

06-013

**SUSTAINABLE DEVELOPMENT GOALS IN PROJECT MANAGEMENT.
QUANTITATIVE AND AUTOMATIC EVALUATION AT THE 24TH ICPME CONGRESS.**

Fariña García, María Consuelo ⁽¹⁾; De Nicolás, Víctor Luis ⁽²⁾; Yagüe Blanco, José Luis ⁽³⁾;
Labrador Fernández, Jesús ⁽²⁾

⁽¹⁾ Universidad Politécnica, ⁽²⁾ Universidad Pontificia de Comillas, ⁽³⁾ Universidad
Politécnica de Madrid

The methods currently used to determine the degree of linkage of the projects of organizations with the goals of the Sustainable Development Goals (SDG) are complex, expensive and sometimes difficult to apply for small organizations and local public administrations. The objective of this study is to find a quick and objective methodology that allows analyzing the linkage of the SDGs with the projects. Using automated quantitative methods can be very helpful when making decisions to achieve the SDGs. For this, text mining techniques will be used in documents. Two methods will be applied to the communications presented at the project management congress ICPME. The result will be contrasted with the opinion of the authors. The results show that there is a greater concordance between the method and the choice of SDGs made by the authors when only the title of the communication is taken into account. This seems to indicate that the identification of the SDGs may have been based on keywords taken from the title of the SDGs without taking into account a more specific level of interrelation such as that contained in the different goals of each objective.

Keywords: Project evaluation; text mining; sustainable development goals.

**OBJETIVOS DE DESARROLLO SOSTENIBLE EN LA DIRECCIÓN DE PROYECTOS.
EVALUACIÓN CUANTITATIVA Y AUTOMÁTICA EN EL 24º CONGRESO CIDIP.**

Los métodos utilizados actualmente para conocer grado de vinculación de los proyectos de organizaciones con las metas de los Objetivos de Desarrollo Sostenible (ODS) son complejos, costosos y en ocasiones difíciles de aplicar para organizaciones pequeñas y administraciones públicas locales. El objetivo de este estudio es encontrar una metodología rápida y objetiva que permita analizar la vinculación de los ODS con los proyectos. Utilizar métodos cuantitativos automatizados puede ser de gran utilidad en la hora de tomar decisiones dirigidas a alcanzar los ODS. Para ello se van a utilizar técnicas de minería de textos en documentos. Se va a aplicar dos métodos a las comunicaciones presentadas en el congreso de dirección de proyectos CIDIP. El resultado será contrastado con la opinión de los autores. Los resultados muestran que hay una mayor concordancia entre el método y la elección de ODS que realizan los autores cuando se tiene en cuenta sólo el título de la comunicación. Esto parece señalar que la identificación de los ODS puede haberse basando en palabras clave tomadas del título del ODS sin tener en cuenta un nivel de interrelación más específico de como el que se recoge en las diferentes metas de cada objetivo.

Palabras claves: Evaluación de proyectos; minería de textos; objetivos de desarrollo sostenible.

Correspondencia: María Consuelo Fariña García farinachelo@gmail.com



© 2021 by the authors. Licensee AEIPRO, Spain. This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introducción

Los Objetivos de Desarrollo Sostenible (ODS) promovidos por la ONU (Naciones Unidas, 2014) y avalados por los gobiernos firmantes son un marco de referencia en la dirección y planificación de proyectos. Este hecho ha generado un creciente interés de las instituciones, empresas y otras organizaciones en el modo de implementarlos y evaluar su cumplimiento. Para ello la ONU ha elaborado un conjunto de indicadores que permiten medir el grado de cumplimiento de los distintos objetivos (United Nations. Sustainable Development Network Solutions, 2014). Sin embargo, para poder calcular todos estos indicadores es necesario recoger gran cantidad de información pluridisciplinar que generalmente es poco accesible y necesita ser analizada rigurosamente con métodos complejos y costosos (Klopp & Petretta, 2017). La dificultad para utilizar los indicadores es aún mayor para las organizaciones de pequeño tamaño y las instituciones locales que no tienen suficientes medios económicos y humanos para desarrollar este tipo de estudios (Nagy, Benedek, & Ivan, 2018). Sin embargo, la importancia de estas instituciones es crucial para el desarrollo sostenible, ya que solamente desde un trabajo local se logran cambios que afecten a la globalidad (Guha & Chakrabarti, 2019).

Por otra parte, existe otra dificultad debida a que la vinculación con los ODS de una organización depende del sector de la misma y del enfoque de la organización. No es lo mismo analizar la vinculación con los ODS de una empresa de alimentación que de una empresa de vehículos. En este sentido existen estudios específicos de muchas organizaciones sectoriales que estudian los ODS desde la perspectiva de su sector centrándose en analizar cómo se pueden mejorar objetivos concretos: el ODS 6, sobre agua y saneamiento (Ait-Kadi, 2016), el ODS 7, sobre energía (Fuso Nerini et al., 2018), el ODS 14, sobre la protección de los ecosistemas marinos (Ntona & Morgera, 2018), el ODS 8, sobre el crecimiento y el empleo digno (Poschen, 2015). También existen estudios no sectoriales que se centran en todos los objetivos (M.C; Fariña García, De Nicolás De Nicolás, V.L; Yagüe Blanco, & Zamorano, 2018; M.C. Fariña García, De Nicolás De Nicolás, Yagüe Blanco, & Labrador Fernández, 2021; Le Blanc, 2015; Nilsson et al., 2018). Estos estudios han demostrado que ningún objetivo está desligado de los demás, siendo distintas las interacciones entre los objetivos.

