



**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación



El futuro  
es de todos

DNP  
Departamento  
Nacional de Planeación

# Rastreo de inversión en alineación con los ODS

**Unidad de Científicos de Datos**  
Dirección de Desarrollo Digital

Abril, 2020

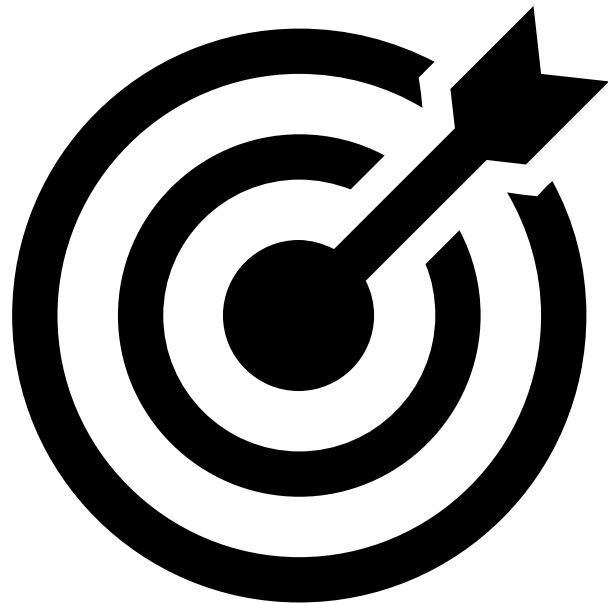


- 1. Descripción del proyecto**
- 2. Metodología**
- 3. Resultados y conclusiones**

# 1. Descripción del proyecto



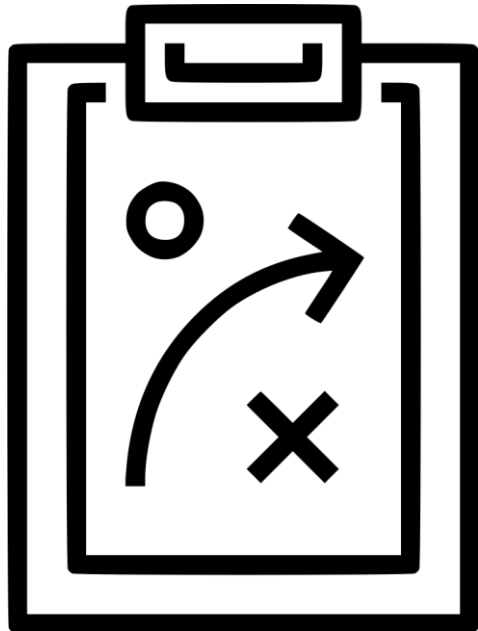
# Objetivo



Analizar la inversión pública realizada en alineación con los Objetivos de Desarrollo Sostenible (ODS) a través de la aplicación de analítica de datos.

# ¿Qué información se utilizó?

1

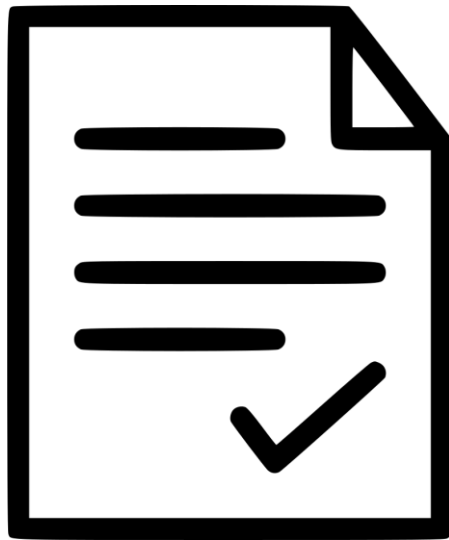


## Textos sobre los ODS

- Palabras clave (proyecto 2018)
- Generalidades y datos destacables de cada ODS (ONU)
- Cartillas sobre “¿Por qué es importante cada ODS?” (ONU)

# ¿Qué información se utilizó?

2



## Proyectos de inversión

- SUIFP PGN:  
Recursos PGN (2020)
- SUIFP SGR: Recursos  
SGR y territoriales (2020)
- APC: Recursos de Cooperación  
Internacional (2018)

