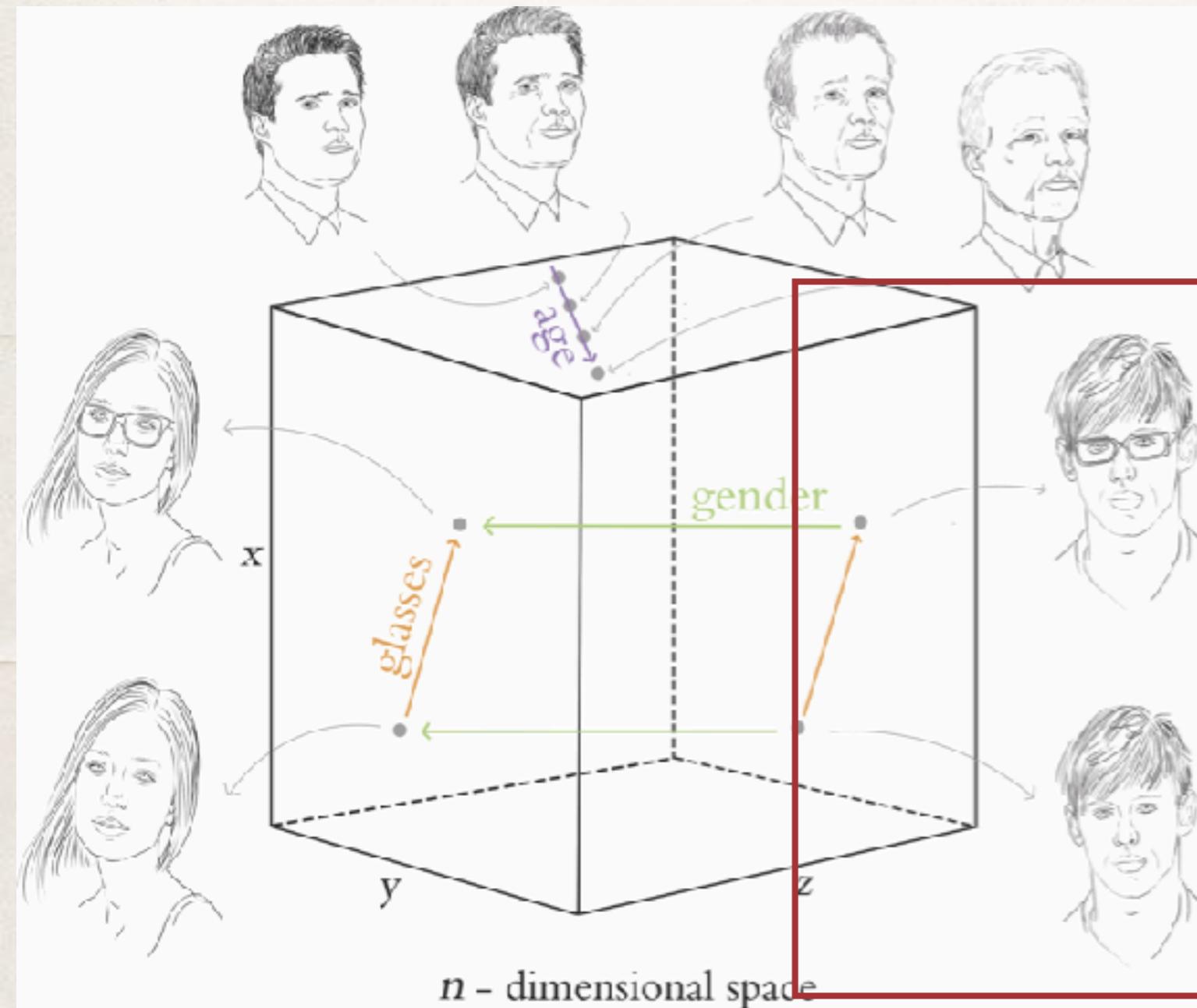
# Representaciones distribuidas



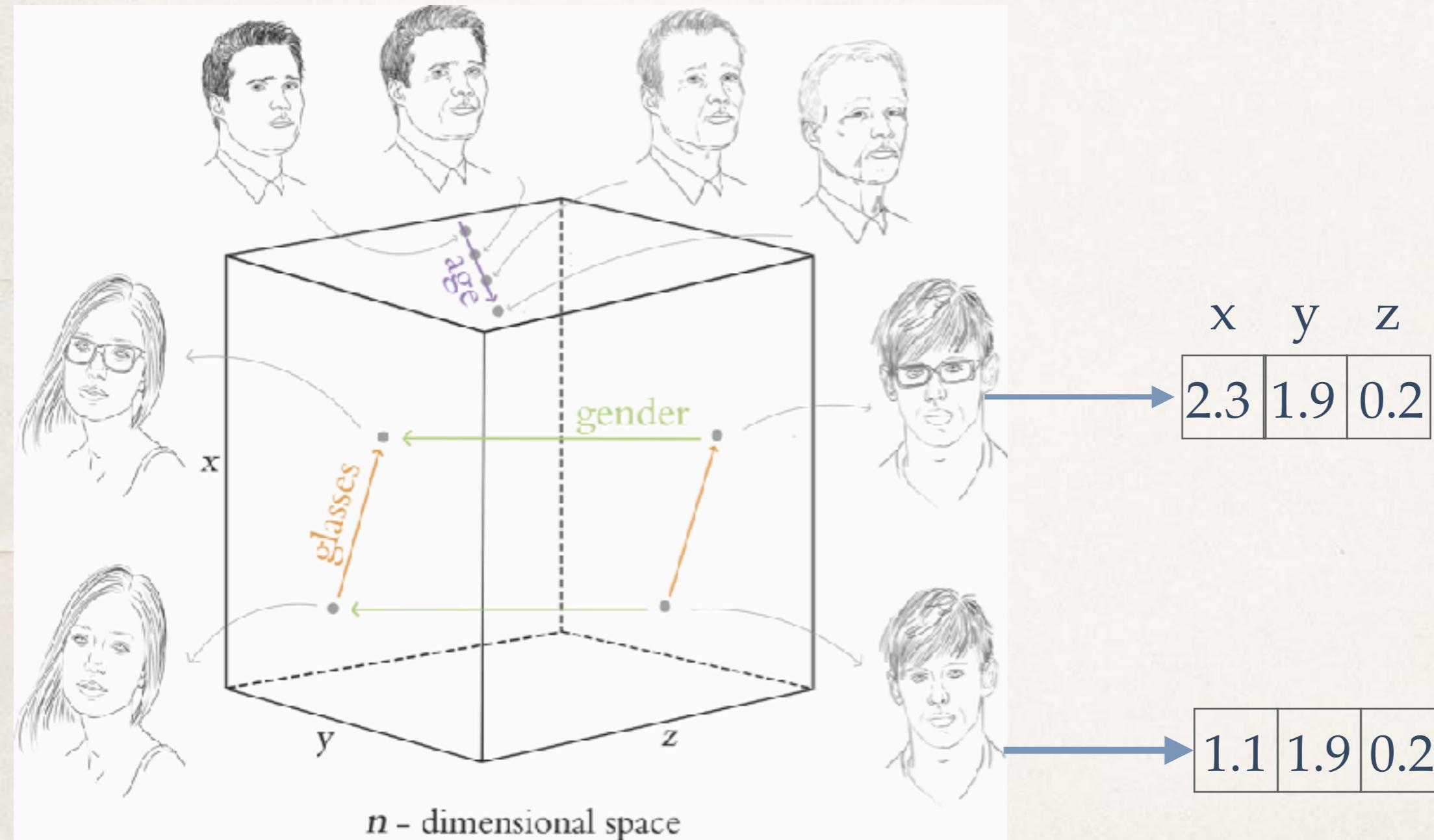
# Representaciones distribuidas



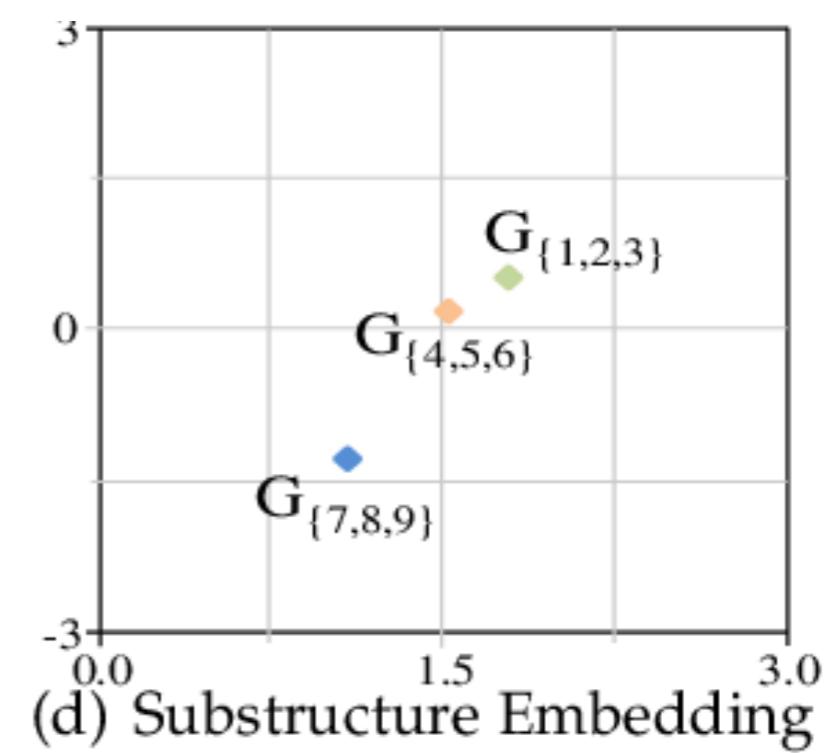