La dirección y planificación de proyectos tiene un papel determinante en la toma de decisiones sobre la aplicación de los ODS y es necesario que los responsables de un proyecto estén informados de su situación lo más rápido posible para obtener conclusiones que ayuden a dirigir los proyectos hacia la Agenda 2030 (Morris Molina, Salazar de Morris, & Barrientos, 2019; Requelme, Narcisa; Cachipundo Ulcuango, Charles Jim; Ortiz Tirado, 2019). Por esta razón, es necesario plantear métodos de evaluación más inmediatos que el riguroso estudio de indicadores. Estos métodos deben ayudar a tener una visión general de la situación de los ODS de las organizaciones. En este sentido, se puede realizar una evaluación cualitativa y cuantitativa previa de los proyectos y otros documentos de una organización. Para ellos se puede realizar un análisis de contenido que clasifique las distintas partes del informe en función de su relación con cada uno de los ODS. Este análisis puede realizarse cualitativa o cuantitativamente, de forma manual o automática. Si se realiza de forma cualitativa y manual, sería un análisis de contenido convencional cuyo resultado es más subjetivo, puesto que la persona que analiza el documento clasificará con su propio criterio las distintas partes del documento. Si se realiza de manera cuantitativa y automática se utilizarían técnicas de minería de textos (MT) que permiten detectar de

manera cuantitativa, a través de entidades clave, las relaciones con cada uno de los ODS. Este último método es el que se va a utilizar en este estudio.

1.1. Evaluación de Documentos con Minería de Textos

La MT sirve para *“identificar hechos y datos puntuales a partir del texto de los documentos; determinar el tema o temas tratados en los documentos mediante la categorización automática; identificar los conceptos tratados en los documentos y crear redes de conceptos; facilitar el acceso a la información repartida entre los documentos de la colección, mediante la elaboración automática de resúmenes, y la visualización de las relaciones entre los conceptos tratados en la colección”* (Eito Brun & Senso, 2004). Las técnicas de MT se desarrollaron para la recuperación de información, aunque actualmente se utilizan en otros ámbitos como la documentación digital, los buscadores de internet, en las bases de datos, la cienciometría, etc. (Boeris, 2013). La MT también se utiliza en sociología para la investigación cualitativa, aunque en este campo suele hablarse de análisis de contenido. Cuando el análisis de contenido tiene un alto volumen de datos y su análisis necesita ser automatizado, es cuando el proceso se convierte en MT (Antons, Grünwald, Cichy, & Salge, 2020). Cuando el volumen de documentos es excesivamente grande entonces la MT se transforma en un estudio de Big Data (Fayyad, Piatetsky-shapiro, & Smyth, 1996). Sin embargo, la base teórica es la misma en los tres análisis. La MT se incluye dentro de la minería de datos (MD), pero se diferencia de esta en que la primera trabaja con datos no estructurados mientras que la segunda trabaja con datos estructurados. A la minería de textos se la conoce también como Knowledge Data Text o KDT (Justicia De la Torre, 2017) siendo este un concepto más amplio, que incluye varias fases entre las que se encuentra la MT.

En la MT existen dos fases bien diferenciadas: el pre-procesamiento y las operaciones básicas de minería (Feldman & Sanger, 2007). En el pre-procesamiento se extraen las entidades de estudio y se reducen al máximo los conjuntos de ellas para que sean manejables. Dentro de esta etapa se realizan operaciones de procesamiento del lenguaje natural (PNL), extracción de información, categorizaciones y agrupamientos de entidades y textos (Krippendorf, 2004). Los expertos en las técnicas de minería han desarrollado diferentes algoritmos para automatizar esas búsquedas de entidades (Antons et al., 2020; Feldman & Sanger, 2007). La segunda fase son las operaciones básicas de minería, que son aquellas que se realizan para el descubrir patrones en las entidades, realizar análisis de tendencias y desarrollar algoritmos de descubrimiento de conocimiento. A continuación se desarrollan más a fondo las dos fases.

1.1.1. Pre-procesamiento

La lingüística computacional y las técnicas de recuperación de información son esenciales en la etapa del pre-procesamiento, ya que elaboran algoritmos para facilitar las tareas de procesamiento del lenguaje natural (PLN), la extracción de información, la categorización y el agrupamiento. Las técnicas dependen del objeto del estudio y la entidad utilizada en el análisis semántico: los caracteres suelen ser poco útiles en minería; las palabras, se consideran la entidad de mayor riqueza semántica; los términos, aunque son más completos necesitan extraerse con algoritmos más complejos, estos pueden ser palabras sueltas o grupos de varias palabras que habitualmente van juntas; y por último los conceptos, que son ideas, expresiones que deben ser extraídas como los conceptos con algoritmos más complejos (Feldman & Sanger, 2007).

Además en el pre-procesamiento de textos existe un concepto imprescindible para la detección de palabras clave que es la palabra vacía o stop word. Estas palabras han de ser detectadas y eliminadas antes de la aplicación del algoritmo a los textos. Propuesto por Luhn, este concepto se refiere a las palabras que no aportan información al estudio (Luhn,

1960). Suelen ser palabras vacías las palabras de categoría gramatical como artículos, pronombres, preposiciones y las formas de verbos muy usados como ser, estar, haber, ir, etc. Sin embargo, también puede haber otras palabras vacías seleccionadas por el investigador en función del objetivo del análisis.

Con respecto a los algoritmos, existen diversos estudios dirigidos a la recuperación de información que han desarrollado diferentes modelos matemáticos para la extracción de entidades. Entre los pioneros, se encuentran Hans P. Luhn y Karen Spärck Jones que juntos crearon el factor TFI_{df} (Luhn, 1960; Spärck Jones, 1972), que extrae palabras clave por su frecuencia en varios documentos. Otro algoritmo contemporáneo al anterior fue el que propuso Gerard Salton (1976) el “Valor de Discriminación de un Término” en el que la clasificación de palabras clave se hacía según la capacidad que tenían dichas palabras para discriminar unos documentos de los demás. Por tanto, las palabras clave eran aquellas que tenían la máxima distancia media posible entre las frecuencias de dichas palabras en cada documento de una colección (Salton et al., 1976). Más actual es el algoritmo RAKE (Contreras Barrera, 2018; Rose, Engel, Cramer, & Cowley, 2010) que ya permite extraer entidades multipalabra. También existen enfoques más concretos para detectar palabras clave como el cálculo de la frecuencia de recopilación, la posición relativa de la primera aparición de la palabra en el texto o el número de veces que se utiliza un término como palabra clave (Turney, 2000). Otras técnicas de minería de textos utilizan las palabras clave recogidas en tesauros sobre un área de conocimiento específico (Callon, M., J. Law, 1986; Carley, 1993), etc.