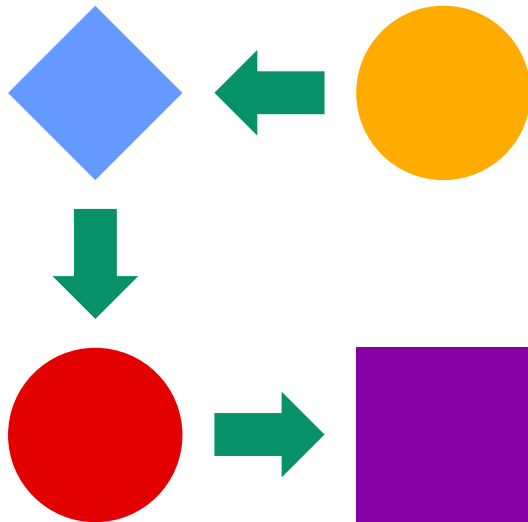


# 2. Metodología





# Metodología



1

Técnicas de procesamiento y vectorización de textos

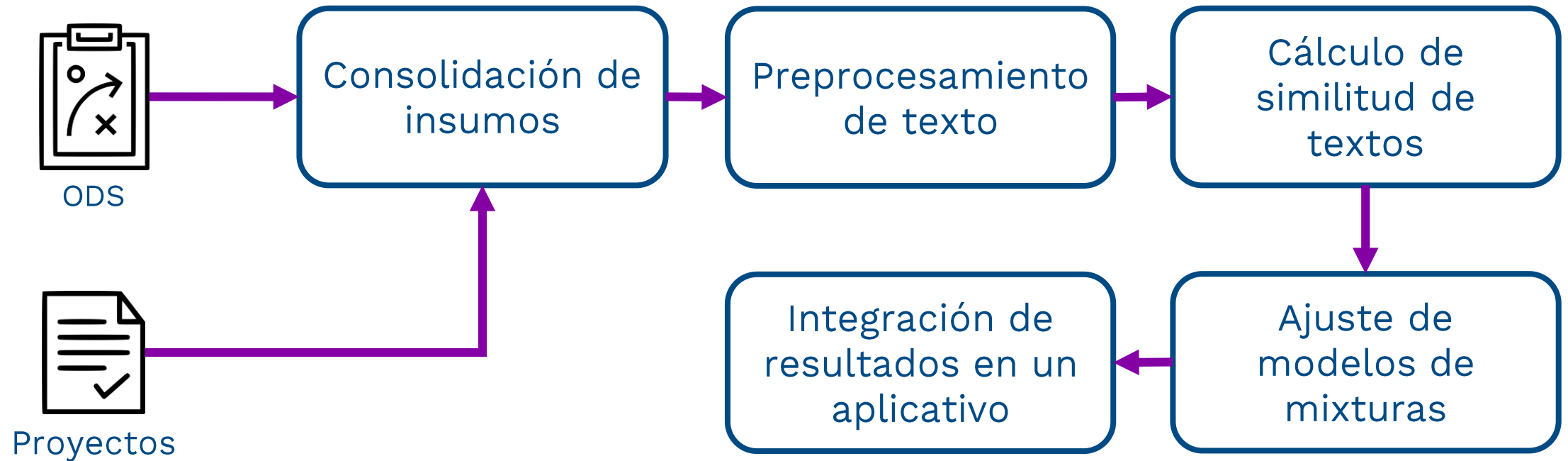
2

Cálculo de similitud entre textos de proyectos y de ODS

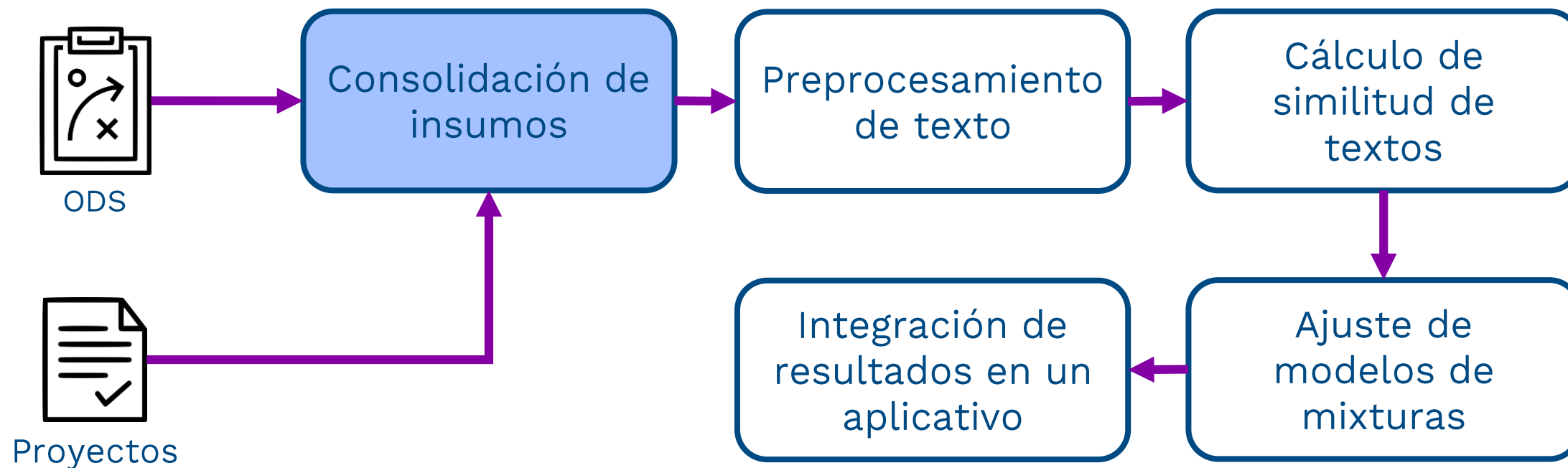
3

Clasificación automática de los proyectos

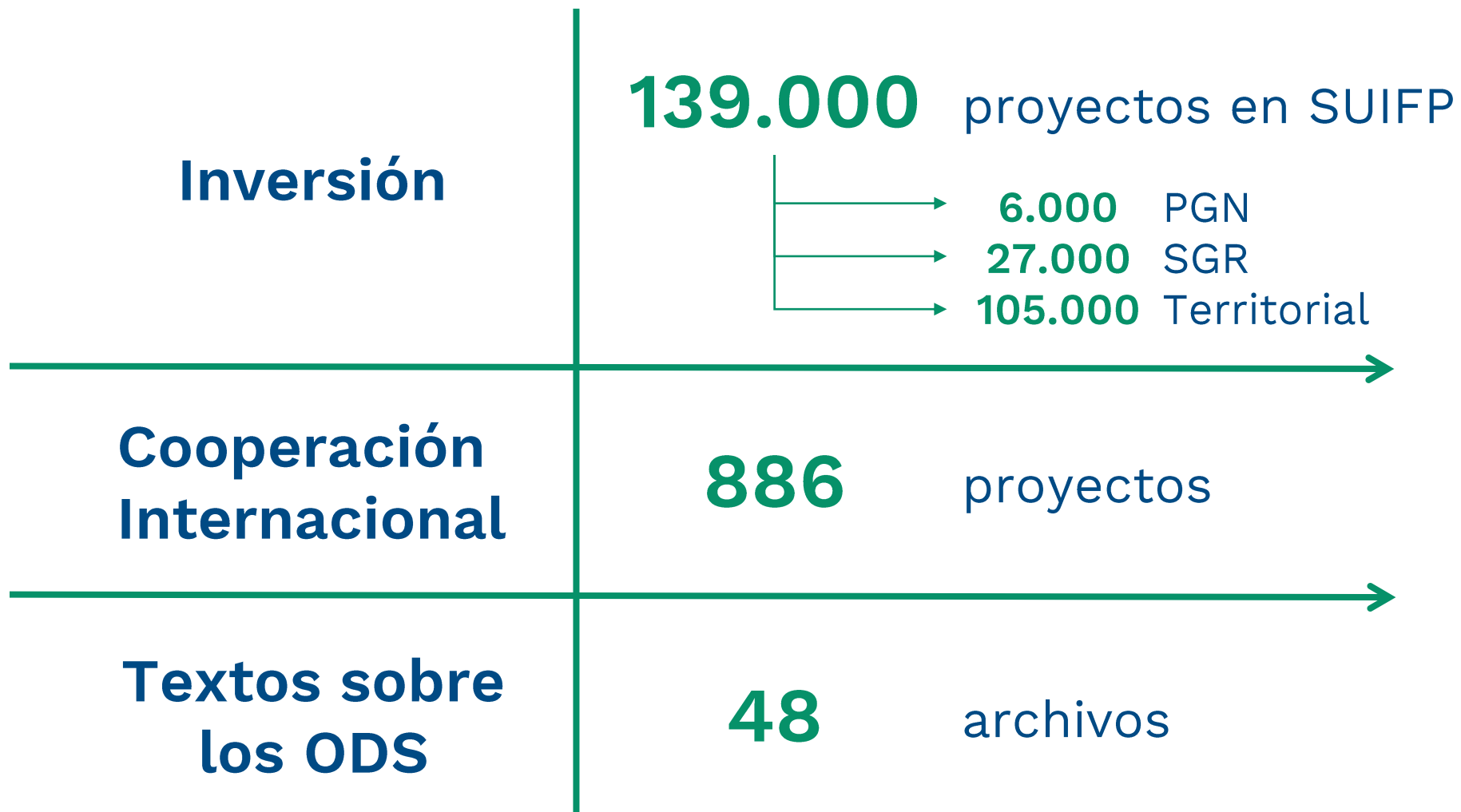
# Vista general del proyecto



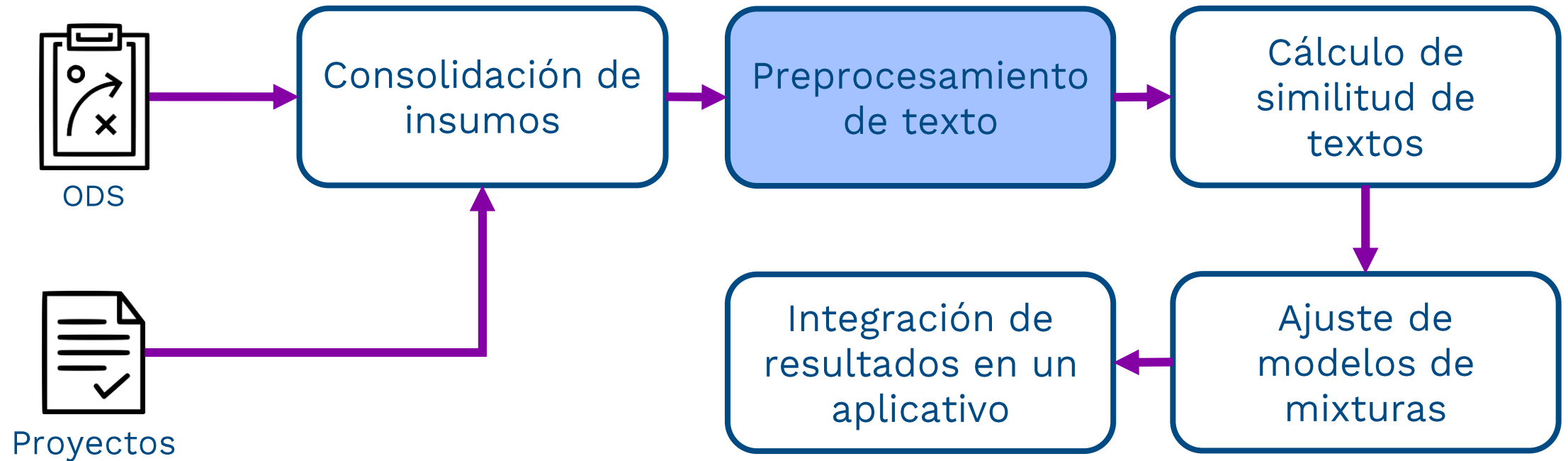
# Vista general del proyecto



# Consolidación de insumos



# Vista general del proyecto



# Preprocesamiento de texto



1

Eliminación de números, signos de puntuación y caracteres especiales

2

Conversión a minúsculas

3

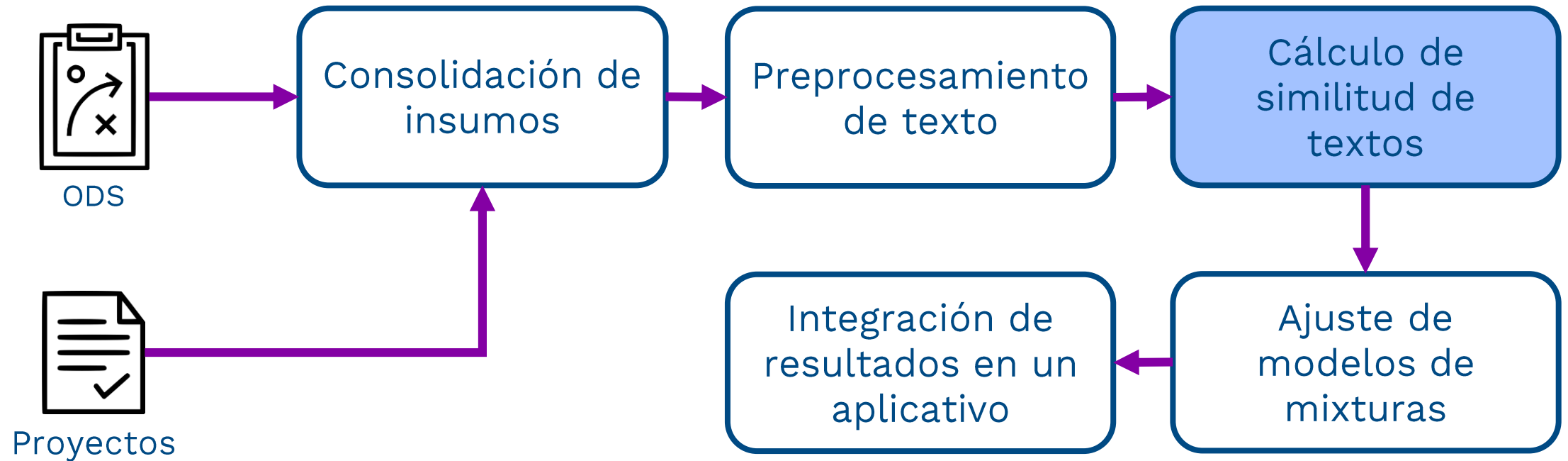
Eliminación de *stopwords*

Términos irrelevantes para el análisis (proyecto, alcaldía, etc.)

