

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación



IDENTIFICACIÓN DE REDUNDANCIAS EN ESPACIOS DE PARTICIPACION CIUDADANA

Entidad

Departamento Nacional de Planeación

- Dirección de Desarrollo Digital.
- Dirección Justicia Seguridad y Gobierno

Sector

Planeación

Lenguaje

R

Fuente de datos

Base de datos de espacios de participación ciudadana (DNP – DJSG)

Presentación

Los espacios de participación son el punto de encuentro entre la administración pública y la ciudadanía en donde a través de procesos de dialogo, deliberación y concertación, se determinan acciones en procura del bienestar general. Aunque en Colombia se tienen 125 espacios de participación institucionalizados (reglamentados a partir de leyes, decretos y resoluciones), el PND 2018-2022 identificó la necesidad de desarrollar una estrategia de racionalización de los espacios, instancias y mecanismos de participación, con el fin de tener un sistema coherente, práctico, articulado y que cuente con mecanismos efectivos. Por este motivo, el DNP requiere datos y herramientas que faciliten la observación de la información disponible sobre los espacios de participación ciudadana para orientar de manera objetiva la propuesta de racionalización mediante la toma de decisiones basadas en datos. Ante esta necesidad, la Unidad de Científicos de Datos desarrolló una herramienta de visualización para identificar las redundancias, utilizando métodos de análisis automático de texto, los espacios de participación diferentes que tienen funciones similares y en los cuales participan los mismos actores (es decir, espacios diferentes con redundancias en composición y en funciones). Además, se desarrolló una herramienta de visualización que permite realizar análisis exploratorios a los objetos y funciones de distintos espacios de participación y comparar la composición y las funciones de cualquier par de espacios de participación.

The citizen participation mechanisms are the meeting point between the public administration and the citizens where, through processes of dialogue, deliberation and agreement, are determined actions in pursuit of the general welfare. Even if in Colombia there are 125 institutionalized participation mechanisms (i.e. regulated by laws, decrees and resolutions), the PND 2018-2022 identified the need to develop a strategy to rationalize the participation mechanisms, in order to have a coherent, practical and articulated system. For this reason, the National Department of Planning requires data and tools that facilitate the observation of available information on citizen participation mechanisms in order to objectively guide the rationalization proposal through data-based decision-making. Given this need, the Data Scientists Team developed a visualization tool to identify redundancies, using automatic text analysis methods, the different participation mechanisms that have similar functions and in which the same actors participate (i.e., different spaces with redundancies in composition and functions). In addition, a visualization tool was developed that allows exploratory analysis of the objects and functions of different participation mechanisms and compares the composition and functions of any pair of them.



Objetivo general

Brindar insumos objetivos que sirvan como apoyo para la formulación de una propuesta de racionalización de los espacios de participación ciudadana.

Objetivos específicos

1. Estructurar las bases de datos de manera que permitan analizar la similitud de los espacios de participación ciudadana desde los enfoques de su composición (actores) y de sus funciones.
2. Identificar similitudes sintácticas entre las funciones de diferentes espacios de participación ciudadana en los que participe un mismo actor.
3. Desarrollar una herramienta de visualización que simplifique el análisis descriptivo y exploratorio de la información disponible sobre los espacios de participación ciudadana.

Metodología

La metodología utilizada para identificar redundancias en diferentes espacios de participación ciudadana se ilustra en la Figura 1. El proceso será descrito a continuación.