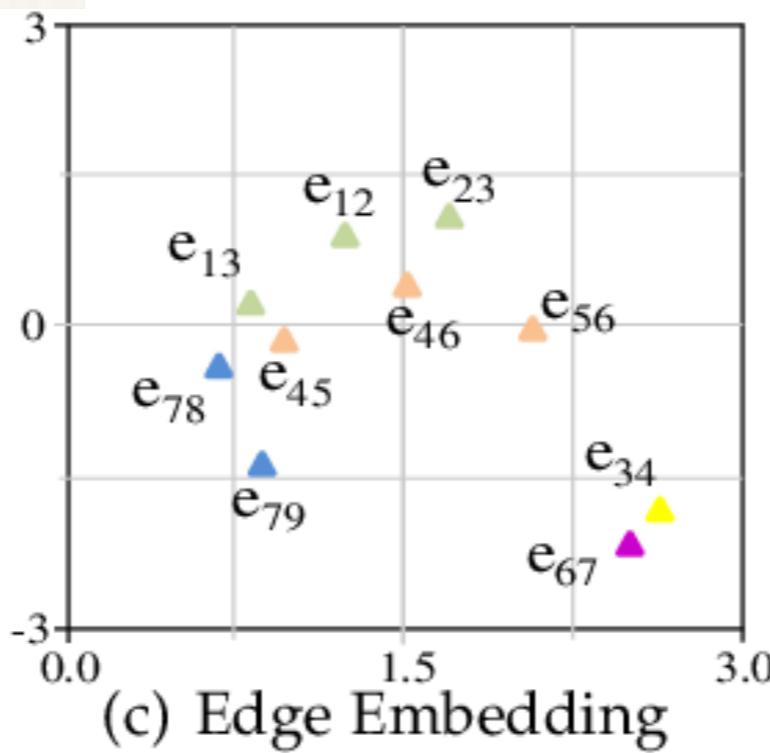
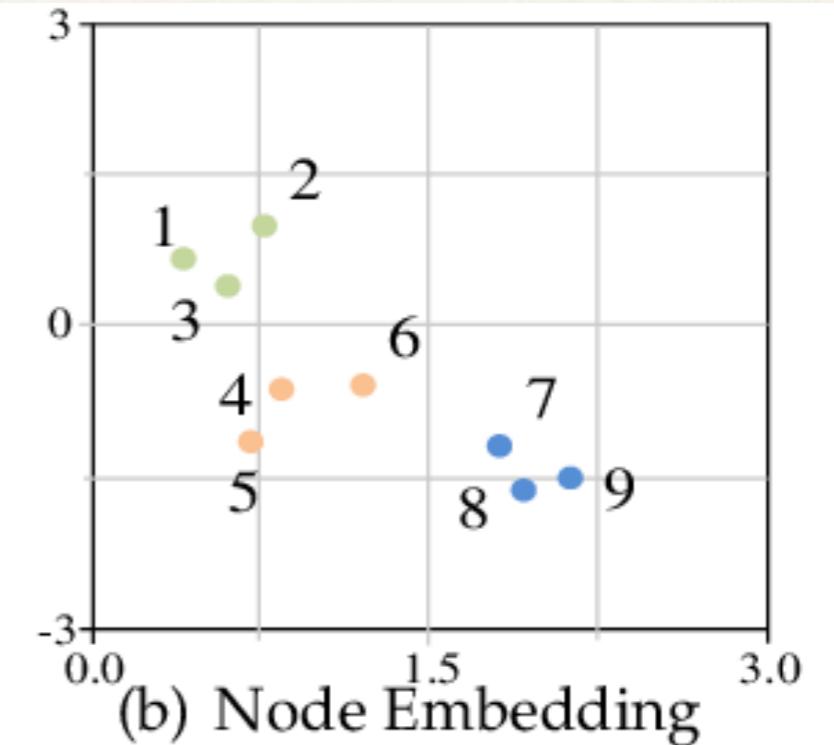
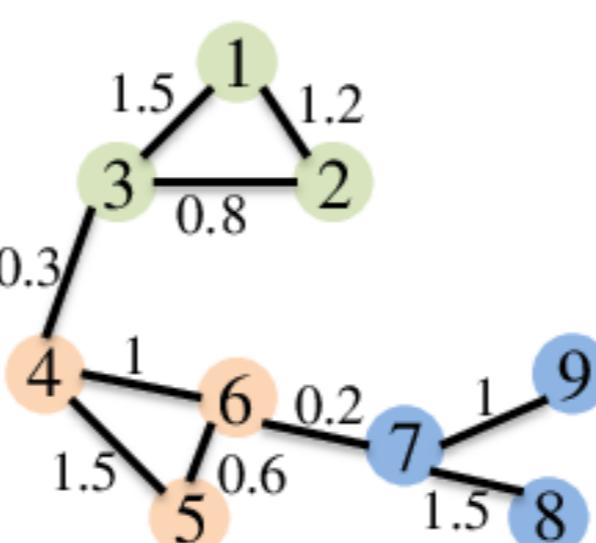
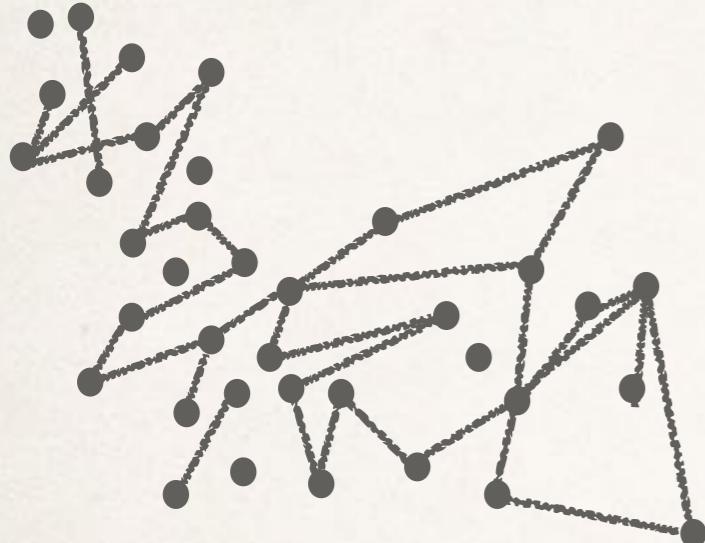
# Representaciones distribuidas