1.1.2. Operaciones Básicas de Minería

Una vez realizada la fase de pre-procesamiento hay que pasar las operaciones básicas de minería. Una opción de análisis interesante es la de establecer vínculos mediante la información coincidente o a través de las relaciones semánticas entre entidades. Esta técnica se denomina Análisis de Redes Semánticas (SNA) y permite que, una vez extraída una lista de palabras clave y conociendo sus frecuencias, se puedan establecer las relaciones entre documentos a través de cálculos y representaciones basadas en la teoría de grafos y el análisis de redes.

Según Sowa (1987), los SNA existen desde la Grecia antigua, sin embargo, no fue hasta los años 80 del siglo pasado cuando se comenzaron a utilizar con mayor frecuencia (Verd Pericas, 2005). Es precisamente a Sowa a quien se le atribuyen los primeros gráficos conceptuales, utilizados para en la automatización del lenguaje natural (Sowa, 1987). Una década antes, autores como Quillian y Collins (1972), Minsky, Longyear, Siklosy ya estaban estudiando el procesamiento de la semántica y del lenguaje dentro de las ciencias de la computación. Por otro parte, algunos psicólogos como Collins o Loftus trabajaban en la psicología cognitiva y la memoria semántica humana (Collins & Loftus, 1975). Collins y Quillian, a su vez, trabajaron juntos en el desarrollo de la inteligencia artificial (Collins & Quillian, 1969). Sin embargo, la terminología de Análisis de Redes Semánticas se comenzó a utilizar sobre todo a partir de que Danowski y Barnett comenzaran a utilizarlos con mayor frecuencia (G. Barnett, Danowski, & Richards, 1993). Carley, por esa misma época, usaba otra terminología cuando se refería al estudio de interacciones entre conceptos: el map analysis. Otras formas de llamar al Análisis de Redes Semánticas menos habituales son los diagramas de conceptos, los grafos conceptuales o las redes conceptuales. Por otra parte, cuando el análisis semántico se ha centrado en el conteo de palabras también se le ha denominado análisis de texto o análisis textual (Carley, 1993).

El SNA permite entender el significado intrínseco de los textos (1997) y se utilizan cada vez más en investigaciones de múltiples disciplinas: el análisis de discursos políticos (Jiang, Anderton, Ronald, & Barnett, 2018); la presencia en redes de social media de un concepto

(Salloum, Al-emran, & Shaalan, 2017) o de una marca (Saez, 2019); en el análisis de temáticas de noticias en determinados países (Jiang, Benefield, Junfei, & Barnett, 2017), en el estudio de la memoria semántica y el estudio del lenguaje (A.M. Collins & Quillian, 1972), etc. Por describir alguno de ellos, el estudio de Jang y Barnett analizaba sobre la cultura de una organización identificando como en el trabajo de Schnegg, grupos de palabras habituales (Jang & Barnett, 1995). En el ámbito de internet, Ruiz y de nuevo Barnett, analizaron cómo se presentaba la información, tanto positiva como negativa, sobre cierta vacuna en la red (Ruiz & Barnett, 2015), etc.

1.2. Caso de Estudio

La metodología propuesta se va a ensayar en las comunicaciones del del 24th Congreso Internacional de Dirección e Ingeniería de Proyectos (CIDIP) celebrado en el año 2020. Este congreso está organizado por la Asociación Internacional de gestión de proyectos (AEIPRO) creada en 1992 para ser un nexo entre los profesionales de la ingeniería de proyectos.

AEIPRO ha llevado a cabo desde su fundación 24 congresos en los cuales se presentan comunicaciones de diferentes áreas de la ingeniería. Concretamente, el congreso se organiza en 10 áreas temáticas: 1. Dirección y Gestión de Proyectos; 2. Ingeniería civil, urbanismo y ordenación del territorio, construcción y arquitectura; 3. Ingeniería de producto, procesos y diseño industrial; 4. Ingeniería ambiental y gestión de recursos naturales; 5. Eficiencia energética y energías renovables; 6. Desarrollo rural y proyectos de cooperación al desarrollo; 7. Tecnologías de la información y las comunicaciones e Ingeniería del software; 8. Gestión de riesgos y seguridad; 9. Innovación educativa en dirección e ingeniería de proyectos; 10. Buenas prácticas y casos de empresa. En este congreso se han presentando varios estudios que vinculan la dirección y planificación de proyectos con los ODS (M.C; Fariña García et al., 2018; Morris Molina et al., 2019; Requelme, Narcisa; Cachipiendo Ulcuango, Charles Jim; Ortiz Tirado, 2019).

En la edición de 2020 se han presentando un total de 191 comunicaciones clasificadas en 10 áreas temáticas como muestra la Tabla 1. Estas comunicaciones han sido clasificadas por sus autores en uno o varios ODS, lo que permitirá tener un dato experto comparable

con los resultados del análisis cuantitativo automático. Hay que señalar que sólo se van a analizar 149 comunicaciones que son las que están escritas en español.

Tabla 1: Número de Comunicaciones por Área Temática en el Congreso CIDIP de 2020.

Área Temática	1. Dirección y Gestión de Proyectos		2. Ingeniería Civil, Urbanismo y Ordenación del Territorio. Construcción y Arquitectura		4. Ingeniería Ambiental y Gestión de Recursos Naturales		5. Eficiencia Energética y Energías Renovables		6. Desarrollo Rural y Proyectos de Cooperación al Desarrollo		7. Tecnologías de la Información y las Comunicaciones. Ingeniería del Software		8. Gestión de Riesgos y Seguridad		9. Innovación Educativa en Dirección e Ingeniería de Proyectos		10. Dirección y Gestión de Proyectos	
Nº de Comunicaciones	27	18	32	39	18	13	6	17	11	10	19	1						

2. Objetivo

Este estudio prueba dos métodos de análisis cuantitativo automático para evaluar el grado de vinculación de las metas de los ODS con los documentos asociados a una organización. Se realiza el ensayo en las comunicaciones presentadas en el 24º Congreso CIDIP.