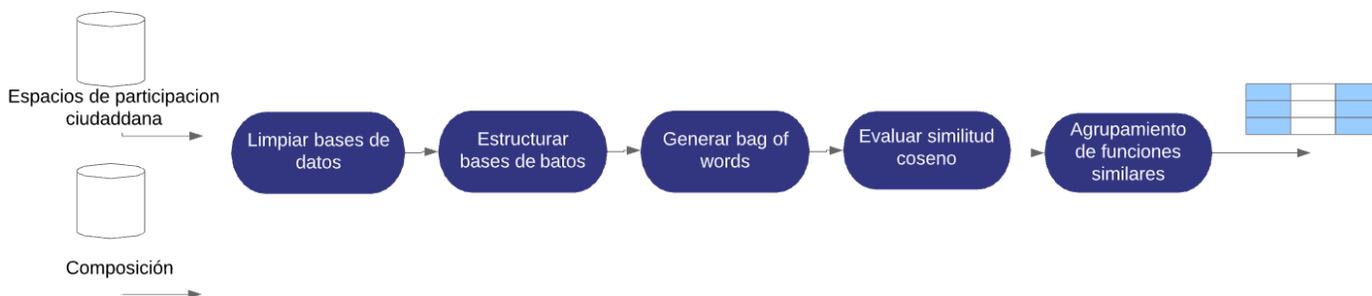


Figura 1: Componentes de la metodología y flujo del proceso.

El procedimiento desarrollado para el análisis puede resumirse en 6 etapas: (1) Limpieza de las bases de datos que relaciona funcionarios que participan, objetivos de cada uno de los espacios de participación ciudadana y la base que relaciona composición con actores, (2) Estructuración de las bases de datos, (3) Generación de Bag Of Words (BOW) para las funciones correspondientes a cada espacio, (4) Evaluación de la similitud coseno con la vectorización generada por el BOW, (5) Agrupamiento de funciones similares, (6) Desarrollo de una herramienta de visualización.

Etapa 1: Limpieza de la base de datos

Las bases de datos que se tomaron como insumo para el desarrollo de este proyecto fueron dos: (1) la base de datos *Espacios de participación ciudadana*, que incluye información sobre los espacios de participación, las normas (leyes, decretos y resoluciones) que regulan cada espacio, su objeto, escala (Nacional, Departamental, Municipal, etc.), alcance (concertación, consulta, fiscalización, iniciativa, etc.), composición y función, y (2) la base *Diccionario de actores*, que contiene el listado de actores, composición limpia y texto original.

Inicialmente se realizó la limpieza de los campos de texto de las bases de datos (composición y funciones), esto incluía limpiar las cadenas de texto (eliminar Stop Words, mayúsculas, puntuaciones, espacios



El futuro es de todos

DNP
Departamento
Nacional de Planeación

innecesarios y conjugaciones). También se incluyó en la limpieza de texto la función *stemming* usando el algoritmo de Porter, este algoritmo evita afectar la frecuencia de una palabra por su conjugación, es decir, toma la raíz de las palabras sin tener en cuenta sufijos de género, número, o conjugaciones de tiempo para una palabra, por ejemplo:

“Aquel es un caballo de la caballería militar, los otros caballos no.”

Al aplicar el algoritmo de Porter se identificaría una raíz “*caball*”, la cual tendría una frecuencia 3 pues esta misma raíz hace referencia al conjunto de palabras caballo, caballería y caballos. También se revisó que las distintas funciones asociadas a cada uno de los espacios estuvieran separadas por *enter*, esto con el fin de dividir las para obtener una función por celda.

Como último paso respecto a esta etapa se eliminó un conjunto de palabras dentro de las funciones, pues se consideraban repetitivas, este conjunto de palabras es:

"colombia", "cupo", "cupos", "uno", "dos", "tres", "cuatro", "cinco", "seis", "siete", "ocho", "nueve", "diez", "once", "doce", "trece", "catorce", "quince", "ser", "haga", "veces", "cada", "deleg\\w+", "represent\\w+", "parágrafo", "preside", "presidirá", "director", "subdirector", "general", "directora", "crédito público", "nacional", "nivel directivo", "viceminist\\w+", "parágrafo", "quien", "quién", "podrá", "podrán".

La expresión regular (\\w+) indica que se evaluará si existe uno o más caracteres alfanuméricos después del texto dado, por ejemplo: deleg\\w+ incluye las palabras delegado, delegación, delegada, etc.