# Representaciones distribuidas



# Graph Embeddings



# Graph2vec

Inspirado en Doc2vec

Grafo completo -> documento

Subgrafos -> palabras

Aprendizaje No supervisado

Representación Agnóstica

## graph2vec: Learning Distributed Representations of Graphs

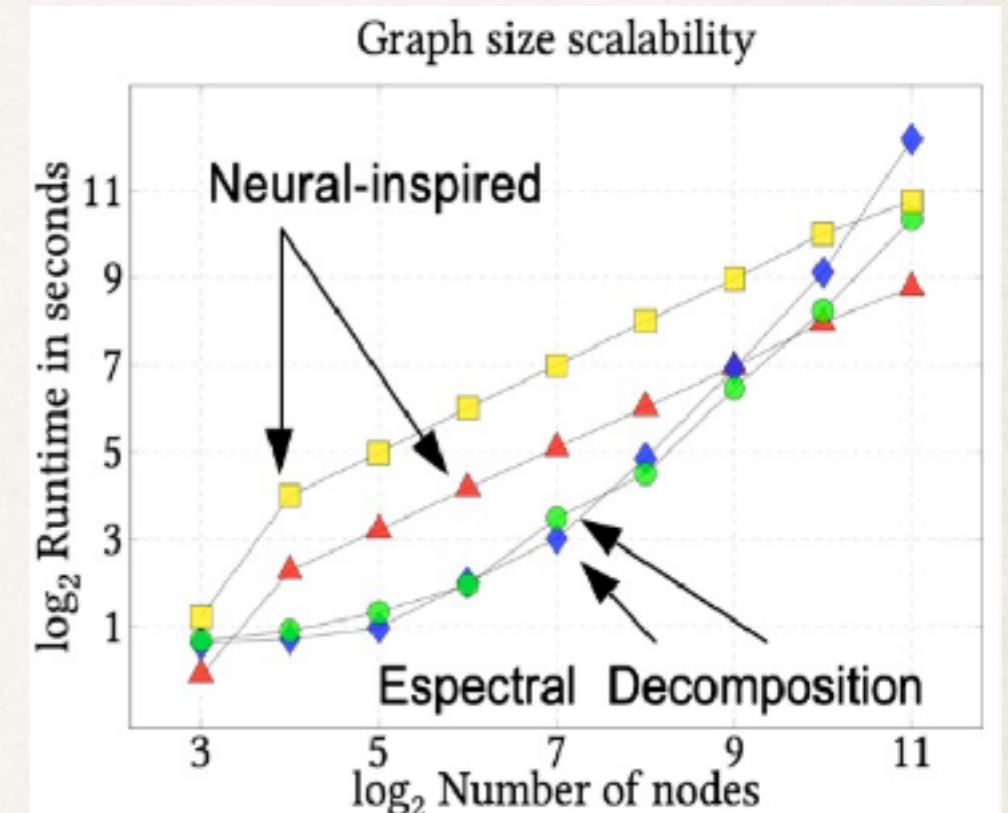
Annamalai Narayanan, Mahinthan Chandramohan, Rajasekar Venkatesan, Lihui Chen, Yang Liu and Shantanu Jaiswal  
Nanyang Technological University, Singapore  
[annamala002@e.ntu.edu.sg](mailto:annamala002@e.ntu.edu.sg),  
 [{mahinthan,rajasekarv,elhchen,yangliu}@ntu.edu.sg](mailto:{mahinthan,rajasekarv,elhchen,yangliu}@ntu.edu.sg), [shantanu004@e.ntu.edu.sg](mailto:shantanu004@e.ntu.edu.sg)