Existe un trabajo relacionado que desarrolla una metodología de procesamiento del lenguaje natural que permite encontrar párrafos que unen conceptos semánticos de cada unos de los ODS (Galsurkar et al., 2018). Este se asemeja al que se va a desarrollar aquí en el objeto de estudio, sin embargo se diferencia del mismo en la metodología utilizada, siendo la nuestra más sencilla de aplicar sin necesidad de tener conocimientos de programación más complejos. Según los autores de este trabajo los métodos de bolsa de palabras clasificadas o el método tf-idf, sólo tienen en cuenta las frecuencias de las palabras pero ignoran el orden de las palabras y el contexto del texto. Sin embargo nosotros consideramos que por su menor complejidad podría ser útil utilizar el TfIdf siempre y cuando se realice un ajuste del método adecuado.

3. Metodología

3.1. Selección de Términos Clave

El primer paso es elegir el método de preprocesamiento. La entidad de estudio va a ser la palabra. Se van a probar dos métodos: selección de palabras clave de los textos oficiales de las metas de los ODS (United Nations, 2015) con el factor TfIdf y selección de palabras clave de los títulos oficiales de los objetivos. Se han elegido estos dos métodos porque el primero tiene una bolsa de palabras mayor y por tanto más rica, pero puede generar más errores al considerar muchas palabras clave. El segundo método, aunque pueda incurrir en menos errores pero tiene menos riqueza de palabras, por lo que puede quedarse corto a la hora de detectar las palabras vinculadas a los ODS. En el método TfIdf se han eliminado las palabras vacías o stop words de categoría gramatical (pronombres, preposiciones, determinantes, etc.)

3.1.1. Método Tf-Idf

En 1960, Luhn desarrolló el factor Tf (term frequency) que permitía medir una frecuencia relativa de las palabras de un texto para identificar las palabras clave. La dificultad que se encontraba en el uso del Tf (1) era encontrar el límite para discriminar qué términos eran los de frecuencia más alta y más baja (Luhn, 1960). Más tarde Spärk Jones creó un nuevo algoritmo, el Idf (2) o inverse document frequency, que indicaba cómo de importante era una palabra en función del número de veces que aparecía en relación con el resto de palabras de varios documentos, desechando aquellas que aparecían en exceso o en defecto. Este indicador se completó más adelante por otros expertos en la materia y ha dado paso a otros (Robertson & Sparck Jones, 1976), como el Tf-Idf (3), muy utilizada en

la actualidad en todos los buscadores de internet y en el posicionamiento web (Johnson, 2019).

$$Tf = \frac{\text{Frecuencia palabra en documento}}{\text{Total palabras documento}} \quad (1)$$

$$Idf = \log_{10} \frac{N^{\circ} \text{ documentos}}{N^{\circ} \text{ documentos que contienen palabra}} \quad (2)$$

$$TfIdf = Tf \times Idf \quad (3)$$

A continuación se calculará el factor Tf-Idf de cada palabra en cada meta de los distintos ODS para conocer cuáles son las palabras clave de cada objetivo según este factor, eliminando aquellas palabras cuyo factor Tf-Idf es cero.

3.1.2. Método Títulos

Se utilizarán las palabras clave de los títulos oficiales de las metas de los ODS en español (United Nations, 2015) que se extraerán de los mismos eliminando las palabras vacías de categoría gramatical y otras que se han considerado no específicas de los ODS.

Para identificar el grado de vinculación de los ODS con las comunicaciones del congreso, una vez seleccionadas las palabras clave a través de los dos métodos, se realizan búsquedas de las mismas en los textos de las comunicaciones con ayuda del programa NVIVO (QSR International Pty Ltd., 2012). Se realizará una búsqueda de las palabras clave de los ODS en los textos completos de cada una de las 149 comunicaciones presentadas en español repartidas entre las diez áreas temáticas. Una vez realizado el pre-procesamiento se realizará un análisis de la red en función de las palabras comunes entre ODS y temáticas.

3.2. Comparación con el Criterio del Autor

Los autores han seleccionado de uno a tres ODS como aquellos vinculados con su comunicación. Estos ODS seleccionados se compararán con los resultados de los métodos cuantitativos automáticos.

Para comparar el criterio de autor con el resultado de los métodos automáticos es necesario seleccionar los ODS con mayor cantidad de palabras en cada comunicación. Se han seleccionado los tres objetivos con mayor número de palabras por comunicación. Estos objetivos se han comparado con los objetivos seleccionados por los autores para cada una de las comunicaciones.

Finalmente se han representado los resultados de los dos métodos cuantitativos automáticos y el criterio del autor en unos grafos de redes semánticas en los que se relacionan los ODS con las distintas temáticas del congreso. Las redes están compuestas por nodos, que son los ODS y las temáticas y por aristas. Los nodos de mayor tamaño son aquellos con mayor grado de centralidad, es decir, aquellos que están más conectados