Etapa 2: Estructuración bases de datos

Para la segunda etapa, se realizó el cruce, partiendo de las bases de datos brindadas por la Dirección de Justicia Seguridad y Gobierno. El cruce mencionado se realizó con el fin de hacer un análisis a actores que deben interactuar en un espacio no a la composición de cada espacio. Por ejemplo:

Si tomamos un espacio cualquiera cuya composición se encuentra a continuación

Composición: El Ministro del Interior o su delegado.
El Ministro de Hacienda y Crédito Público o su delegado.
El Ministro de Justicia y del Derecho o su delegado.
El Ministro de Defensa Nacional o su delegado
El Ministro de Agricultura y Desarrollo Rural o su delegado, quien lo presidirá.
El Ministro de Ambiente y Desarrollo Sostenible o su delegado.

El listado de actores será el siguiente:

Actores: Ministerio del Interior
Ministerio de Hacienda y Crédito Público
Ministerio de Justicia y del Derecho
Ministerio de Defensa



Ministerio de Agricultura y Desarrollo Rural

Ministerio de Agricultura y Desarrollo Rural

La tabla resultante tiene en la primera columna los actores seguido por una columna que relaciona los espacios en los cuales el actor participa, por último, la tabla tiene las funciones de cada uno de los espacios, esto con el fin de identificar funciones similares entre diferentes espacios donde participa el mismo actor. Para realizar este cruce en la base de datos se tomó la variable composición.

Etapa 3: Representación del texto con el modelo “Bag Of Words”

Esta etapa toma como insumo la salida de la etapa anterior (diccionario de actores asociados a los espacios en que participan y las funciones de cada uno de los espacios). Del insumo se realiza un Bag Of Words (BOW), respecto a todas las funciones de todos los espacios en que participa cada uno de los actores.

Un Bag Of Words consiste en construir una matriz de frecuencia, típicamente cada una las palabras (o unidad mínima de texto analizar) de todos los textos se ubican en las columnas y cada fila representa cada uno de los textos, lo anterior permite vectorizar (es decir, transformar a una representación numérica) el texto o los textos de acuerdo con la frecuencia de la palabra dentro de un conjunto de textos, a continuación, la Figura 2 ofrece un ejemplo para entender cómo funciona un BOW.

	este	es	un	texto	otro	último
este es un texto	1	1	1	1	0	0
este texto es otro texto	1	1	0	2	1	0
un último texto	0	0	1	1	0	1

Figura 2: Ejemplo de una representación de texto usando Bag Of Words.

Etapa 4: Evaluar similitud coseno

Con la vectorización de textos se aplica la similitud coseno, una medida que tiene como argumento el ángulo entre dos vectores, dado que exista una similitud máxima, es decir, que los dos vectores evaluados apunten a la misma dirección el coseno toma valor de 1, en este caso particular se interpreta la similitud máxima como dos funciones iguales en diferentes espacios que un mismo actor debe asistir. La figura (3) muestra la fórmula para evaluar la similitud coseno.

$$\text{similitud} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

Figura 3: Formula similitud coseno.

Respecto al análisis se estableció un umbral de 0.5, es decir, que para los valores de las similitudes coseno mayores a 0.5 se consideraba similares las funciones evaluadas, es importante aclarar que se podían



considerar un grupo de tres o más funciones similares, no se evaluaba únicamente parejas de funciones. A manera de ejemplo con el fin de entender mejor la similitud coseno, se tomará los textos:

A: *este es un texto.*

B: *este texto es otro texto.*

Del ejemplo anterior sabemos que la vectorización de A y B son las siguientes:

$$A = [1,1,1,1,0,0]$$

$$B = [1,1,0,2,1,0]$$

Por lo tanto, podemos hacer el producto punto de los dos vectores y el producto de las normas, obteniendo el siguiente resultado:

$$\text{Cos}\theta = \frac{4}{2 * 2.64} = 0.75$$

Es decir, con un umbral de similitud igual a 0.75, se considera que los textos A y B son similares.