### ABSTRACT

Recent works on representation learning for graph structured data predominantly focus on learning distributed representations of graph substructures such as nodes and subgraphs. However, many graph analytics tasks such as graph classification and clustering require representing entire graphs as fixed length feature vectors. While the aforementioned approaches are naturally unequipped to learn such representations, graph kernels remain as the most ef-

malware [6] and those of chemical compounds could be used to predict their properties such as solubility and anti-cancer activity [7].

**Graph Kernels and handcrafted features.** Graph kernels are one of the most prominent ways of catering the aforementioned graph analytics tasks. Graph kernels evaluate the similarity (aka kernel value) between a pair of graphs  $G$  and  $G'$  by recursively decomposing them into atomic substructures (e.g., random walks, shortest paths,



# Preprocesamiento

Freq muestreo [0-10]/min

Datos de 1 año del sistema SCADA (Mayo 2018, Mayo 2019)



Datos de tanques -> Nivel



La granularidad es de 1 hora

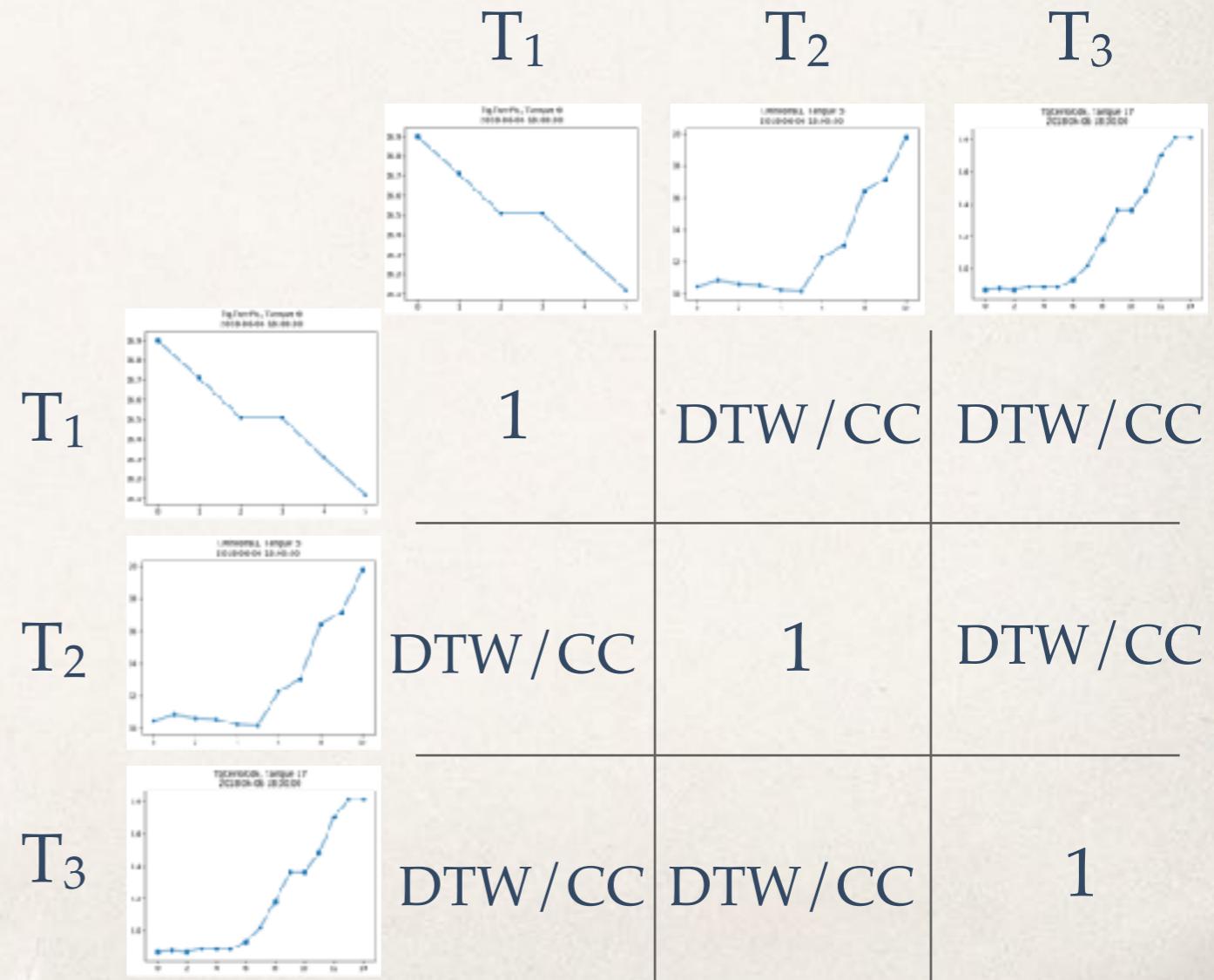
5592 grafos

Grafo

->Matriz de Adyacencia

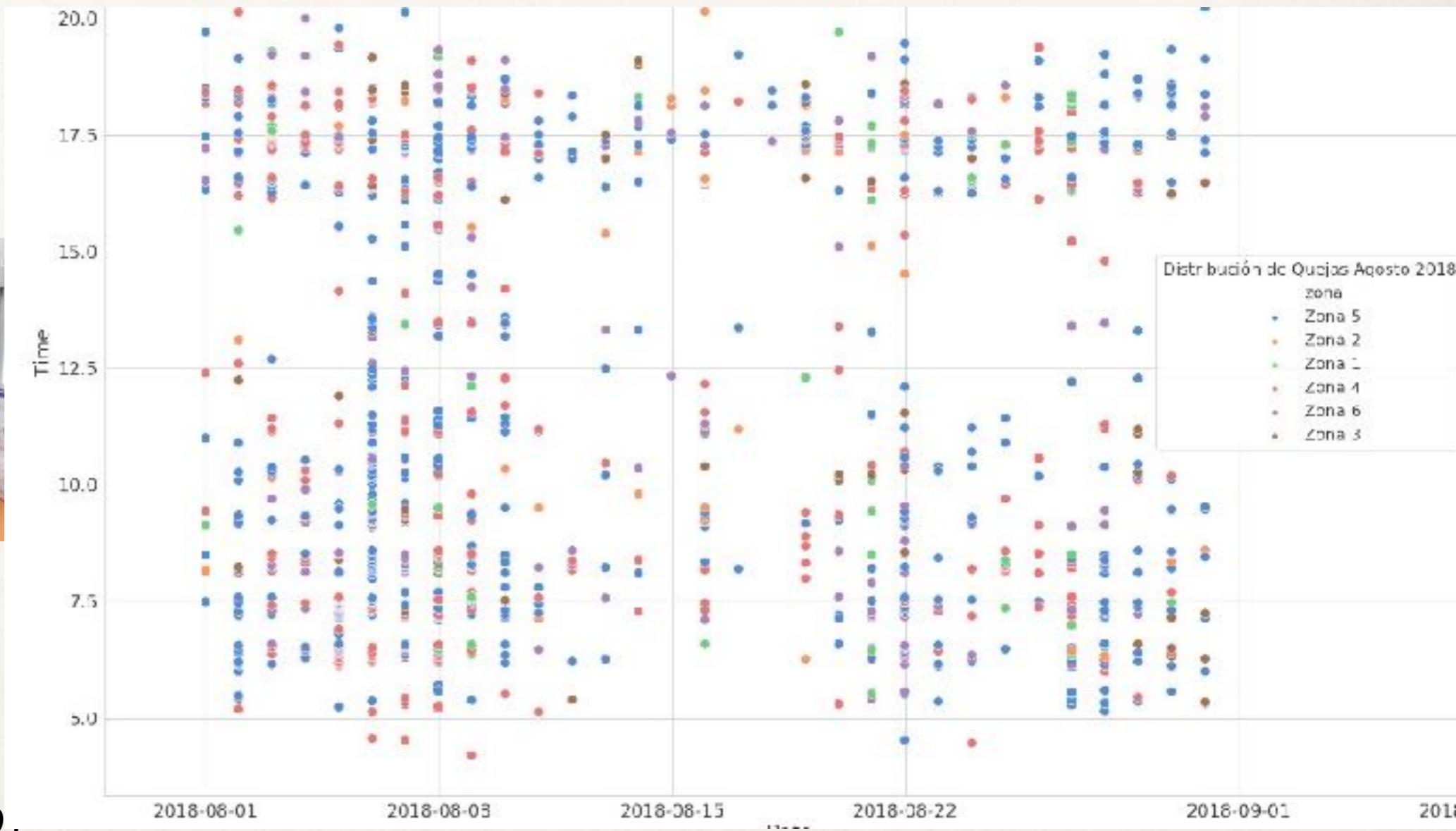
Dynamic Time Warping (DTW)

Coeficiente de Correlación (CC)



# ¿y las etiquetas de los datos?

## Centro de Información y Servicio (CIS)

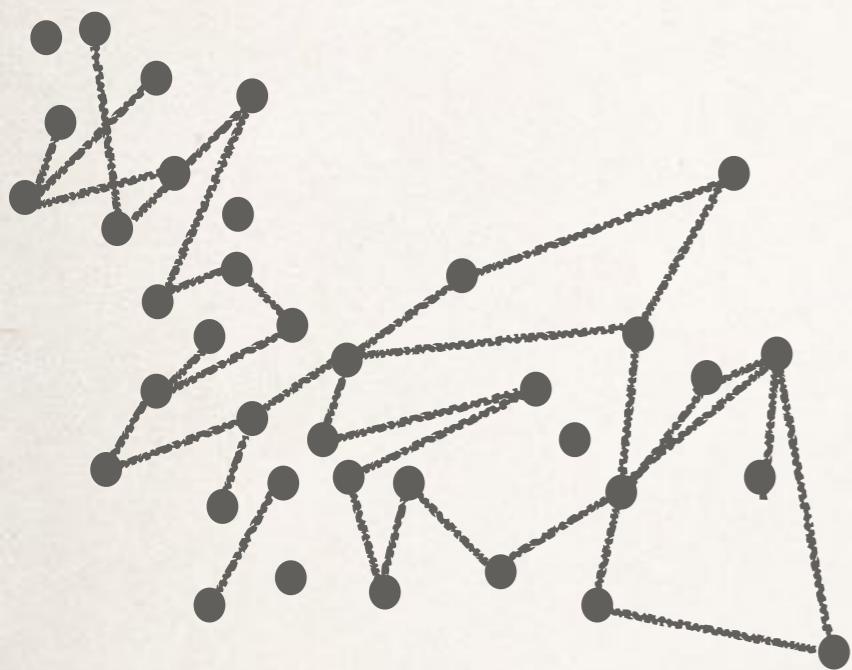


(Mayo 2018, Mayo 2019,

# Fotografía completa

SCADA

DTW;CC



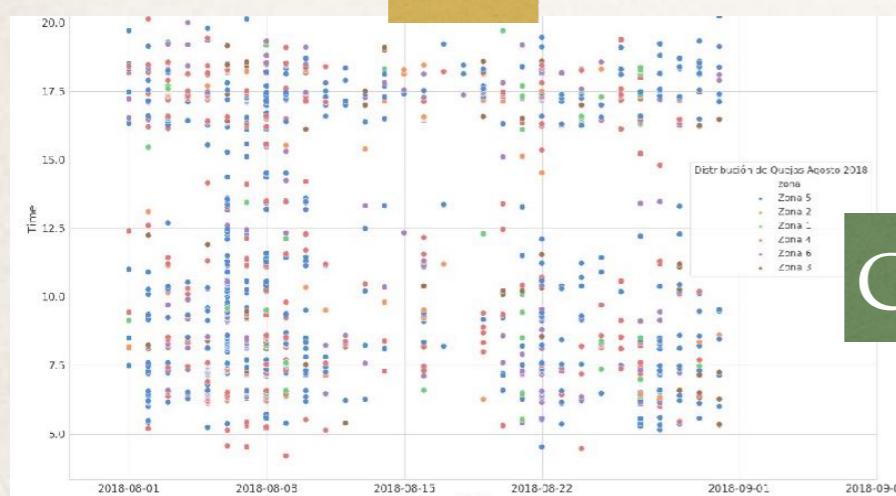
Graph2vec

Método  
Computacional

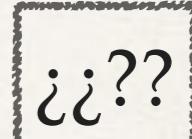


SVM, RF

Clustering



No  
Problema



Falta  
de Agua | Fuga  
de Agua

Fuga  
de Agua

# Data Science (our two cents)

---



Diseño de la base de datos  
Ruido  
Inconsistencias  
Campos con mismo nombre  
Versiones de los datos  
Falta de documentación