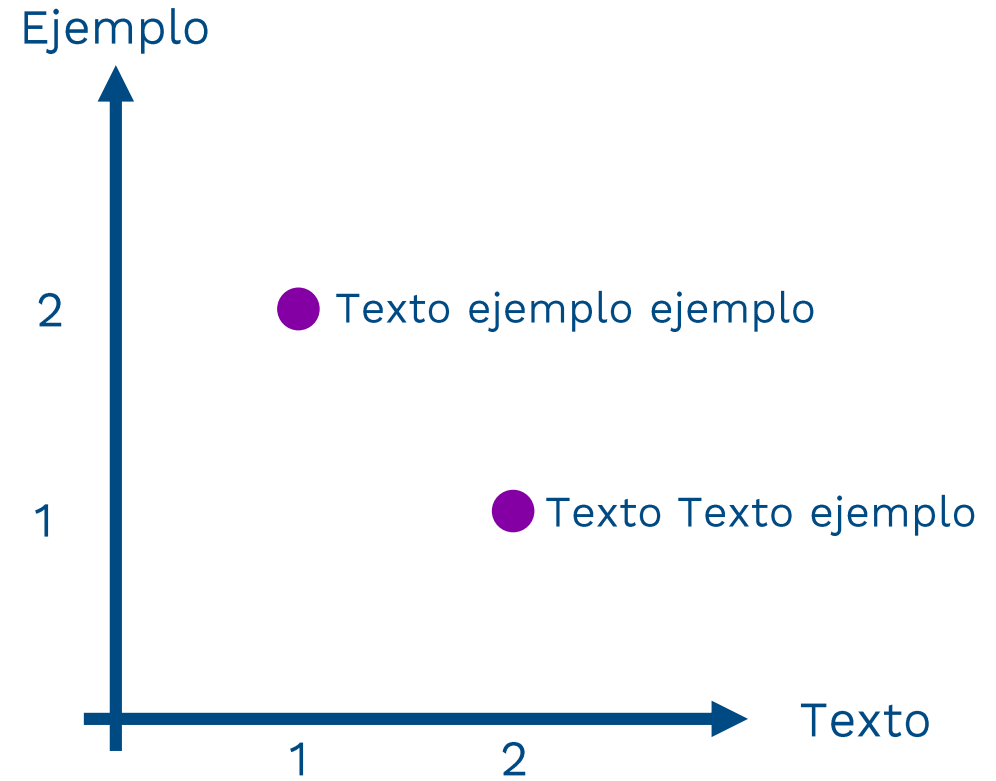
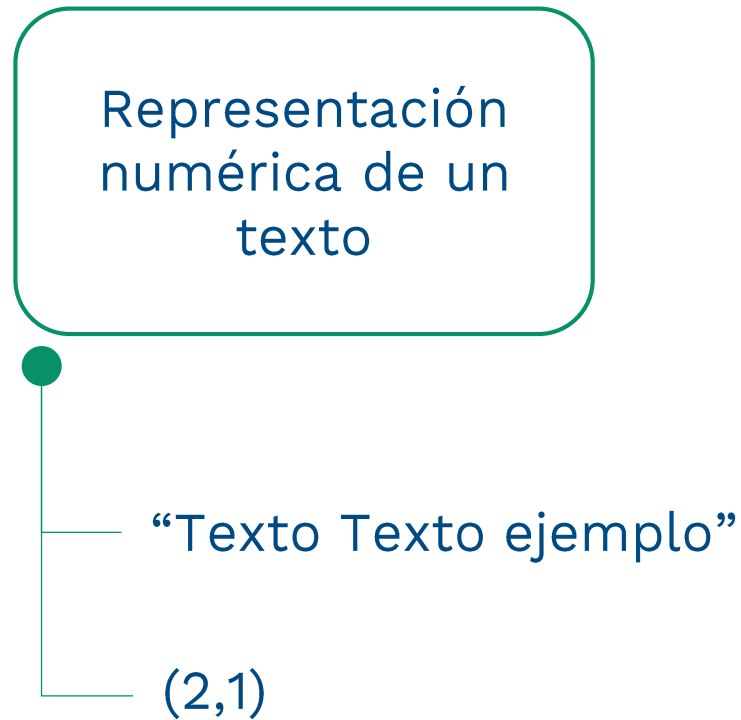
Palabras que no agregan significado (conectores, etc.)



# Vista general del proyecto



# Vectorización del texto





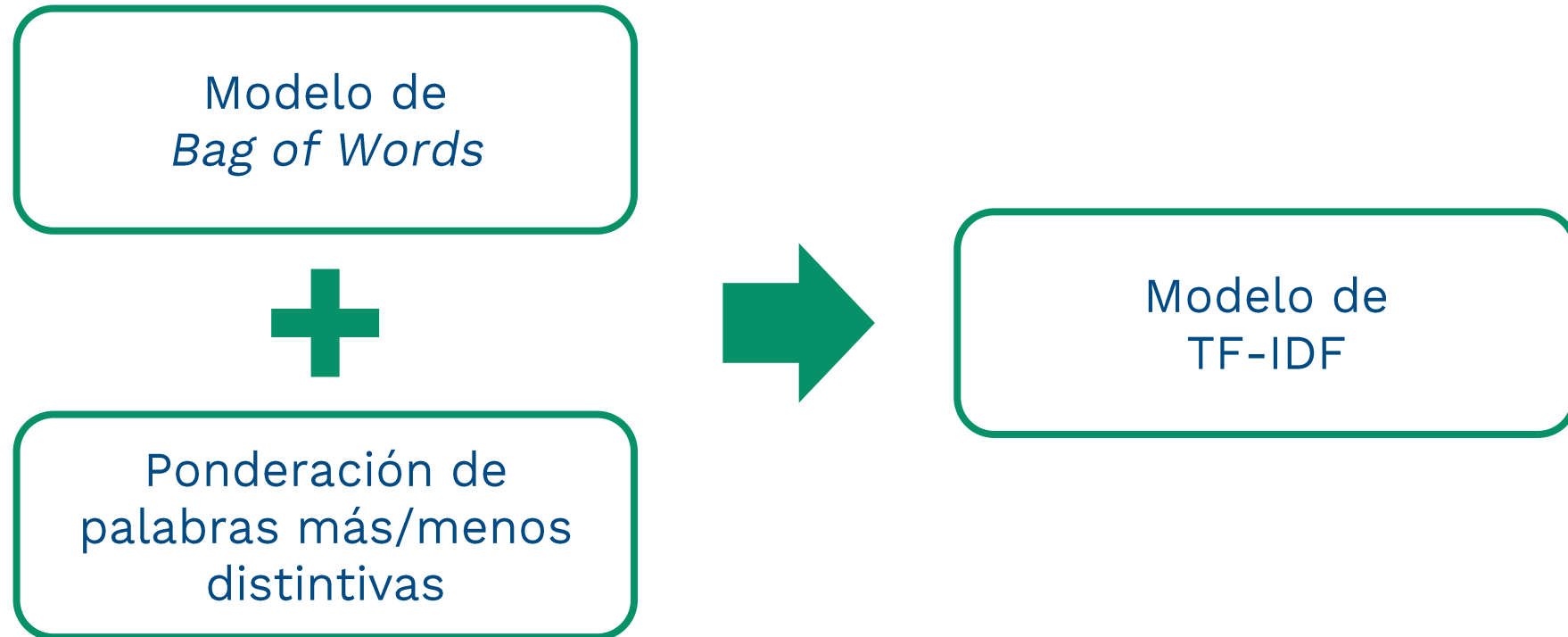
# Modelo de *Bag of Words* (BOW)

	un	texto	ejemplo	otro	de	más
un texto ejemplo	1/3	1/3	1/3	0	0	0
otro texto ejemplo	0	1/3	1/3	1/3	0	0
ejemplo de texto	0	1/3	1/3	0	1/3	0
un texto más texto	1/4	2/4	0	0	0	1/4