Etapa 5: Agrupamiento de funciones similares (en espacios diferentes donde participa el mismo actor)

Mediante la técnica de *clustering* jerárquico por aglomeramiento se realizó el agrupamiento de funciones similares, esta etapa toma en consideración las similitudes entre textos (funciones de los espacios de participación) analizadas con la similitud coseno. La jerarquización se realizó bajo el criterio del mínimo.

La Figura 4 es una visualización de proximidades entre puntos individuales, cada punto representa una función y cada enlace la similitud coseno entre las funciones.

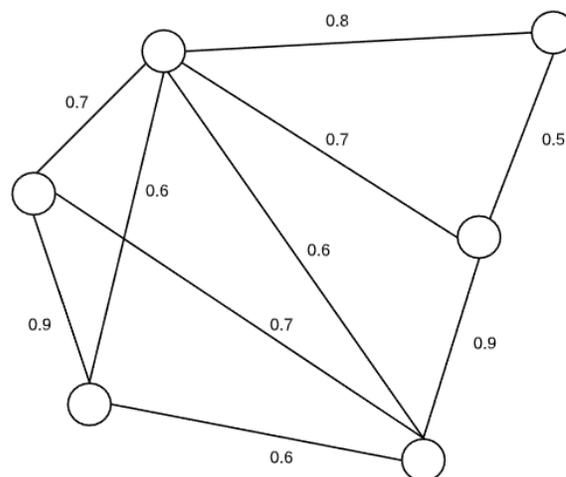


Figura 4. Representación similitud de funciones(puntos)



Dado el caso anterior con valores determinados para las similitudes, se realizará un ejemplo de *clustering* jerárquico aglomerativo con un umbral de 0.9. Para realizar el *clustering* primero se evalúan las relaciones que tiene un valor mayor o igual al umbral (0.9). Después se establecen enlaces entre los puntos con mayor similitud generando clúster, es decir, un conjunto (dos o más) de puntos es interpretado como un punto individual, las relaciones restantes se establecen con la menor similitud entre los puntos contenidos en el clúster y el resto de los puntos.