Librerías muy soportadas



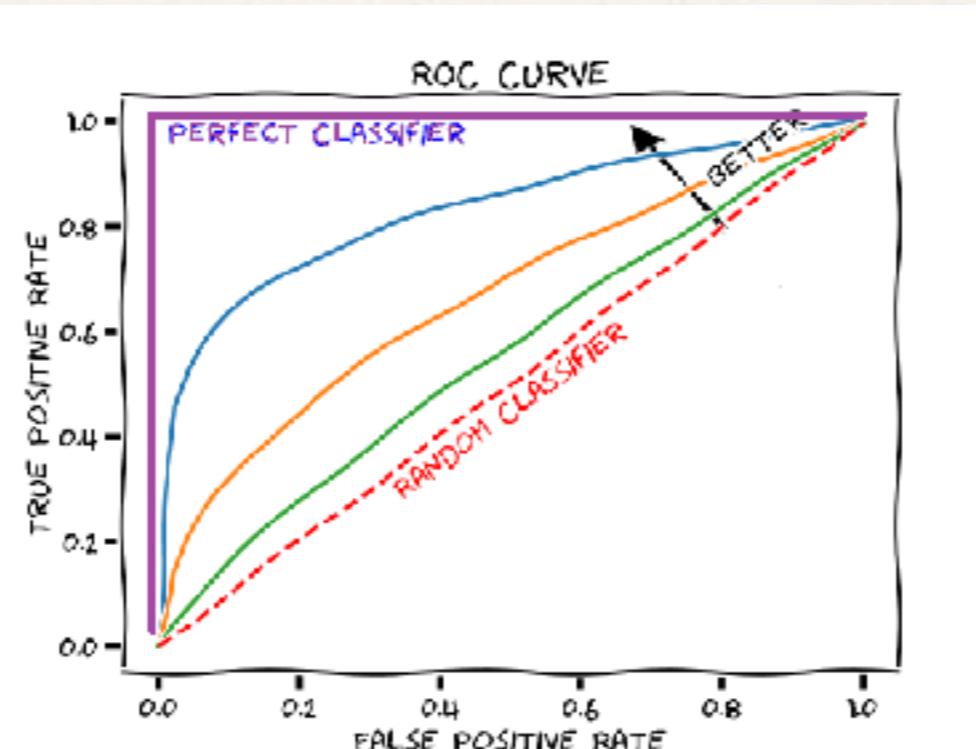
# Resultados preliminares

$$ACC = \frac{tp + tn}{tp + fp + tn + fn}$$

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

$$precision = \frac{TP}{TP + FP}$$

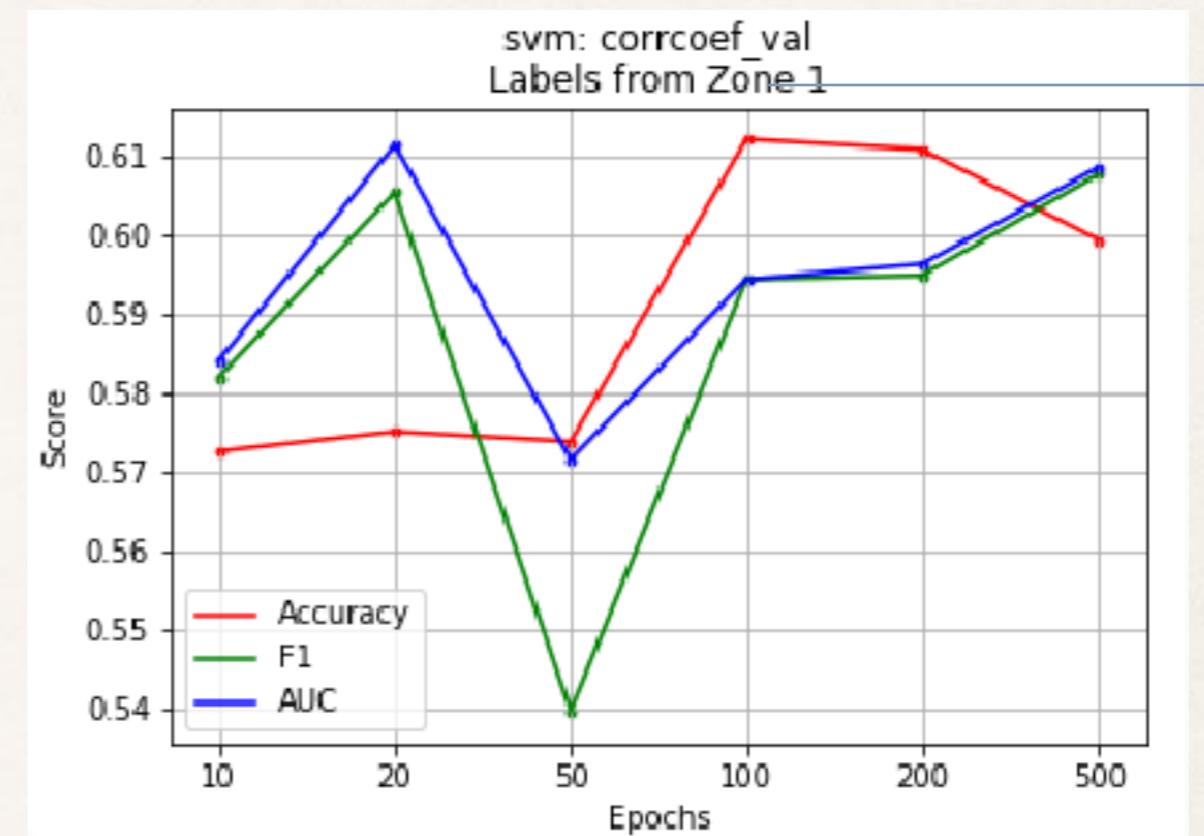
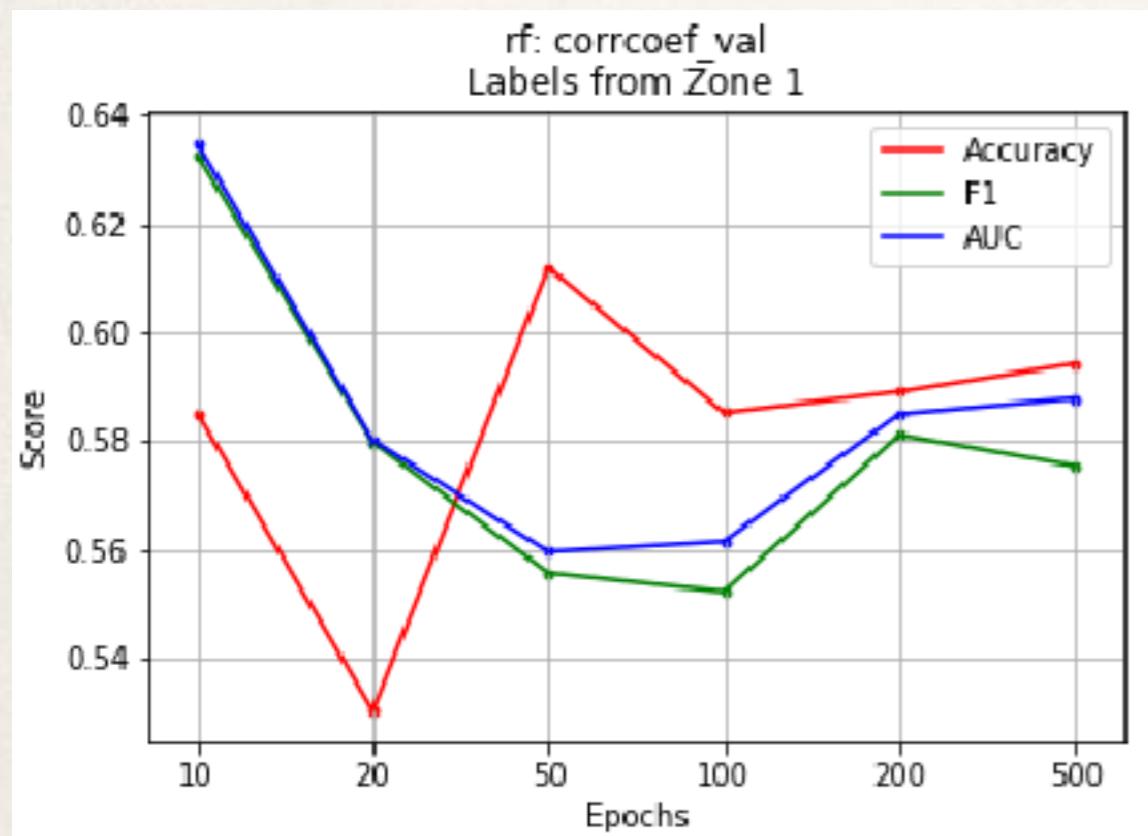
$$recall = \frac{TP}{TP + FN}$$