Frecuencia sobre el número total de palabras:

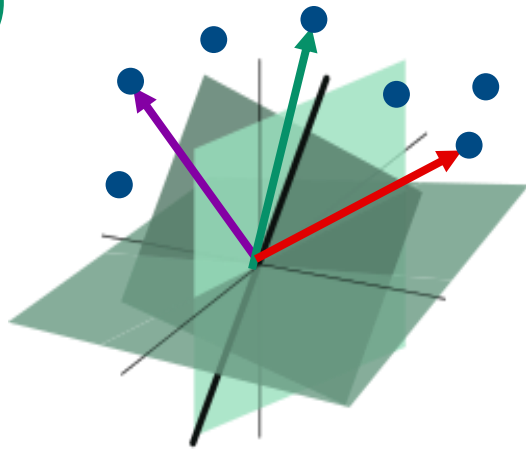
Se puede interpretar como la probabilidad de que una palabra esté en un texto

# Modelo de TF-IDF



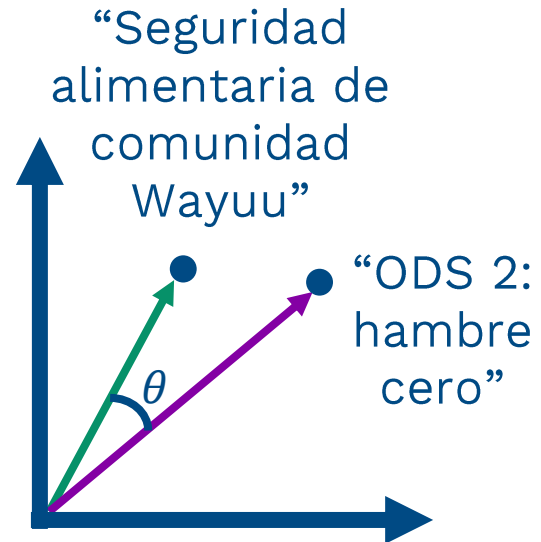
# Medida de similitud

1



Textos vectorizados:  
“Bag of Words” o TF-IDF

2



Se calcula el coseno del ángulo  
entre los vectores ( $\theta$ )

1: totalmente  
idénticos

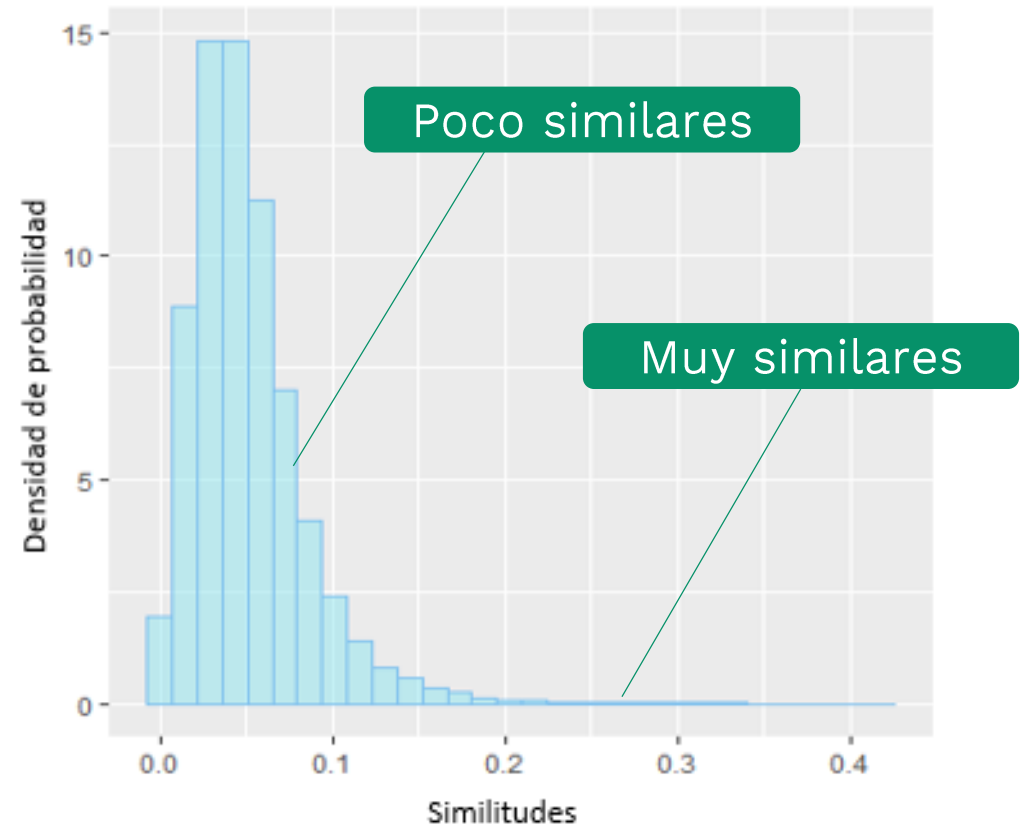
0: totalmente  
diferentes

# Medida de similitud

3

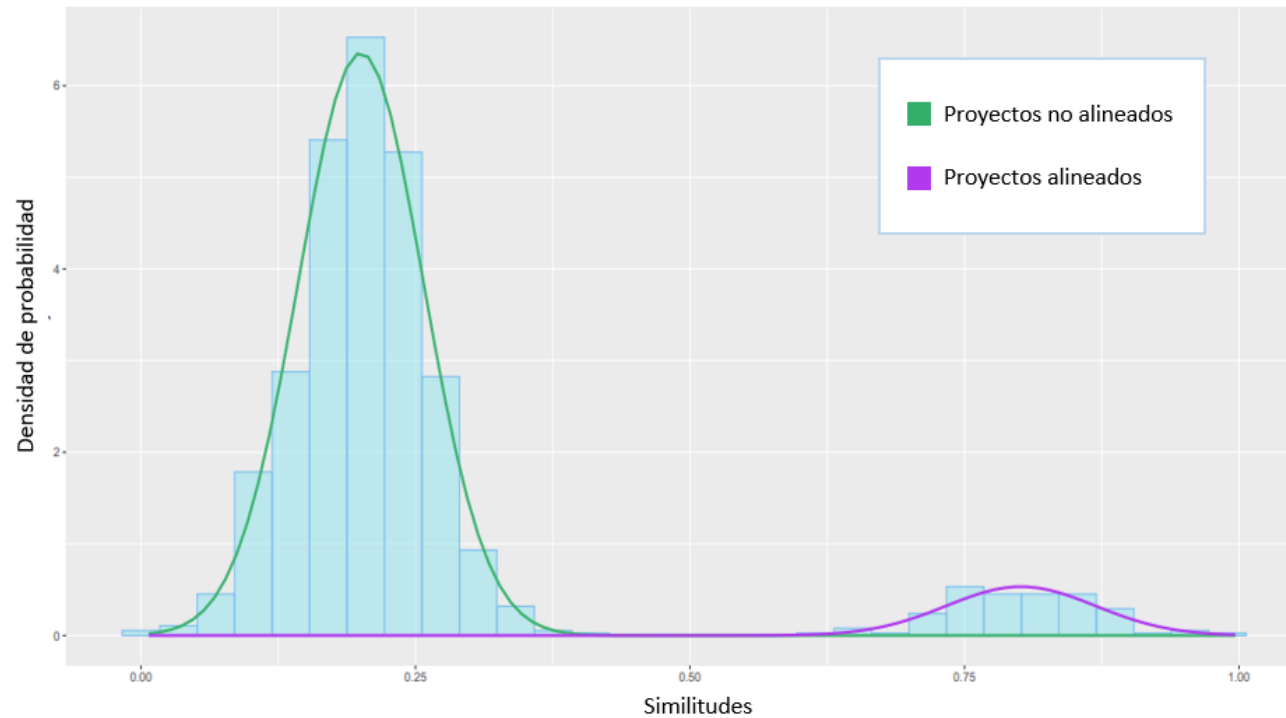
Se calcula el promedio de las medidas:

$$\frac{\text{Similitud}_{\text{BoW}} + \text{Similitud}_{\text{TF-IDF}}}{2}$$



Distribución de la similitud de los proyectos con respecto al ODS 1

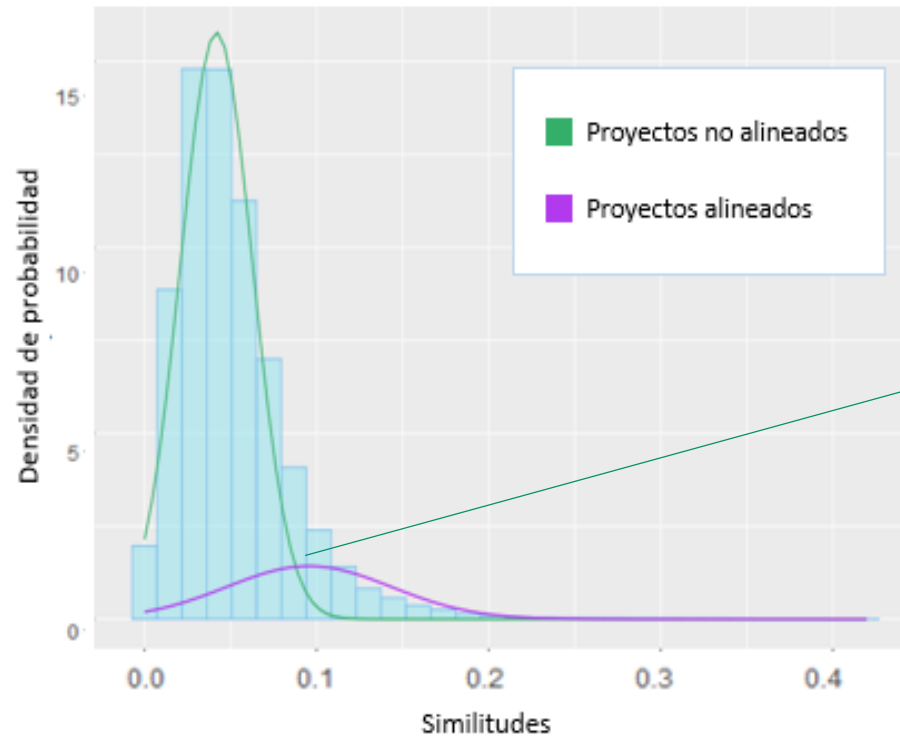
# Definición del umbral



Distribución de la medida de similitud en un escenario **ideal**

Ajuste de modelos de mixturas

# Definición del umbral



Se tomó la intersección como el umbral de clasificación

Distribución de la medida de similitud en el escenario **real**

# 3. Resultados



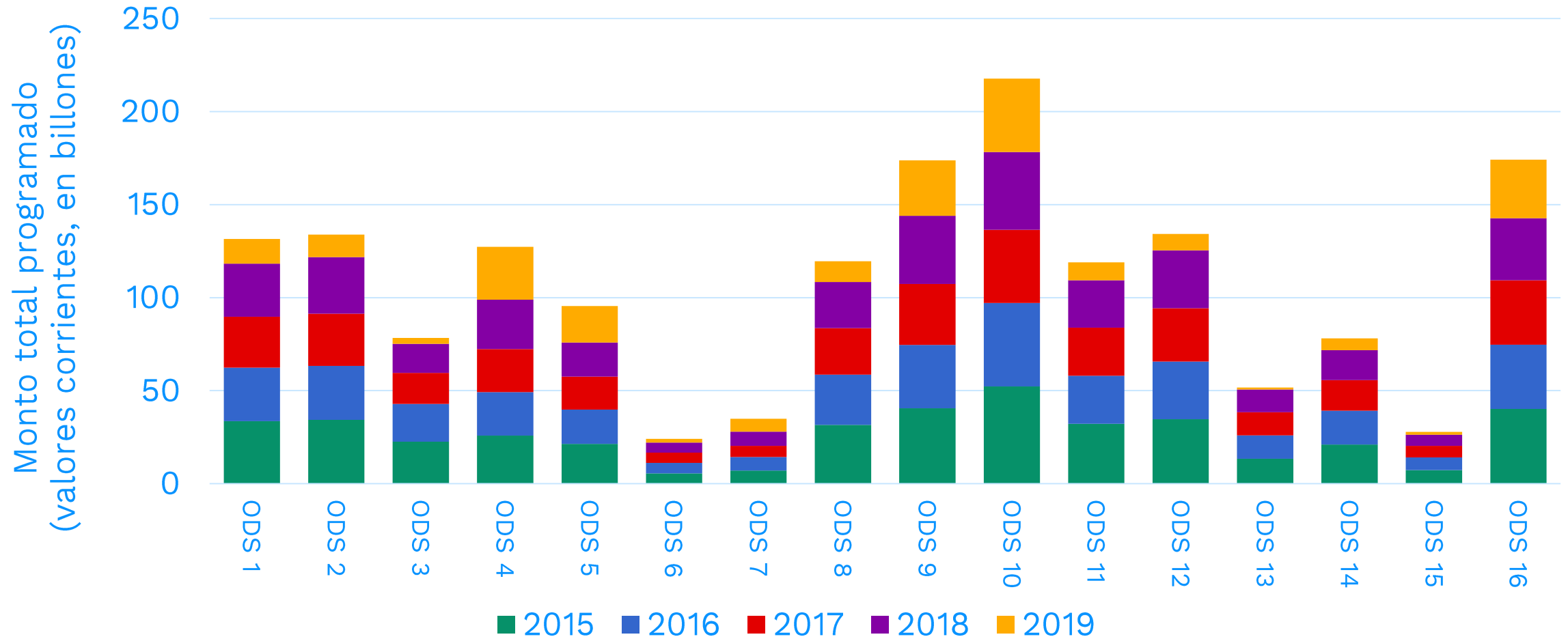
# Proyectos clasificados: ODS 1

Nombre	Similitud ODS 1
Mejoramiento de viviendas como estrategia de erradicación de las privaciones multidimensionales de familias en condición de pobreza extrema y pobreza monetaria en la zona rural y urbana del municipio de Cumaral.	0.42
Apoyo para la superación de la pobreza extrema a familias acompañadas por la ANSPE en el departamento de Arauca.	0.40
Implementación de unidos red de protección social para la superación de la pobreza extrema – nacional.	0.40