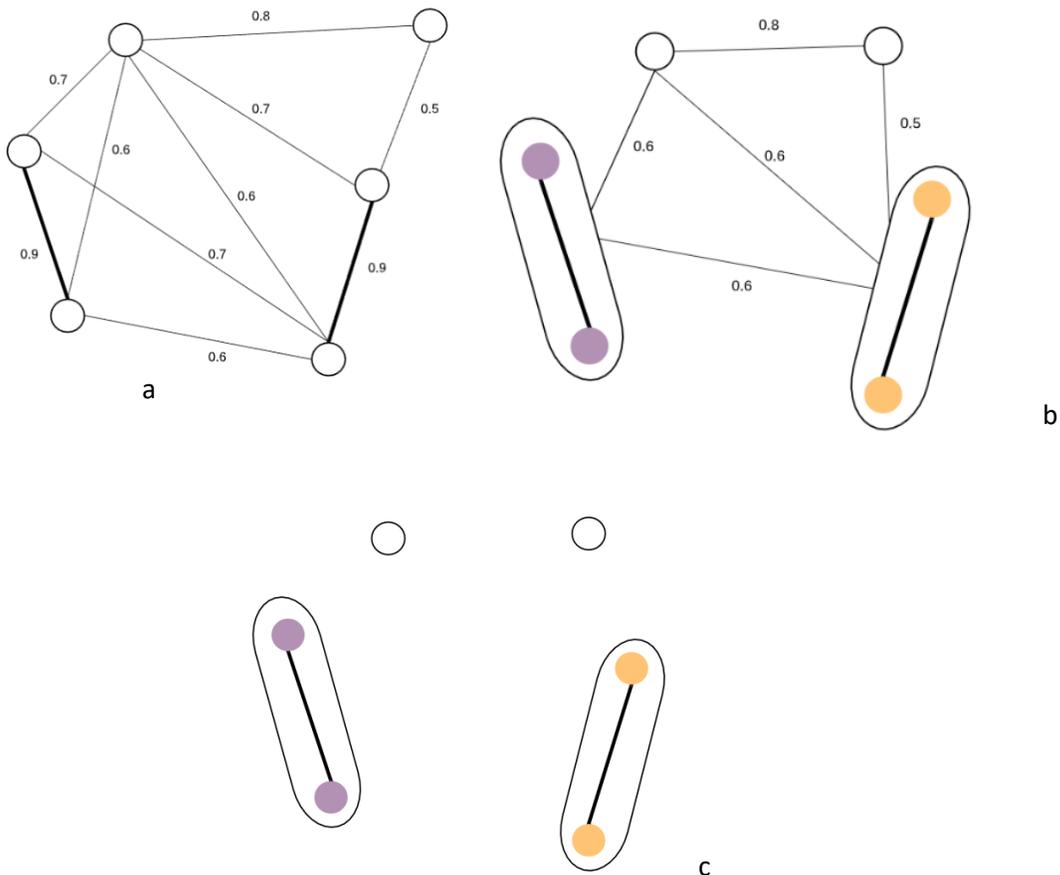


Figura 5. a) Establecer enlaces entre puntos con mayor similitud. b) Generación de clúster y nuevos enlaces. c) Visualización final del clustering

Ahora el caso anterior se evaluará bajo las mismas condiciones con un umbral de 0.8. Los pasos para realizar el *clustering* son los mismos, primero se evalúan las relaciones que tiene un valor mayor o igual al umbral (0.8). Después se establecen enlaces entre los puntos con mayor similitud generando clúster, es decir, un conjunto



El futuro
es de todos

DNP
Departamento
Nacional de Planeación

(dos o más) de puntos es interpretado como un punto individual, las relaciones restantes se establecen con la menor similitud entre los puntos contenidos en el clúster y el resto de los puntos.

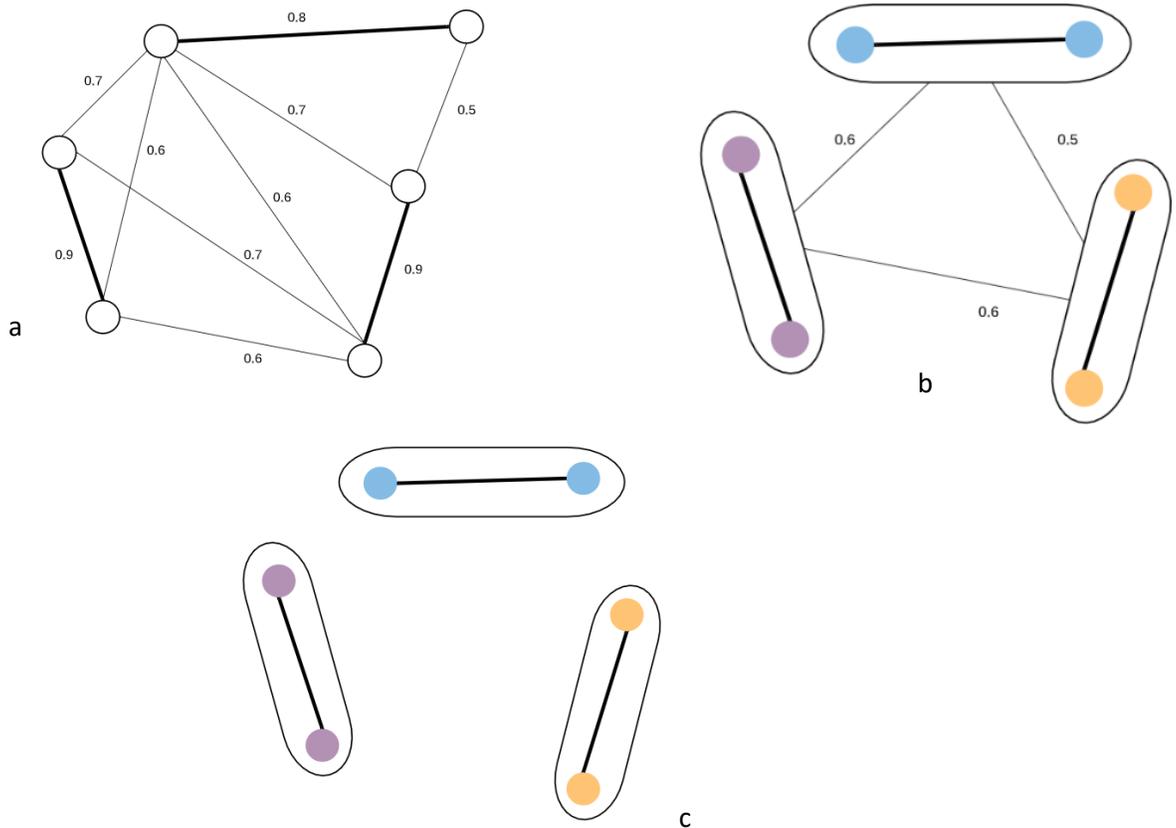


Figura 6. a) Establecer enlaces entre puntos con mayor similitud. b) Generación de clúster y nuevos enlaces. c) Visualización final del clustering

Etapa 6: Creación de la herramienta de visualización

Para visualizar los resultados más relevantes, se presentó la información en un tablero interactivo. El objetivo del tablero fue poder realizar comparaciones entre pares de espacios de participación y caracterizarlos con base en su objeto y sus funciones. Para las comparaciones por pares, se definió un indicador de “compatibilidad en composición” que se presenta en el aplicativo y que se define como:

$$\% \text{ Compatibilidad en composición} = \frac{\# \text{ de actores presentes en ambos espacios}}{\# \text{ de actores en el espacio con menos actores}}$$

El detalle de este aplicativo se encuentra en el manual de usuario.



Resultados

Se analizaron 117 espacios, en promedio había 11 actores y 9 funciones por cada espacio. Mediante la comparación de las funciones 1 a 1 de los espacios en los que actúan los mismos actores con un umbral de 0.5 se identificaron 49 similitudes sintácticas (es decir, relacionadas con la escritura de los textos como tal y no semánticas, es decir, relacionadas con su significado), por lo que se recomienda que la Dirección de Justicia Seguridad y Gobierno evalúe en contexto los resultados con el fin de decidir si estos pueden realmente representar redundancias en las funciones de espacios diferentes donde participe el mismo actor. Dos ejemplos de las similitudes sintácticas encontradas se ilustran en las figuras 7 y 8.

Espacios	Composición	Funciones
 Consejo Nacional Ambiental (Ley 99 de 1993 y Decreto 1124 de 1999)	a) El Ministro del Medio Ambiente, quien lo presidirá. b) El Ministro de Minas y Energía.	1. Recomendar al Gobierno Nacional la política y los mecanismos de coordinación de las actividades de todas las entidades y organismos públicos y privados cuyas funciones afecten o puedan afectar el medio ambiente y los recursos naturales renovables.
 Consejo Asesor de Política Minera (Ley 685 de 2001)	a) El Ministro de Minas y Energía, quien lo presidirá. b) El Ministro del Medio Ambiente.	a) Recomendar al Gobierno Nacional la política y los mecanismos de coordinación de las actividades de todas las entidades y organismos públicos y privados cuyas funciones afecten o puedan incidir en la industria minera.

Figura 7: Ejemplo de similitud sintáctica identificada. Ministerios de Medio Ambiente y Minas y Energía participan en ambos espacios y ambos espacios tienen una función que es sintácticamente (textualmente) parecida, pero su significado puede ser diferente.

Espacios	Composición	Funciones
 Comisión Nacional Intersectorial de Vigilancia en Salud Pública (Decreto 3518 de 2006)	a) El Ministro de la Protección Social o su delegado, quien la presidirá; b) El Director del DNP o su delegado.	a) Asesorar y apoyar al Ministerio de la Protección Social en la definición de lineamientos para las acciones de vigilancia en salud pública que sean prioritarias para el país, especialmente, tratándose de decisiones para preservar la seguridad nacional en salud.
 Instancia de coordinación y asesoría dentro del SGSSS² (Decreto 2478 de 2014)	1. El Ministro de Salud y Protección Social, quien podrá delegar en uno de los Viceministros y la presidirá; 2. El Director del DNP, quien podrá delegar en uno de los Subdirectores Generales.	1. Asesorar y efectuar recomendaciones al Ministerio de Salud y Protección Social (...) en asuntos relacionados con los programas y proyectos que son de su competencia y, específicamente, en materia de prestación de servicios de salud, salud pública, aseguramiento en salud y promoción social en salud, que se presenten a su consideración.

Figura 8: Ejemplo de similitud sintáctica identificada. El Ministerio de Salud y Protección Social y el Departamento Nacional de Planeación participan en ambos espacios y ambos espacios tienen una función que es sintácticamente (textualmente) parecida, pero su significado puede ser diferente.



El futuro
es de todos

DNP
Departamento
Nacional de Planeación

En el proceso también se encontró que existen entidades que participan en muchos espacios, por ejemplo, el Ministerio de Educación es quien participa en más espacios (32), seguido del Departamento de Planeación Nacional que participa en 30 espacios. Las entidades presentes en más espacios se presentan en la figura 9.

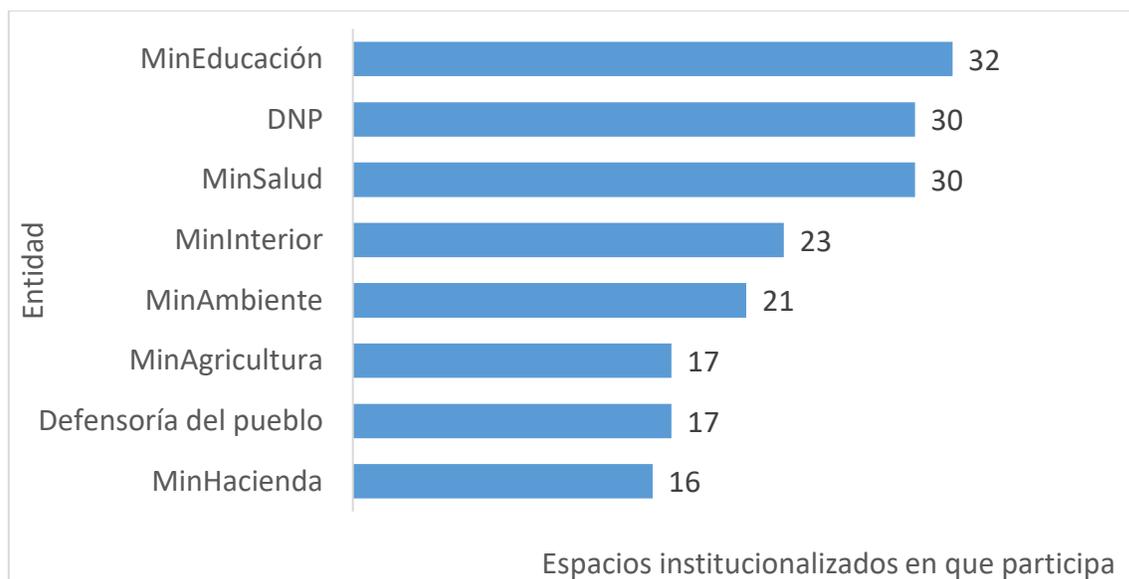


Figura 9: Entidades que hacen parte de más espacios institucionalizados de participación.