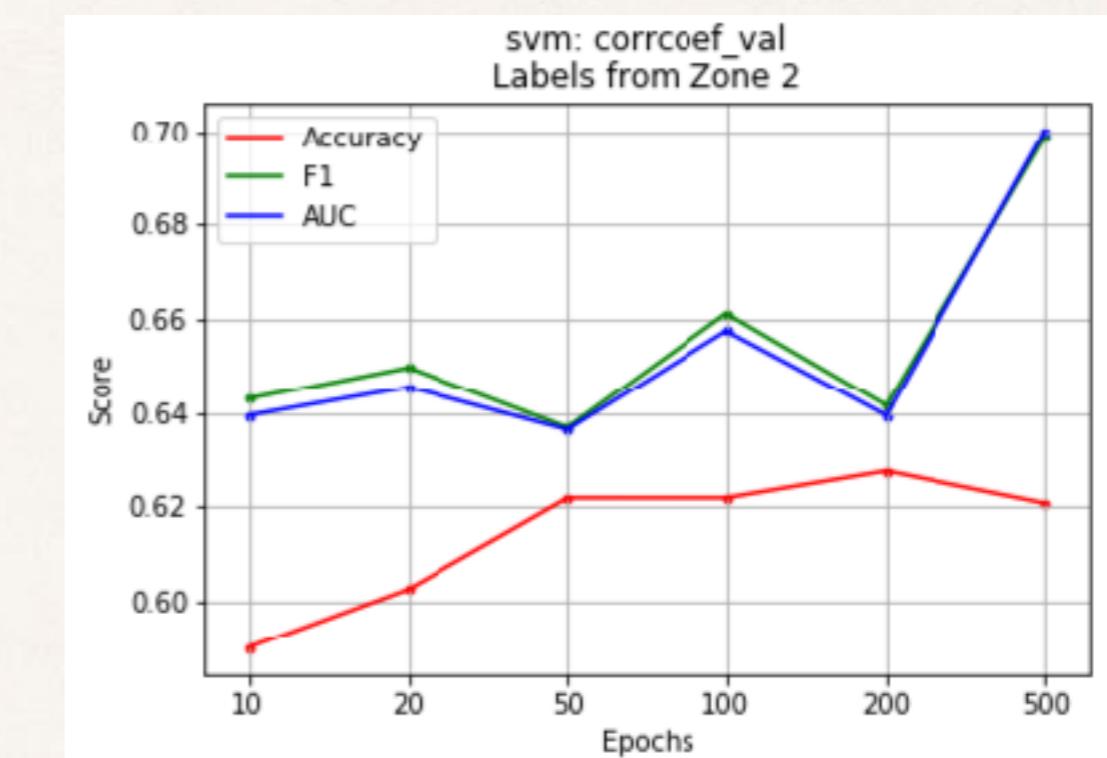
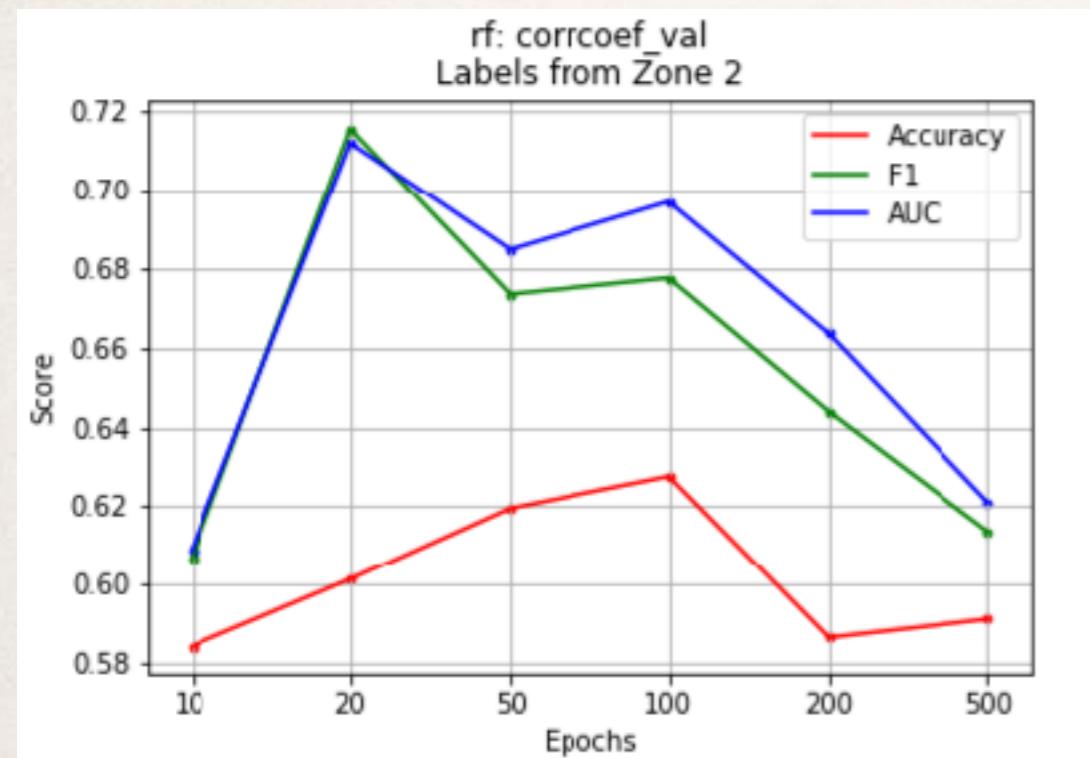
90% clase 0  
10% clase 1