Proyectos con mayor puntaje de similitud con respecto al ODS 1

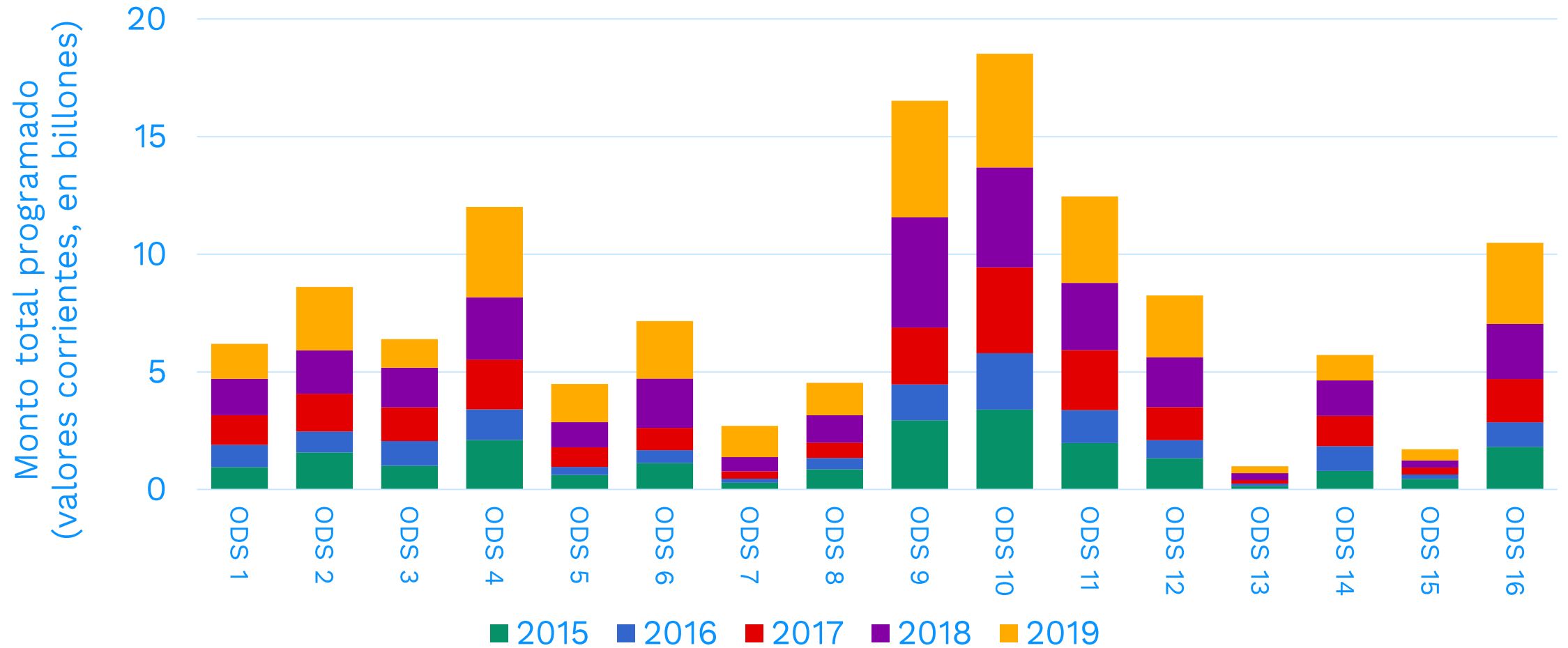


## Monto total programado para los proyectos clasificados con recursos PGN



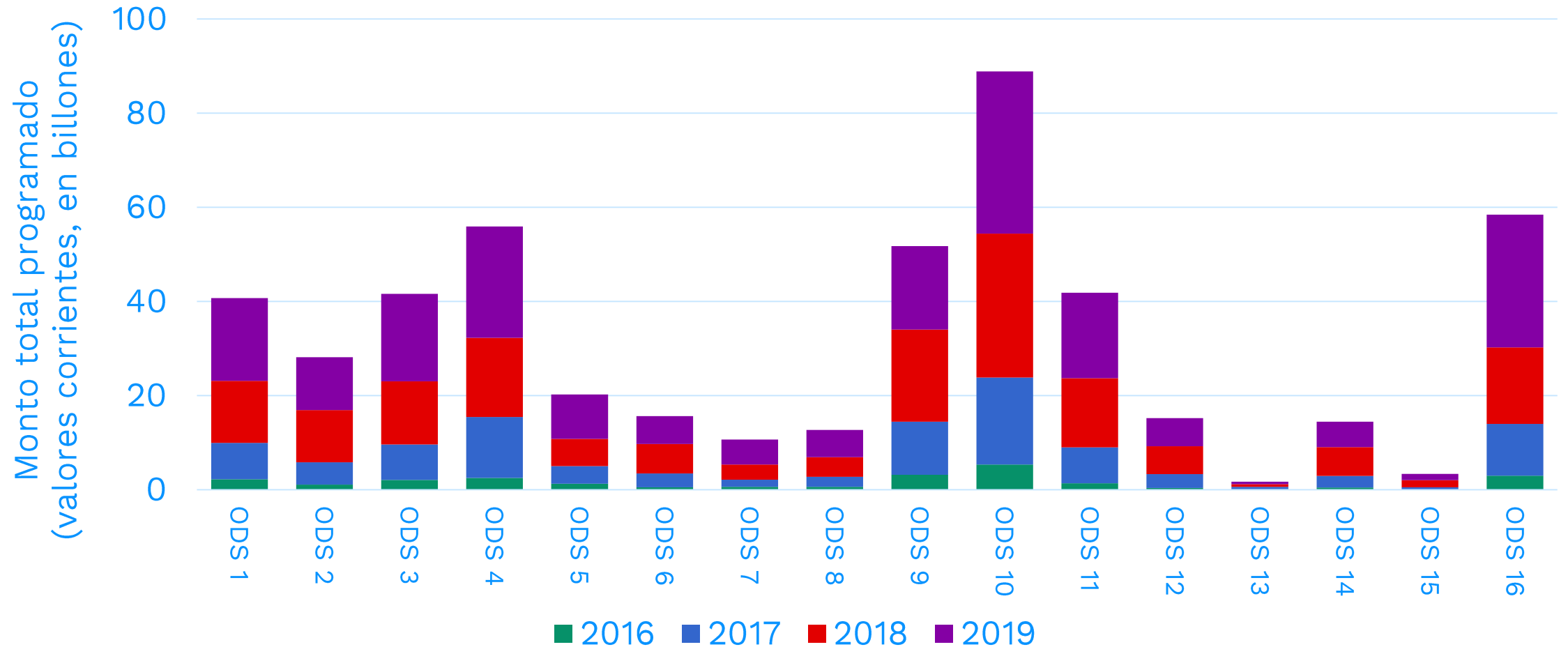
Nota: Se muestran los resultados obtenidos con los umbrales escogidos automáticamente

## Monto total programado para los proyectos clasificados con recursos SGR



Nota: Se muestran los resultados obtenidos con los umbrales escogidos automáticamente

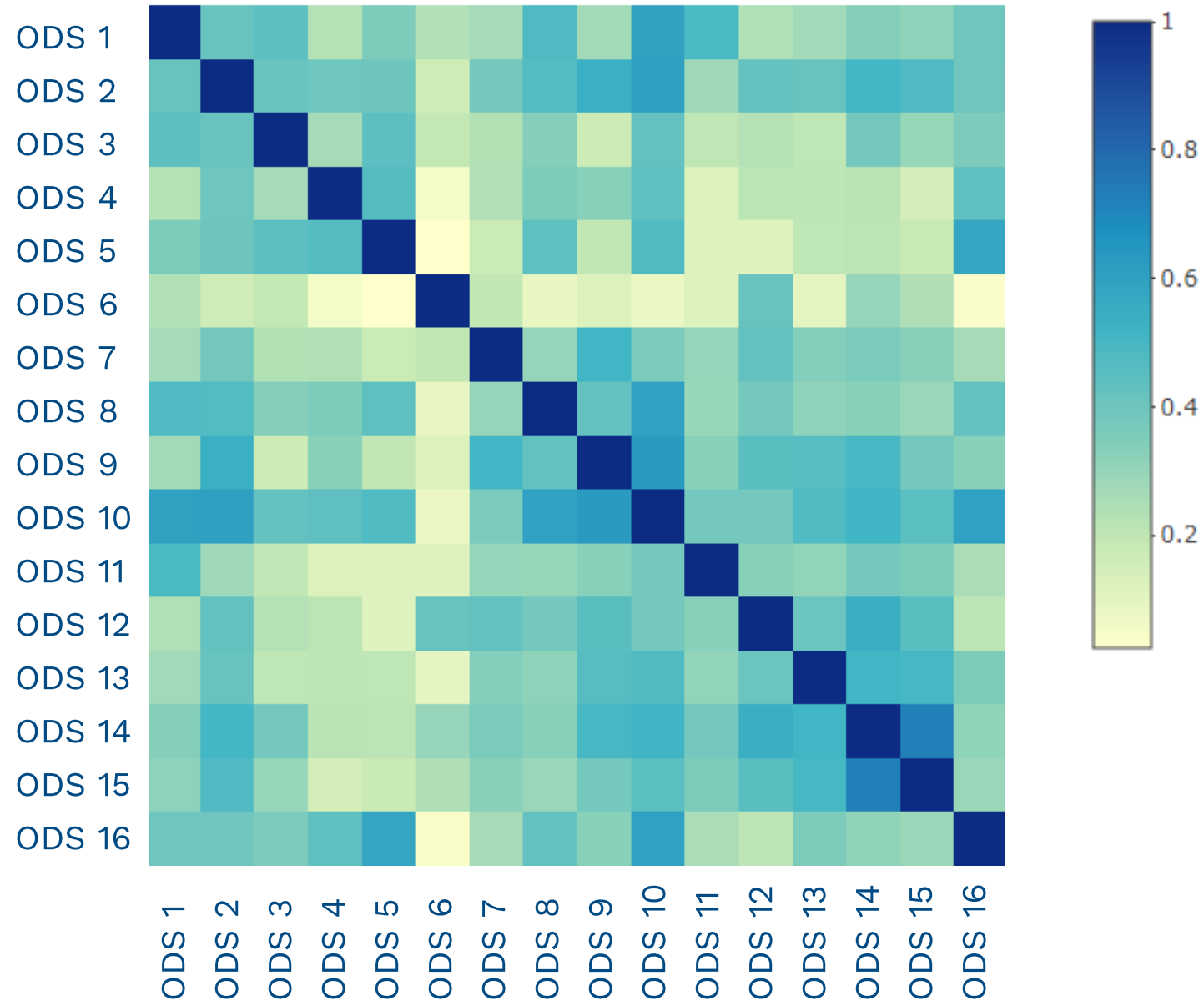
## Monto total programado para los proyectos territoriales (excepto SGR) clasificados