Finalmente, la herramienta de visualización fue creada a partir de la librería *Shiny*. Esta herramienta permite filtrar por tipo de actor (por ejemplo, todos los espacios en que participan actores del sector privado) por actor (por ejemplo, todos los espacios en que participa el Ministerio de Salud o el DNP), por sector (por ejemplo, todos los espacios en que participan actores del sector Salud), por clasificación de los espacios (por ejemplo, todos los espacios de Gestión Pública) y por subcategoría de los espacios (por ejemplo, todos los espacios para las Víctimas del conflicto armado). También genera nubes de palabras y redes de palabras con base en los objetos y las funciones de los espacios de participación escogidos por el usuario. A la herramienta se puede acceder a través de intranet (red interna del DNP) por medio del siguiente enlace:

http://vdatascience:3838/participacion_ciudadana/

La interfaz del aplicativo se ilustra en la figura 10. La interfaz y funcionalidades de la herramienta se presentan con más detalle en el manual de usuario.



El futuro es de todos

DNP Departamento Nacional de Planeación

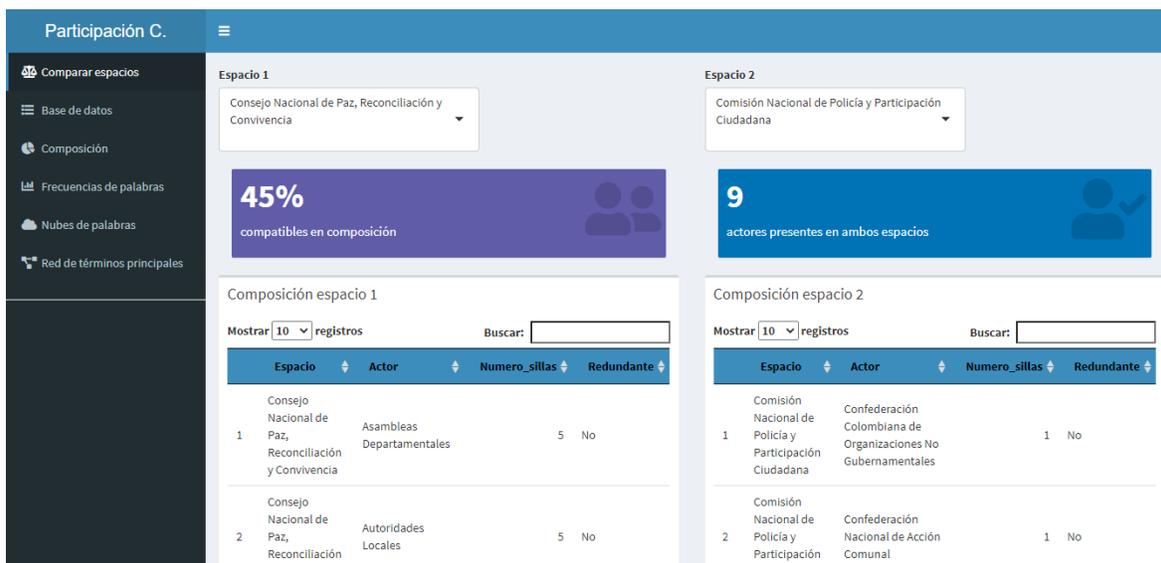


Figura 10: Interfaz del aplicativo de visualización en el módulo de comparación de espacios.

Conclusiones y recomendaciones

1. El procedimiento permitió identificar 49 similitudes sintácticas de interés, es decir, 49 casos en que existen funciones similares en espacios diferentes en los que participa (por lo menos) un mismo actor.
2. El análisis brinda dos herramientas para apoyar la toma de decisiones en la formulación de la política de participación ciudadana: un listado de redundancias sintácticas identificadas y un aplicativo de visualización que facilita la comparación entre espacios y la realización de análisis descriptivos y exploratorios.
3. Se recomienda que los resultados se analicen en contexto por parte de los técnicos de la DJSG para determinar si las similitudes sintácticas corresponden a verdaderas redundancias en las funciones.
4. Se recomienda analizar la articulación entre los espacios en que participan entidades como el Ministerio de Educación, el DNP y el Ministerio de Salud y Protección Social, quienes están presentes en 30 espacios o más.

Socialización

Los resultados de este proyecto fueron presentados a la Dirección de Justicia Seguridad y Gobierno, a quienes se entregó la herramienta y las bases de datos estructuradas para facilitar futuros análisis de la información.