Subsampling  
5 FCV

# Resultados preliminares

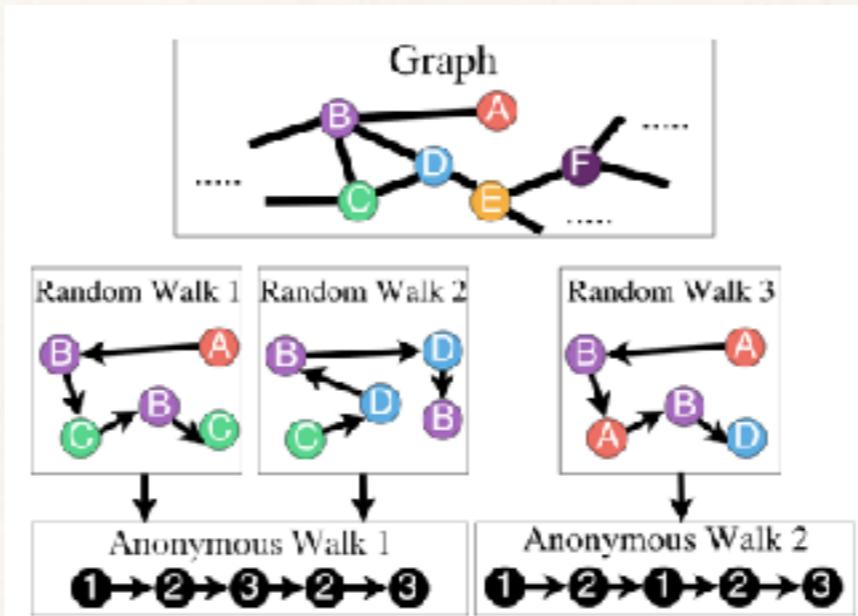


# Resultados preliminares



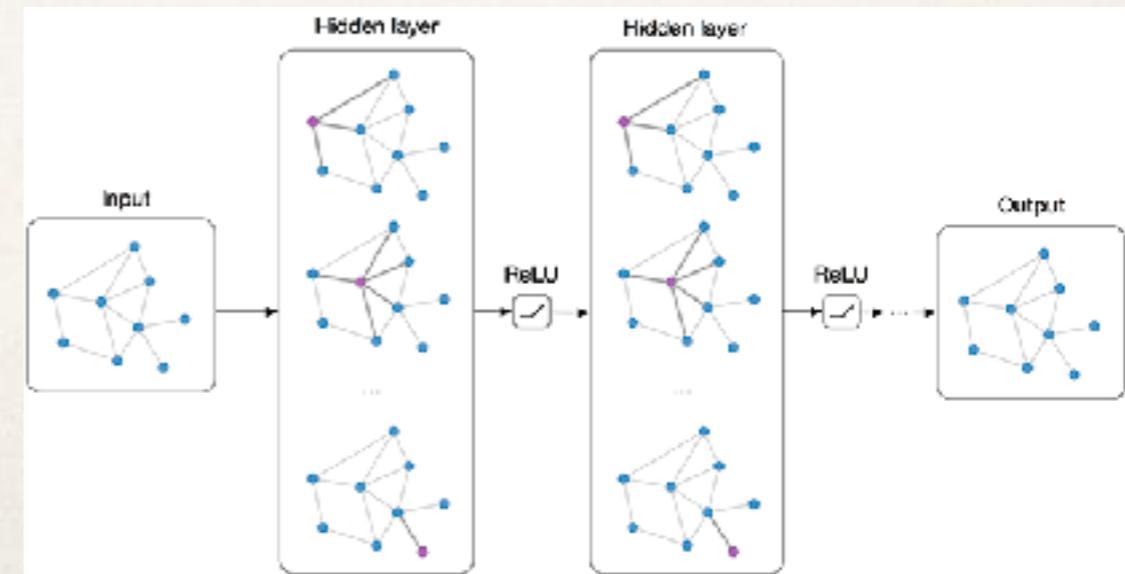
# ¿Qué más estamos haciendo?

- ✿ Anonymous Walks



- ✿ Métodos basados en descomposición espectral

- ✿ Graph Neural Networks



# Smart Water Management via Machine Learning



UNIVERSIDAD AUTÓNOMA DE  
**CHIHUAHUA**

**THANKS**