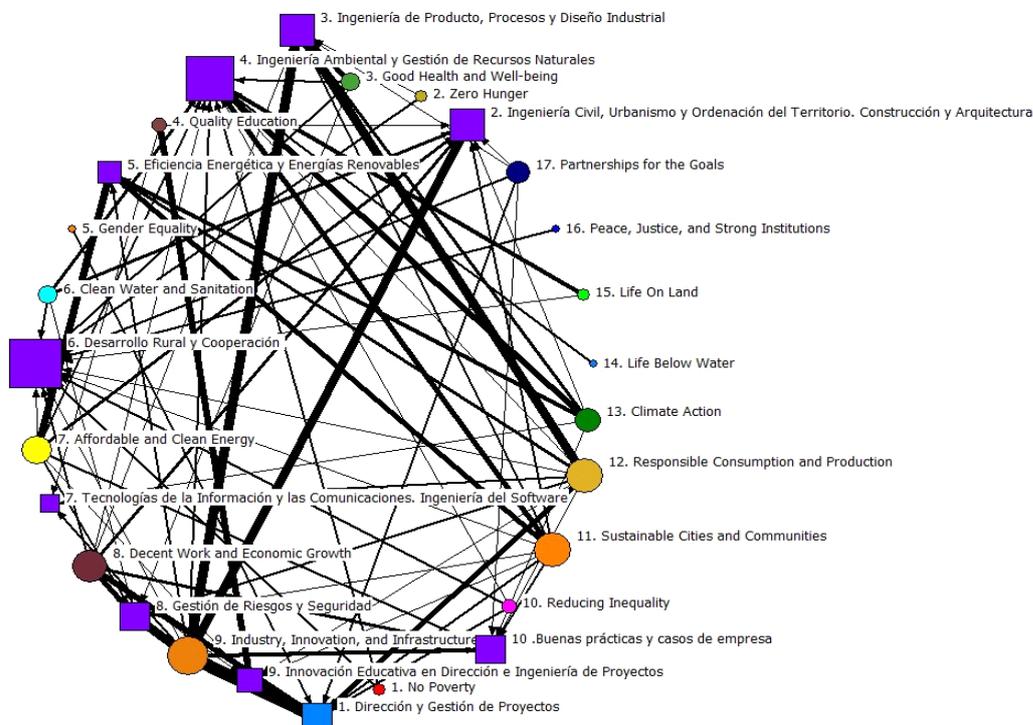
con otros nodos. Las aristas más gruesas reflejan aquellas relaciones con mayor cantidad de palabras y por tanto más fuertemente ligadas.

4. Resultados

La comparación de los métodos cuantitativos automáticos con el criterio de autor permite tener una primera visión de cómo se comportan los métodos automáticos. La Figura 1 muestra la red del número de temáticas de las comunicaciones vinculadas con cada ODS según el criterio de autor. Según este criterio el ODS más conectado es el ODS 9, sobre Industrias, Innovación e Infraestructura. Esto tiene bastante sentido puesto que es un congreso de investigación e innovación en ingeniería en el que participan sobre todo estudiantes e investigadores del área industrial. Le siguen aunque con bastante distancia el ODS 12, sobre Producción y Consumo Sostenible; el 8, sobre crecimiento económico y empleo digno; y el 11 sobre ciudades y asentamientos sostenibles. Los objetivos menos conectados son en este congreso según el criterio de autor son el 16, sobre paz, justicia y

buena gobernanza, el 14, sobre protección de ecosistemas marinos y el 5, sobre género y mujer.

Figura 1: Red según criterio de autor



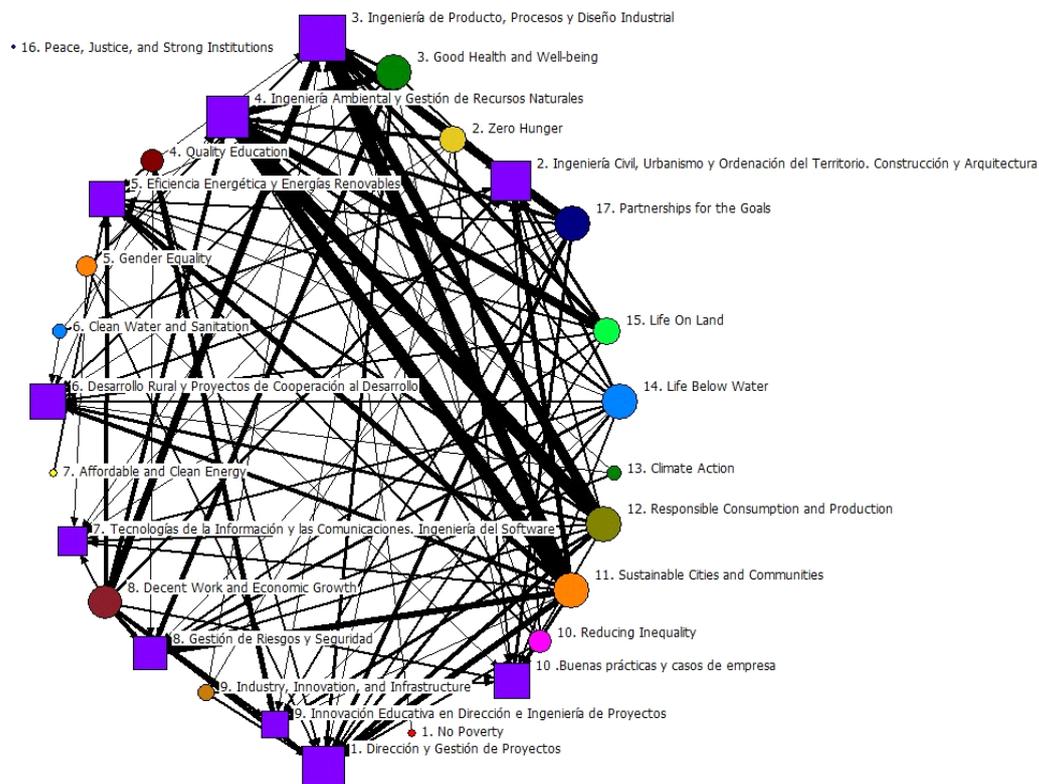
Con respecto a las temáticas más conectadas con los ODS, según el criterio de autor, se observa también en la Figura 2 que son la 6, sobre Desarrollo Rural y Cooperación; la 4, sobre Ingeniería Ambiental y Gestión de Recursos Naturales; la 3, sobre ingeniería de producto, procesos y diseño industrial; y la 2, sobre Ingeniería civil y ordenación del territorio, construcción y arquitectura. Las conexiones más fuertes se dan entre el ODS 9 y las temáticas 2 y 3 y el ODS 12 y la temática 3. Las temáticas menos conectados son la 7, de tecnologías de la información y la 9, sobre innovación educativa y la 5 sobre eficiencia energética.

En la Figura 2. Se representan los resultados en red utilizando el método Tf-Idf. El resultado muestra que los objetivos más conectados son el 11, sobre ciudades sostenibles, el 12, sobre producción y consumo sostenible y el 8 sobre crecimiento y empleo digno. En cambio, en este método el ODS 9 tiene muy pocas conexiones. Entre los menos

conectados están el 16 que tiene cero conexiones y está fuera de la red, el 7, sobre energía sostenible y el 1 sobre la erradicación del hambre.

Con respecto a las temáticas, se observa también en la Figura 2, que entre las más conectadas se encuentran al igual que en resultado del criterio de autores la 3, la 2, la 4. Sin embargo, en vez de aparecer la 6 aparece la 10, sobre buenas prácticas y casos de empresa. Las temáticas con menos conexiones son la 9 y la 7, igual que en el criterio de autor, sin embargo la 5 a diferencia del criterio de autor sale con un valor más alto. Las conexiones más fuertes en este método están entre entre las temáticas 3 y 4 y los ODS 11 Y 12.

Figura 2: Red según TFIDF



La Figura 3 muestra la red con el método títulos. En este método el ODS más conectado es el ODS 12, sobre producción y consumo sostenible y el 9, sobre infraestructura, industrias e Innovación, el 8, sobre crecimiento económico y empleo digno; y el 11 sobre ciudades y asentamientos sostenibles. El resultado se parece bastante al del criterio de autor. Los objetivos menos conectados son el 10, sobre igualdad; el 15 sobre la protección

de ecosistema y el 14 sobre la protección de ecosistemas marinos. En los ODS menos conectados no se parece al criterio de autor salvo en el en el ODS 14.

Con respecto a las temáticas más conectadas se encuentran la 4, la 3 y la 2 de nuevo. Pierde peso en comparación con el criterio de autor la temática 6. Las temáticas menos conectadas son la 9, la 10 y la 5 coincidiendo en la 9 el criterio de autor.

Figura 3: Red según criterio títulos

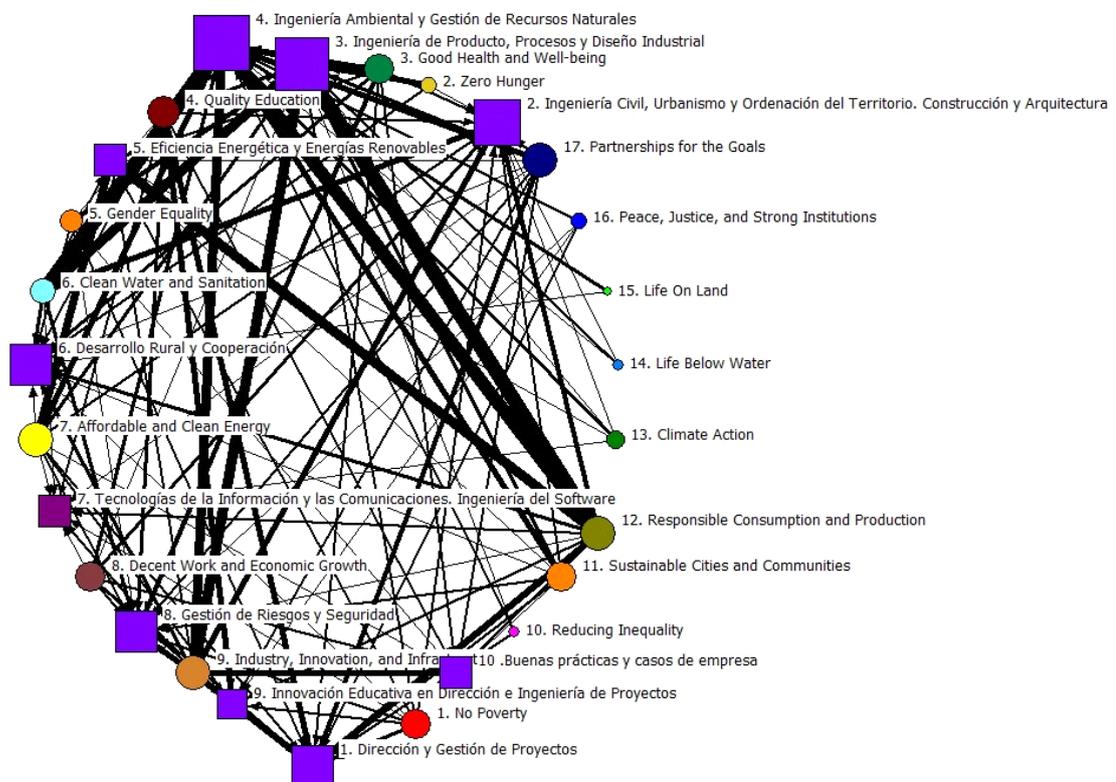


Tabla 2: Porcentaje de coincidencia de los métodos con el criterio del autor.

	% Coincidencia de al menos 1 ODS con el criterio de Autor	% Coincidencia de al menos 2 ODS con el criterio de Autor
Criterio TfIdf	47,10	4,34
Criterio Títulos	78,98	10,14

5. Discusión y Conclusiones

Se han encontrado diferencias entre los dos métodos a la hora de vincular los ODS con las temáticas del congreso AEIPRO. De entrada se supone que el criterio de autor es una opinión experta y por tanto la información sobre los ODS es fiable, cuanto más se parezcan los resultados de los métodos automáticos a los datos del criterio de autor, más fiable será el método.

Sin embargo, aunque los resultados han mostrado algunas semejanzas con el criterio de autor en ODS y temáticas más conectadas, también existen algunas diferencias. El método títulos parece asemejarse más a la red del criterio de autor aunque apenas se parece en los ODS menos conectados. Hay que tener en cuenta que las conexiones dependen del número de palabras clave, a más conexión mayor es el número de palabras clave coincidentes entre las comunicaciones y los ODS y a menos conexión menor es el número de palabras coincidentes. A mayor riqueza de palabras en el caso del método Tf-Idf las conexiones más bajas se asemejan más al criterio de autor. Esto significa que aunque haya pocas conexiones estas son de buena calidad. En el caso del método títulos, las conexiones de los ODS menos conectados son menos parecidas al criterio de autor, lo cual significa que esas pocas palabras que conectan son de peor calidad. Las altas conexiones, sin embargo, reflejan lo contrario, al tener más palabras clave en el método Tf-Idf el resultado se aleja del criterio de autor y el método títulos se acerca. Esto puede deberse a que los autores a la hora de elegir los ODS vinculados a sus comunicaciones no lean nada más que el título y no conozcan la gran cantidad de interrelaciones entre los ODS que se pueden tener en cuenta y además, al elegir tan solo dos o tres ODS relacionados, la información aportada pierde determinadas vinculaciones que si refleja el Tf-Idf.

Se observa también que en ambos métodos entre los ODS más conectados se encuentran el 12, sobre producción y consumo sostenible y el 11, sobre ciudades y asentamientos sostenibles. El hecho de que el 9 salga más bajo en el método Tf-Idf puede deberse a tres razones: La primera causa podría ser que este congreso realmente este más vinculado a ellos en contra de los decían los autores; La segunda causa es que este ODS esté relacionado con el 12 y el 11 en las palabras clave hayan detectado las comunicaciones vinculadas al mismo con los objetivos 12 y 11. La tercera causa es que el método Tf-Idf no funcione para detectar las vinculaciones de las comunicaciones con los ODS. Para poder ajustar este método será necesario confirmar la fiabilidad del criterio de autor o probar este mismo método en organizaciones.

No obstante, este estudio ha servido para detectar diferencias entre métodos de automatización que permiten valorar por donde debe continuar en una futura línea de

investigación para encontrar un método fiable que pueda ser utilizado en los documentos de cualquier organización.

6. Referencias Bibliográficas

- A.M. Collins, & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. *L. W. Gregg (Ed.), Cognition in Learning and Memory*, 117–138.
- Ait-Kadi, M. (2016). Water for Development and Development for Water: Realizing the Sustainable Development Goals (SDGs) Vision. *Aquatic Procedia*, 6, 106–110. <https://doi.org/10.1016/j.aqpro.2016.06.013>
- Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. *R&D Management*, 50(3), 329–351.
- Barnett, G., Danowski, J. A., & Richards, W. D. (1993). Communication networks and network analysis : A current assessment. In G. A. Barnett & W. D. Richards (Eds.), *Progress in Communication Sciences* (p. 90). New Jersey: Ablex Publishing Corporation.
- Boeris, C. E. (2013). Aplicación de técnicas de análisis de redes sociales y de co-ocurrencia de palabras en la determinación de frentes de investigación. En *55a Reunión Nacional de la Asociación Argentina de Astronomía*. La Plata: Instituto Argentino de Radioastronomía.
- Callon, M., ,J. Law, y A. & R. (1986). *Mapping the dynamics of science and technology: Sociology of science in a real world*. The Macmillan Press Ltd.
- Carley, K. (1993). Coding choices for textual analysis: A comparison of content analysis and map analysis. *Sociological Methodology*, 23:75-126. *Sociological Methodology*, 23, 75–126.
- Collins, A. M., & Loftus, E. F. (1975). A Spreading Activation Theory of Semantic Processing. *Psychological Review*, 82, 407–428.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval Time from Semantic Memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240–247.
- Contreras Barrera, M. (2018). Aplicación del algoritmo RAKE en la indización de documentos digitales. *Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información*, 32(75), 109. <https://doi.org/10.22201/iibi.24488321xe.2018.75.57951>
- Eito Brun, R., & Senso, J. A. (2004). Minería textual. *El Profesional de La Información*, 11–27.
- Fariña García, M.C.; De Nicolás De Nicolás, V.L; Yagüe Blanco, J. L., & Zamorano, R. (2018). Los Objetivos de Desarrollo Sostenible como Marco para los Proyectos: Análisis de du Influencia a partir de Redes Semánticas. XXII Congreso Internacional de Dirección e Ingeniería de Proyectos (pp. 1757–1769).
- Fariña García, M.C., De Nicolás De Nicolás, V. L., Yagüe Blanco, J. L., & Labrador Fernández, J. (2021). Semantic network analysis of sustainable development goals to quantitatively measure their interactions. *Environmental Development*, 37(November 2020). <https://doi.org/10.1016/j.envdev.2020.100589>
- Fayyad, U., Piatetsky-shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge

Discovery in, 37–54.

- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook : advanced approaches in analyzing unstructured data*. Cambridge University Press.
- Fuso Nerini, F., Tomei, J., To, L. S., Bisaga, I., Parikh, P., Black, M., Mulugetta, Y. (2018). Mapping synergies and trade-offs between energy and the Sustainable Development Goals. *Nature Energy*, 3(1), 10–15. <https://doi.org/10.1038/s41560-017-0036-5>
- Galsurkar, J., Singh, M., Wu, L., Vempaty, A., Sushkov, M., Iyer, D., Varshney, K. R. (2018). Assessing national development plans for alignment with sustainable development goals via semantic search. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, (February), 7753–7758.
- Guha, J., & Chakrabarti, B. (2019). Achieving the Sustainable Development Goals (SDGs) through decentralisation and the role of local governments: a systematic review. *Commonwealth Journal of Local Governance*, 1–21. <https://doi.org/10.5130/cjlg.v0i22.6855>
- Jang, H., & Bamett, G. A. (1995). Cultural differences in organizational communication: A semantic network analysis. *Bulletin de Methodologie Sociologique*, 44, 31–59.
- Jiang, K., Anderton, B. N., Ronald, P. C., & Barnett, G. A. (2018). Semantic Network Analysis Reveals Opposing Online Representations of the Search Term “ GMO .” *Global Challenges*, 2(1), 1700082. <https://doi.org/10.1002/gch2.201700082>
- Jiang, K., Benefield, G. A., Junfei, Y., & Barnett, G. A. (2017). Mapping Articles on China in Wikipedia : An Inter-Language Semantic Network Analysis. In *Proceedings of Hawaii International Conferences on System Science (HICSS-50)* (pp. 2233–2242).
- Justicia De la Torre, M. del C. (2017). *Nuevas Técnicas de Minería de textos: aplicaciones*. Universidad de Granada. Escuela Técnica Superiores de Ingeniería Informática y de Telecomunicación.
- Klopp, J. M., & Petretta, D. L. (2017). The urban sustainable development goal: Indicators, complexity and the politics of measuring cities. *Cities*, 63. <https://doi.org/10.1016/j.cities.2016.12.019>
- Krippendorff, K. (2004). *Content Analysis. An Introduction to Its Methodology*. T. (Sage, Ed.), Thousand Oaks. California.
- Le Blanc, D. (2015). Towards Integration at Last? The Sustainable Development Goals as a Network of Targets. *Sustainable Development*, 23(3), 176–187. <https://doi.org/10.1002/sd.1582>
- Luhn, H. P. (1960). Keyword-in-context index for technical literature American Documentation, 1960: 11(4):288–295. *American Documentation*, 11(4), 288–295.
- Morris Molina, L. H., Salazar de Morris, O. J., & Barrientos, E. (2019). XXIII Congreso Internacional de Dirección e Ingeniería de Proyectos, 029(July), 1584–1594. Recuperado de: http://dspace.aepro.com/xmlui/bitstream/handle/123456789/2408/AT10-001_2019.pdf?sequence=1&isAllowed=y
- Naciones Unidas. Sustainable Development Network Solutions. (2014). *Indicators and a monitoring framework for Sustainable Development Goals Launching a data revolution for the SDGs*. Recuperado de: <http://unsdsn.org/wp->

content/uploads/2014/11/141125-Indicator-working-draft-WEB.pdf

- Naciones Unidas (2015). Resolución de la Asamblea *General A/RES/70/1: Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible*. Recuperado de: https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E
- Naciones Unidas, (2014). Resolución de la Asamblea *General A/RES/69/700: El camino hacia la dignidad para 2030: acabar con la pobreza y transformar vidas protegiendo el planeta*.
- Nagy, J. A., Benedek, J., & Ivan, K. (2018). Measuring sustainable development goals at a local level: A case of a Metropolitan Area in Romania. *Sustainability (Switzerland)*, 10(11). <https://doi.org/10.3390/su10113962>
- Nilsson, M., Chisholm, E., Griggs, D., Howden-Chapman, P., McCollum, D., Messerli, P., ... Stafford-Smith, M. (2018). Mapping interactions between the sustainable development goals: lessons learned and ways forward. *Sustainability Science*, 13(6), 1489–1503. <https://doi.org/10.1007/s11625-018-0604-z>
- Ntona, M., & Morgera, E. (2018). Connecting SDG 14 with the other Sustainable Development Goals through marine spatial planning. *Marine Policy*, 93(July 2017), 214–222. <https://doi.org/10.1016/j.marpol.2017.06.020>
- Poschen, P. (2015). *Decent Work, Green Jobs and the Sustainable Economy: Solutions for Climate Change and Sustainable Development*. Leeds(England): Green Leaf Publishing.
- QSR International Pty Ltd. (2012). *NVivo qualitative data analysis Software*. Recuperado de: <http://www.qsrinternational.com>
- Requelme, Narcisa; Cachipuendo Ulcuango, Charles Jim; Ortiz Tirado, P. X. (2019). La Sostenibilidad en la Construcción de Políticas Públicas, Planes, Programas y Proyectos de Desarrollo Rural en Ecuador. XXIII Congreso Internacional de Dirección e Ingeniería de Proyectos (pp. 1584–1594). Recuperado de: http://dspace.aepro.com/xmlui/bitstream/handle/123456789/2408/AT10-001_2019.pdf?sequence=1&isAllowed=y
- Robertson, S. E., & Sparck Jones, K. (1976). Relevance Weighting of Search Terms. *Journal of the American Society for Information Science*, 27(3), 129–146.
- Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic Keyword Extraction from Individual Documents. *Text Mining: Applications and Theory*, (October 2017), 1–20. <https://doi.org/10.1002/9780470689646.ch1>
- Ruiz, J. B., & Barnett, G. A. (2015). Exploring the presentation of HPV information online: A semantic network analysis of websites. *Vaccine*, 33(29), 3354–3359. <https://doi.org/10.1016/j.vaccine.2015.05.017>
- Saez, A. (2019). Marcas que hacen noticias y diarios que hacen marcas. *Ad Comunica. Revista Científica de Estrategias, Tendencias e Innovación En Comunicación*, (17), 215–217. <https://doi.org/10.6035/2174-0992.2019.17.13>
- Salloum, S. A., Al-emran, M., & Shaalan, K. (2017). Mining Social Media Text : Extracting Knowledge from Facebook. *International Journal of Computing and Digital Systems*, 6(2), 73–81. <https://doi.org/10.12785/ijcds/060203>
- Salton, G., Wong, A., & Yu, C. (1976). Automatic indexing using term discrimination and

term precision measurements. *Information Processing & Management*, 12(1), 43–51.

Sowa, J. F. (1987). Semantic networks. Recuperado de:
<http://www.jfsowa.com/pubs/semnet.htm>

Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28, 11–21.

Turney, P. D. (2000). Learning algorithms for keyphrase extraction. *Information Retrieval*, 2(4), 303–336.

Verd Pericas, J. M. (2005). El uso de la teoría de redes sociales en la representación y análisis de textos. De las redes semánticas al análisis de redes textuales. *Revista de Metodología de Ciencias Sociales*, 10, 129–150.

Comunicación alineada con los Objetivos de

Desarrollo Sostenible