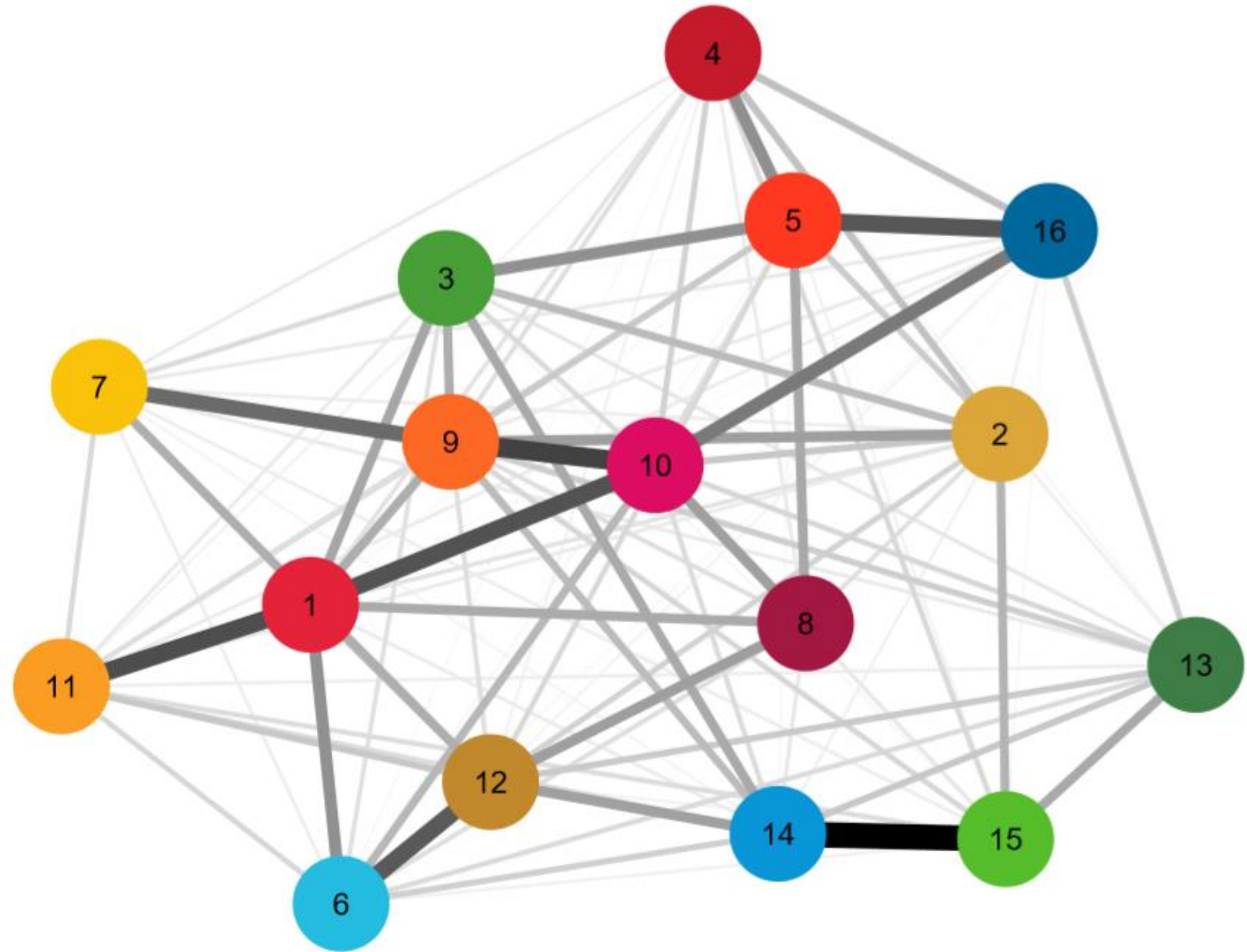
Nota: Se muestran los resultados obtenidos con los umbrales escogidos automáticamente

# Correlación entre los puntajes asignados

## Relaciones entre los ODS

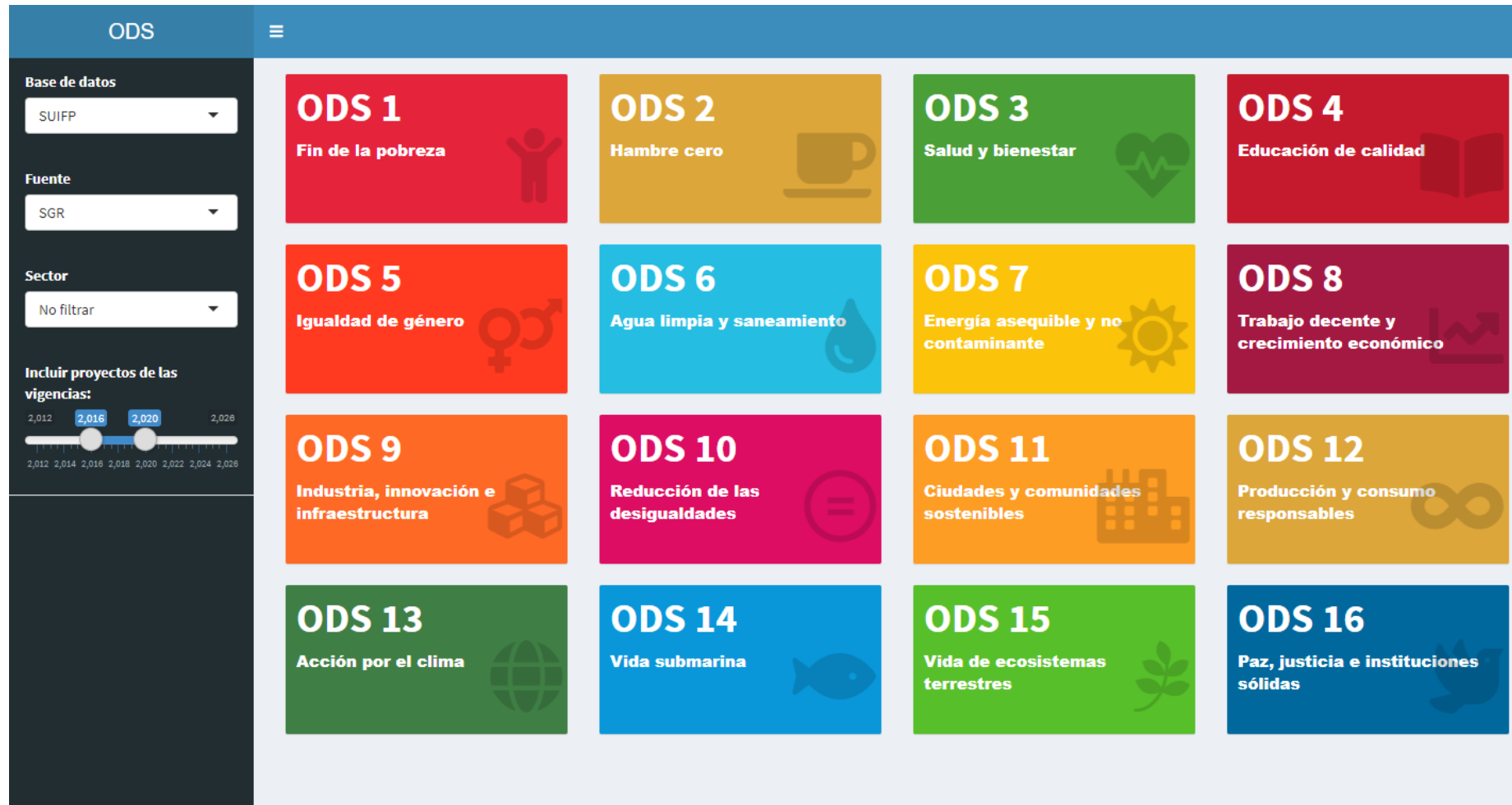


# Relaciones entre los ODS



Nota: Los enlaces representan la correlación entre los puntajes de similitud.

# Aplicativo desarrollado

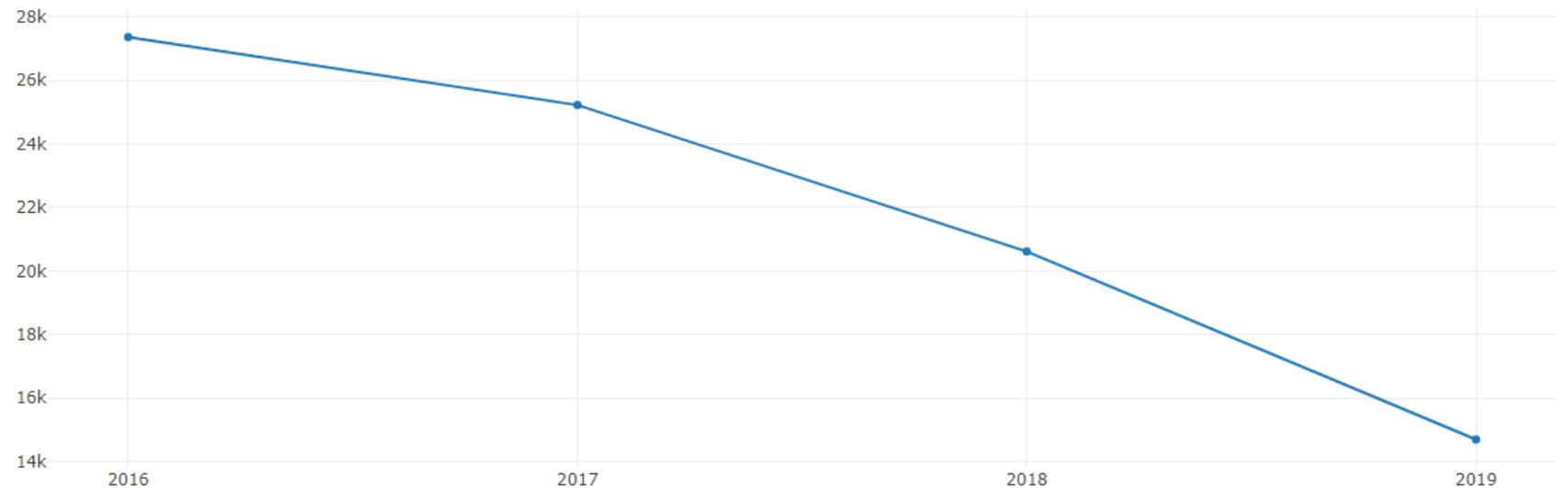


# Monto asignado total y por vigencia para los proyectos clasificados

Monto total programado para los proyectos de inversión escogidos (valores corrientes, miles de millones)

\$ 87.897 miles de millones

Monto total programado por vigencia (valores corrientes, miles de millones)

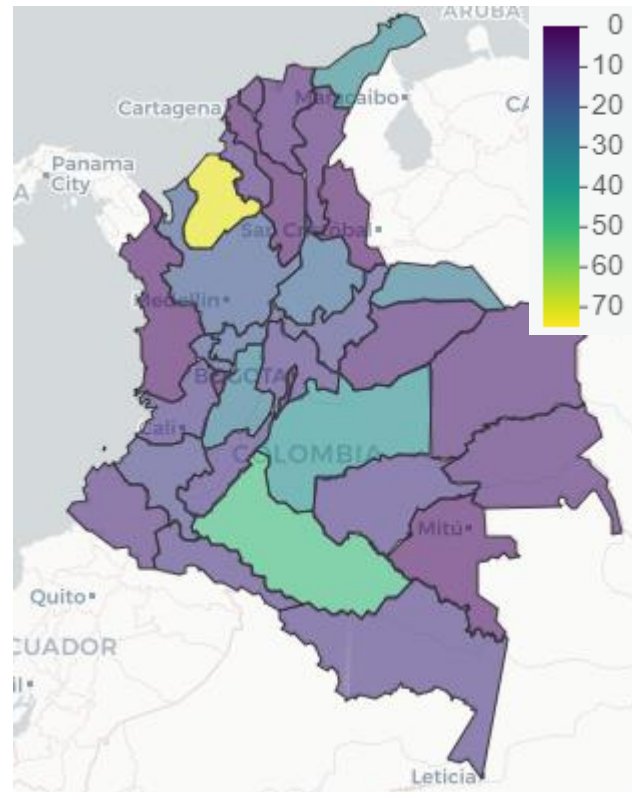


Nota: Se muestran los resultados obtenidos con el umbral escogido automáticamente

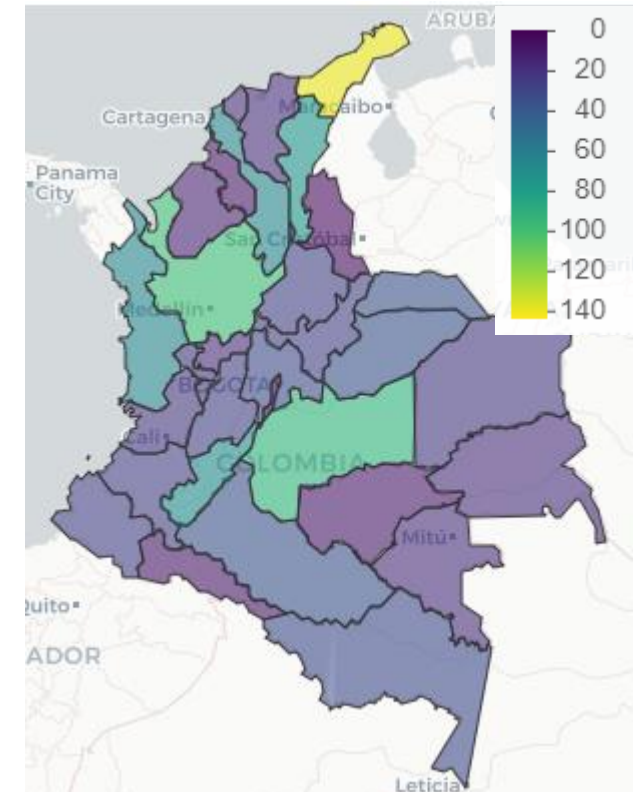
Gráficos del  
aplicativo:  
ODS 1

## Inversión por departamento entre los proyectos clasificados

Gráficos del  
aplicativo:  
ODS 1



**2016-2017**

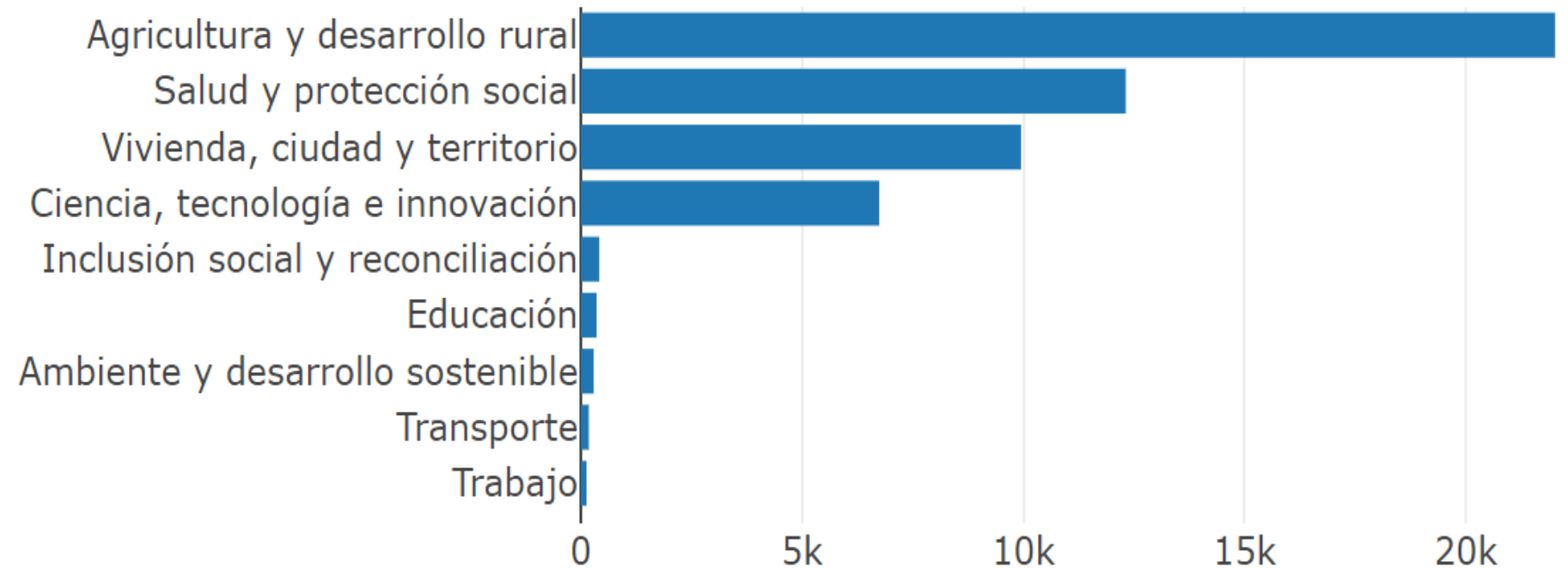


**2018-2019**

Nota: Se muestran los resultados obtenidos con el umbral escogido automáticamente



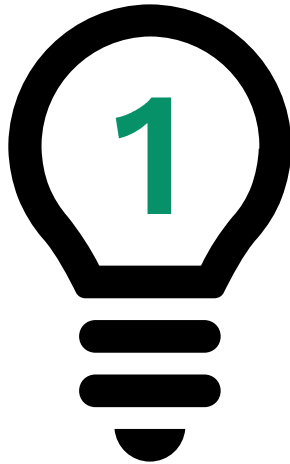
## Inversión por sector (miles de millones) entre los proyectos clasificados



Gráficos del  
aplicativo:  
ODS 1

Nota: Se muestran los resultados obtenidos con el umbral escogido automáticamente

# Conclusiones



Las técnicas de minería de texto representan una alternativa para estimar la inversión pública en alineación con los ODS.

# Conclusiones



Los análisis tienen en cuenta las relaciones existentes entre los ODS, pues un mismo proyecto de inversión puede estar alineado con más de un ODS.

# Conclusiones



El aplicativo desarrollado resume la información más relevante para la toma de decisiones.





**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación