**Tema 4: Valor de los datos​**

Sonia Castro Yama

MsC. Economía, Especialista en Analítica de Datos

Luis Carlos Torres Soler

PhD. Pensamiento complejo, MsC. Ing. Sistemas, Matemático

Millones y millones de datos se generan a cada segundo en el mundo, quizá más de un millón en cada organización; se procesan por medios computacionales por su volumen y con el fin de generar otros a utilizar en la toma de decisiones como información. Sin embargo, estos son de tipo cuantitativo. Los de carácter cualitativo solo están en la mente de las personas. En general, para esos procesos computacionales, en las organizaciones, se considera solamente datos que son numéricos; es decir, cantidades (cuantitativos). Hay que reflexionar bastante cuando se presentan datos con diferentes características no cuantificables, los cuales también proveen información y conocimiento.

Variadas preguntas surgen cuando se debe analizar datos cualitativos, una de las cuales es: ¿se pueden usar métodos categóricos (no numéricos) para el aprovechamiento del valor de los datos en el sector público?

1. **Metodología**

El primer aspecto en el estudio es la caracterización de la fuente de información. Se toman los datos que arrojan las encuestas que realiza la Organización para la Cooperación y el Desarrollo Económico (OCDE), la cual se realiza cada dos años la encuesta *STIP Compass*. Esta encuesta se hace desde el 1993 en más de 50 países en temas de iniciativas para monitorear y analizar políticas públicas en las áreas de CTI (Ciencia, Tecnología e Innovación). Esta encuesta incluye iniciativas que se generan en diferentes ministerios e instituciones nacionales del sector gobierno, con competencia en dominios tan amplios como investigación, innovación, educación, industria, medio ambiente, trabajo, entre otros, y su fin es identificar las iniciativas de políticas relevantes sobre transferencia de conocimiento y co-creación en una región o en el país.

**¿De dónde provienen los datos de STIP Compass?**

La principal fuente de datos de *STIP Compass* son las respuestas a la encuesta sobre políticas de CTI de la CE y la OCDE, la cual se realiza cada dos años. La edición más reciente es 2021**[[1]](#footnote-1)**. Los datos se recopilan bajo los auspicios del Comité de Política Científica y Tecnológica de la OCDE (CSTP) y el Comité del Espacio Europeo de Investigación e Innovación (ERAC) de la Unión Europea. La encuesta se realiza a funcionarios del gobierno nacional que trabajan en políticas de CTI en una variedad de organizaciones de la administración pública (Ministerios, instituciones centralizadas y decentralizadas, agencias, institutos y organizaciones). Se emplea un cuestionario en línea (software de código abierto LimeSurvey).

La encuesta tiene un alcance amplio y abarca cuestiones de política en torno a la investigación pública, la innovación y el espíritu empresarial, la transferencia de conocimientos, las habilidades para la innovación, la innovación para los desafíos sociales y la gobernanza del sistema de CTI. En su mayor parte, se pide a las personas responsables que enumeren y caractericen las iniciativas de política que están implementando para abordar un desafío particular, como el déficit de innovación que se encuentra típicamente en las pymes o las brechas de transferencia de conocimiento entre universidades y empresas. Todas las iniciativas de políticas se caracterizan mediante una plantilla estándar. Esa plantilla utiliza taxonomías de instrumentos de política y grupos destinatarios que permiten la comparación de iniciativas de política dentro y entre países.

El estudio realizado (investigación) lleva a determinar que puede proponerse como una metodología para la utilización de datos categóricos (no numéricos –cualitativos), también emplear programas que identifican textos y utilizar software con diferentes algoritmos para generar redes. Por tanto, este estudio utilizó el programa Gephi**[[2]](#footnote-2)** (uno de los varios que existen). Este programa se trata de una herramienta que genera a partir de los datos que se incluyan, redes (gráficos), con la característica de poder adherirse colores o tamaños que permitan identificar patrones, formas, concentraciones. El principal fin es ayudar a quien analiza datos (cuantitativos y cualitativos), realizar observaciones que le permitan inferir hipótesis, patrones, como también aislar singularidades o fallas de las estructuras de los datos. Ayuda, de cierta forma, a completar lo que se halla por estadística descriptiva, básicamente es un motor de renderizado. Se trata de una herramienta para el reconocimiento del pensamiento visual. Posee variedad de interfaces, las cuales intentan facilitar el razonamiento visual de los hacedores de políticas públicas con benchmarking entre los países.

El estudio muestra cómo se puede realizar análisis de datos cualitativos a partir de emplear la ciencia de redes, al emplear programas computacionales (software) que faciliten la generación de las redes, que de manera simple son grafos donde deben percibirse diferentes aspectos, en particular a cómo se agrupan formando interrelación de iniciativas, que a la vez se puede inferir de cómo se desarrollan diferentes políticas públicas a través de la colaboración y cooperación.

La ciencia de redes (Network Science), es una disciplina bastante nueva y emergente, surge por los años 1980, a partir de la necesidad de elaborar interconexiones e interrelaciones diversas entre una gran variedad de categorías, clases, o datos, con miras a generar conocimiento. La idea inicial es obtener cierto tipo de información que se genera a partir de los gráficos en las redes neuronales biológicas, y que se reproducen en procesos cognitivos, y de allí se empezaron a generar distintas redes sociales de manera computacional.

Esta área de la ciencia ayuda a descubrir principios, aspectos, cualidades, como a elaborar algoritmos y herramientas que mejoren la visualización de los gráficos (redes) y, en lo posible, mostrar el comportamiento de las redes que se forman.

La identificación y caracterización de la información que puede obtenerse al procesar un gran volumen de datos conduce a determinar patrones, para el estudio, los temas de las iniciativas de políticas públicas. Se trabaja con la escogencia de los países a estudiar, y se realiza la respectiva descarga de la plataforma STIP COMPASS de: Canadá, Chile, Colombia, Japón, Corea, México, EE. UU, Reino Unido, España.

Posterior a esto, se determinan las variables relevantes para el estudio, considerando que no se observan estudios que hayan realizado análisis sobre las iniciativas públicas; se incorpora el presupuesto de cada iniciativa en millones de euros en la investigación, para la Agenda sólo se toman aquellos que muestran mayor relevancia para el objetivo de la convocatoria. En otras palabras, se acata la propuesta que hace el equipo consultor de la CAF, los cuales se enumeran a continuación:

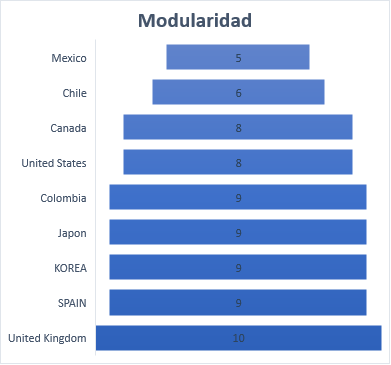
**Temas**

1. Inteligencia artificial (IA)
2. Estrategias de política de innovación empresarial
3. Políticas de clúster
4. Investigación e innovación colaborativas
5. Transformación digital de empresas
6. Ética de las tecnologías emergentes
7. Apoyo financiero a la I + D + i empresarial y la innovación
8. Coordinación de políticas horizontales
9. Movilidad intersectorial
10. Compromiso de múltiples partes interesadas
11. Plan o estrategia nacional de CTI
12. Apoyo no financiero a la I + D e innovación empresarial
13. Ciencia abierta y mejor acceso a publicaciones y datos de investigación
14. Investigación e innovación para países en desarrollo
15. Estrategia de investigación e innovación para la sociedad
16. Cultura de ciencia, tecnología e innovación
17. Estimular la demanda de innovación y creación de mercado.
18. Inteligencia de políticas estratégicas
19. Apoyo específico a las pymes
20. Transferir y compartir estrategias.

El programa Gephi permite modificar modularidades en una matriz totalmente cualitativa, para lo cual se homogeniza la base de datos (incluida la mayúscula y minúscula que se empleen). La metodología parte de efectuar primero una minería de datos a cada uno de los temas de iniciativas que existen sobre los distintos países, esta escogencia y transformación es la base para el programa Gephi, que mediante los diversos algoritmos que tiene involucrados (en nuestro caso el algoritmo Yifan Hu), se espera hallar los patrones que se generan de esas iniciativas.

PATRONES

Las redes que se generan por medio del programa Gephi lleva a concebir que se forman patrones de las iniciativas públicas; por tanto, permite visualizar que estos brindan una gama de posibilidades para identificar nodos o clústeres, es decir, esos nodos donde se podría desarrollar mayor articulación entre los actores y, por ende, se podría generar modelos de políticas públicas más eficientes, eficaces y, desde luego, mayor cobertura y focalización, lo que va a brindar mayores beneficios.

****

Cada Modularidad o Clúster se representa con colores en las gráficas, en cada país. Con los 9 países analizados se crearon 73 modularidades.

Los nodos son los puntos que representan las iniciativas con sus políticas y las aristas con las líneas que conectan un nodo con el otro, en su generalidad son curvas.

En total los Nodos estudiados son 814 y 1010 aristas.

Variedad de gráficos (redes) se pueden formar a partir de emplear el programa Gephi, empleando en él, el algoritmo Yifan Hu. De las redes que se visualizan, se puede percibir que existen nodos de los temas focalizados en el valor de los datos de las iniciativas de los gobiernos de los países que se tomaron para el estudio: Canadá, Chile, Colombia, Japón, Corea, Estados Unidos, Reino Unido, México y España.

**Canadá**

Texto, Carta

Descripción generada automáticamenteCanadá, presenta 111 nodos y 147 aristas, un grafo muy robusto sobre las iniciativas que desemboca en ocho (8) tipos de modularidad o clústeres con 127 iniciativas a partir de los datos considerados. La iniciativa más recurrente en este país es la Investigación e innovación colaborativa (Collaborative Research and Innovation).

**Chile**

Texto

Descripción generada automáticamente

Chile presenta 79 iniciativas, 78 nodos y 101 aristas. En este país la iniciativa más recurrente es Apoyo financiero a la I+D+i empresarial (Financial Support to Business R&D and Innovation), con 22 iniciativas enfocadas durante el periodo analizado y 6 modularidades.

**Colombia**

Imagen que contiene Texto

Descripción generada automáticamente

Colombia, lo más particular del país cafetero, es la gran cantidad de iniciativas enfocadas en inteligencia artificial. Estas alcanzan 10 iniciativas. Este indicador no se presenta en ningún otro país analizado, aunque se reportan iniciativas con el nombre, no son catalogadas como de inteligencia artificial. Lo otro muy particular, es la cantidad de modularidad que reportan, muy similares a los reportadas por Japón, Corea y España que reportan 9, por encima de Estados Unidos y Canadá, que reportan 8 y, por debajo, del Reino Unido que reporta 10. Colombia reporta 83 Nodos y 95 artistas.

**Japón**

Imagen que contiene Texto

Descripción generada automáticamente

Japón a diferencia de los demás países desarrollados presenta una disminución sustancial en el número de nodos y aristas (53 y 55 respectivamente), por lo que se presume una focalización sistemática de sus iniciativas; sin embargo, una modularidad de 9 y 35 iniciativas focalizadas. La iniciativa que más se incentiva es Ética de las tecnologías emergentes, con 6 iniciativas, sobre el tema y 9 modularidades.

**Corea**

Texto

Descripción generada automáticamente

Corea presenta 124 iniciativas enfocadas. La más recurrente es la de temas relacionados con la Coordinación de Políticas Horizontales. Corea también cuenta con 9 modularidades, y un robusto conjunto de nodos y aristas (105 y 141 respectivamente).

**México**

Imagen que contiene Carta

Descripción generada automáticamente

México tiene (comparado con el resto de los países) 42 nodos y 49 aristas, algo más bajo que los demás países analizados; su modularidad solo llega a 5 y no hay una focalización de iniciativas. Se evidencia una atomización con todas las iniciativas y 29 iniciativas en temas de datos.

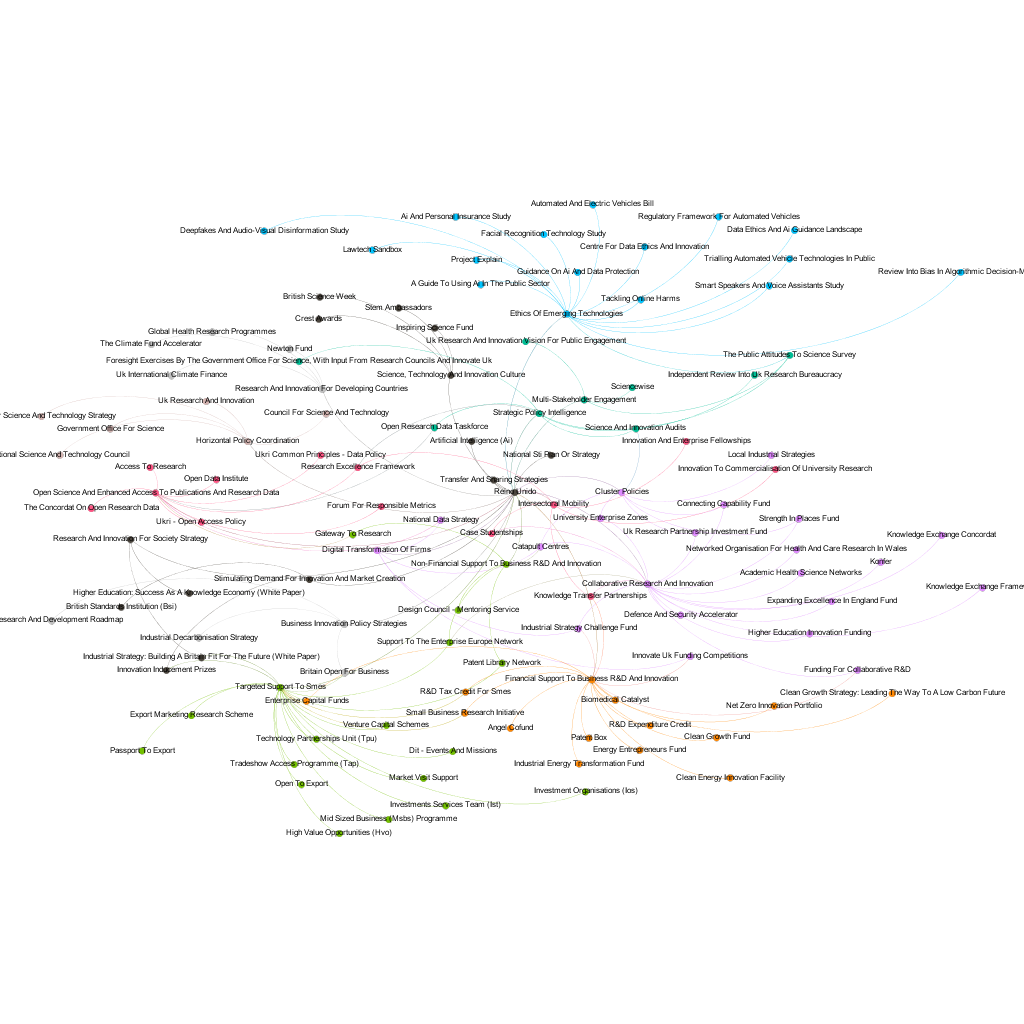
**España**

Imagen que contiene Texto

Descripción generada automáticamente

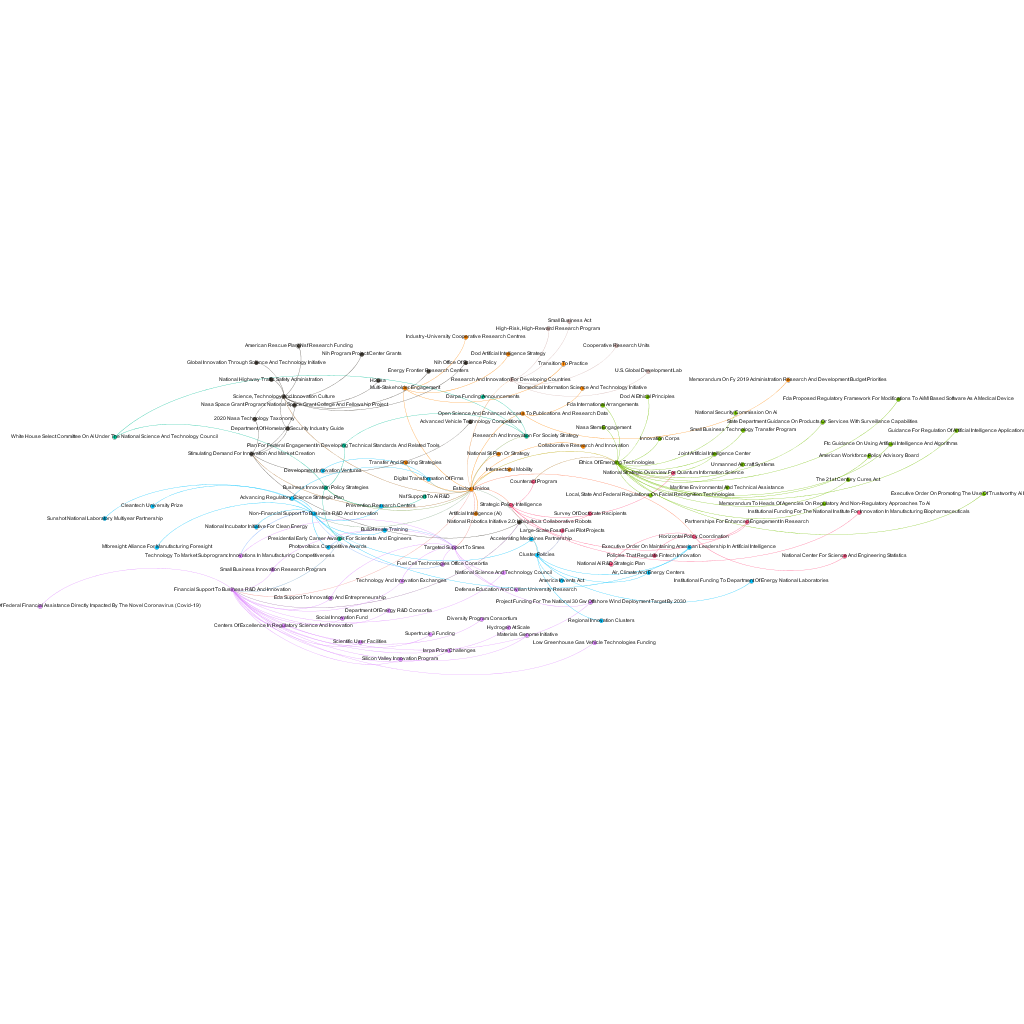
España presenta 9 modularidades, 112 nodos y 141 aristas y un nutrido sistema de iniciativas, 121 en temas de datos, de las cuales la más recurrente es la de Apoyo financiero a I+D+i empresarial, la cual presenta 18 iniciativas en este tema.

**Reino Unido**



Reino Unido presenta un robusto sistema de iniciativas focalizadas en datos (133), de las cuales la más recurrente es la de “Apoyo específico a las Pymes” (Targeted Support to SMEs) y una modularidad de 10, 121 nodos y 151 aristas.

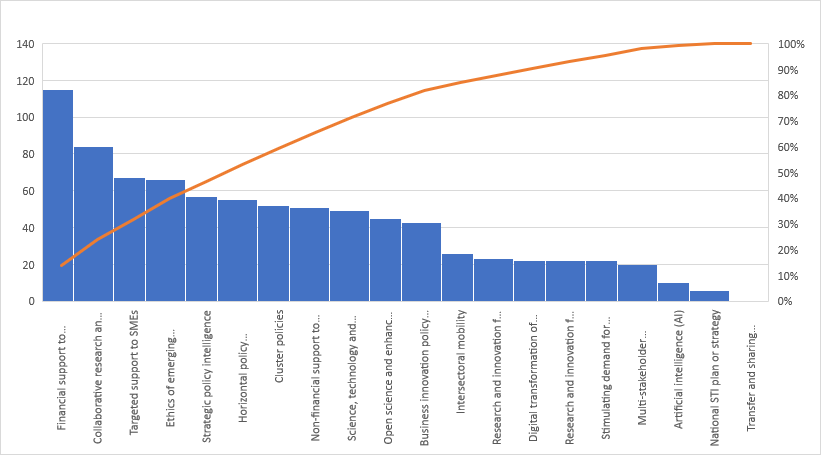
**Estados Unidos**



Estados Unidos presenta 110 iniciativas, la más recurrente es la de Ética de las tecnologías emergentes (Ethics of Emerging Technologies) con 20 iniciativas; presenta 8 modularidades, 109 nodos y 130 aristas.

El conjunto de países analizados (Canadá, Chile, Colombia, Japón, Corea, México, España, Reino Unido, Estados Unidos) permite generar un Pareto de recurrencia de las iniciativas, que lo conforman el Apoyo financiero a la I+D+i empresarial, Investigación e innovación colaborativa, Apoyo específico a las pymes y Ética de las tecnologías emergentes.

**Análisis de recurrencia de iniciativas del grupo de países analizados.**



En la gráfica se percibe la recurrencia de iniciativas que ocurren en los países, esto, como quizá las que se consideran de mayor relevancia; sin embargo, como se indicó, no a todas se les asigna adecuados recursos económicos.

**Presupuestos invertidos**

Es importante identificar cuál es la inversión realizada por los gobiernos para las iniciativas de políticas en CTI, y compararlas entre ellos. Se observa que el país que más invierte en datos es Canadá, en el cual, la iniciativa que más recursos o la que más se financia es: Apoyo Financiero a la I+D+i Empresarial (Financial Support to Business R&D and Innovation). Le sigue Corea y Reino Unido. En los países latinos, el de mayor inversión en políticas de CTI es Colombia, y el que menos invierte es México. La iniciativa en que más se invierte, en promedio con foco en datos es: Apoyo Financiero a la I+D+i Empresarial (Financial Support to Business R&D and Innovation).

Además, ante el desarrollo tecnológico para diferentes ámbitos: exploración espacial y marítima, defensa nacional, apropiación de recursos para ‘reducir’ pobreza, conflictos sociales, hambruna, la mejor forma es tener conocimiento sobre las distintas dimensiones en que se desarrollan las situaciones, y las distintas dimensiones en que la tecnología se requiere y avanza; considerando que la implementación de tecnologías conducen a mejorar contextos como la educación, la administración pública y la seguridad.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Conclusiones**

La mirada cualitativa a los datos, en particular a los gráficos (redes) que se obtienen por el software, permite inferir que en los diferentes países existen clústeres (nodos) que poseen gran influencia sobre otros, siendo los mayores generadores de iniciativas en políticas de CTI. No es sencillo determinar cuál es el impacto que de una o varias políticas; sin embargo, se conoce que existe inversión para el desarrollo. Es importante concebir qué países como Colombia, país en desarrollo, le aporte buen presupuesto para inclusión de las nuevas tecnologías.

Determinándose cuáles son los clústeres de mayor ascendencia, permite que se piense en cuáles son las alianzas estratégicas que pueden realizarse con varios propósitos: primero, focalizar las iniciativas de política en CTI-DATOS; segundo, establecer mejores planes de acción de manera que la iniciativa tenga mayor impacto en la población (región o país). Esto ayuda a que las políticas en CTI-DATOS sean más aprovechadas.

El uso de los programas computacionales (software) permite generar variedad de redes (gráficos), a partir de realizar simulaciones diversas, variando los parámetros que el mismo software permite. La visualización facilita considerar cuáles clústeres se alteran al variar ciertos parámetros, cuál es la incidencia de una política, cómo se altera la red a la inclusión o borrado de algunos nodos; además, al establecer o quitar relaciones entre nodos. Es decir, los programas computacionales conducen a desarrollar planes estratégicos de mayor relevancia, no solo para este tipo de iniciativas, sino también se pueden emplear en diferentes acciones, en especial de la administración pública.

Hoy día se genera variedad de software para establecer comunicabilidad entre personas de regiones muy apartadas, con distintas lenguas y con diferentes propósitos, que determina que se forman redes sociales (Facebook, Instagram, WhatsApp, …), sin embargo, también se establecen otro tipo de redes, como son de conocimiento, de investigación, de gobierno, entre muchas otras, y estas se desarrollan por y para determinados grupos de población.

En las unidades de gestión tecnológica o de Innovación son variadas las decisiones estratégicas que deben considerarse, pues es importante determinar el momento en que debe crearse, adquirirse, perfeccionarse, asimilarse la inclusión y uso de nuevas tecnologías, lo cual se hace por medio de diferentes proyectos estratégicos de investigación o desarrollo, considerando cuáles son las variables esenciales como necesidades para una región.

Por último, este tipo de estudios, puede ser el insumo para estudios macroeconómicos o econométricos que se quieran desarrollar, pues se propone, determinar correlaciones entre el número de nodos o modularidades de los países y el desarrollo en un escenario dinámico.

Hay que ver cómo el valor de los datos toma relevancia al emplearse el análisis por medio de las redes, pues se alcanza a percibir grandes diferencias y ventajas entre la investigación cualitativa y cuantitativa, en particular para estudios macroeconómicos, donde se puede incorporar variables como el PIB, desempleo; además, realizar correlaciones con el fin de hacer una identificación de los factores que determinan el crecimiento de los países directamente con iniciativas y/o políticas públicas.

Como punto final, varias concepciones existen sobre la inteligencia, se le considera como la habilidad para detectar variaciones en el ámbito y en el entorno, aprender y ajustar lo que es o se posee; en pocas palabras, saber dónde vive, cómo es su entorno, qué puede hacer, cuándo debe actuar, por qué y para qué de una forma y no de otra. Esta premisa lleva a mirar que existe un mundo de posibilidades; por tanto, este estudio puede servir de motivación para realizar más estudios sobre el tema.

Los datos con que se puede trabajar son millones, y aunque solo se toman unos países considerados relevantes en el campo de CTI, aun así, el procesamiento de datos debe realizarse por medios computacionales, aun tratándose de datos cualitativos, pero en este caso se emplea las redes, donde se pueden percibir aspectos de trascendencia para generar conocimiento como cómo se desarrollan las políticas institucionales con miras al desarrollo tecnológico de una región o de un país, y con esto, el valor de los datos.

**Glosario**

**STIP Compass:** es una iniciativa conjunta de la Comisión Europea **(CE)** y la **OCDE** que tiene como objetivo recopilar datos cualitativos y cuantitativos sobre las tendencias nacionales en las políticas de ciencia, tecnología e innovación (CTI). El portal apoya el monitoreo y análisis continuo de las políticas de CTI de los países y pretende ser una plataforma central para la investigación de políticas y el asesoramiento que apoye a los funcionarios gubernamentales, analistas y académicos. A través de diversos vínculos se puede explorar y descargar datos sin inconveniente con miras, eso sí, para analizar las políticas sobre una amplia gama de cuestiones de políticas de CTI. En otras palabras, los datos son de libre acceso, pero deben seguirse los principios FAIR (localizable, accesible, interoperable y reutilizable). En su definición de la OCDE

**La ciencia de redes**: se define como "el conocimiento organizado de las redes basado en su estudio utilizando el método científico". En este contexto, la visualización de las redes ofrece una forma complementaria al análisis estadístico para descubrir, extraer y clasificar nuevos patrones en la estructura y los datos de la red**[[3]](#footnote-3)**. Los desarrollos actuales incluyen el análisis dinámico de redes (DNA) y la importación de spigots (correos electrónicos, Twitter, Facebook, …). Definido por El Consejo Nacional de Investigación

**Iniciativa de Política**: medio para alcanzar determinado objetivo concreto de política pública, a través del cual se promueve la realización de acciones de regulación, coordinación, financiación, provisión directa o indirecta de bienes y servicios, actividades o beneficios en áreas, sectores o ejes transversales que se consideran estratégicos para los países, entidades de financiamiento, organizaciones privadas y públicas, clientes y consumidores, comunidades, empresas y otros usuarios (OCDE, 2012; CEPAL, 2014; UNESCO, 2016; OCDE, 2019).

**Instrumento de Política**: mecanismo de operación de la iniciativa. Es el medio que permite definir las condiciones y requisitos a través de los cuales se realiza la selección de propuestas, programas, proyectos, beneficiarios o población objetivo de la intervención pública, con el fin de acceder a los beneficios, incentivos, apoyos financieros, acompañamientos o asesorías derivadas de la programación de la oferta institucional en coherencia con el objetivo y característica de la iniciativa de política a ejecutar.

**Algoritmo Yifan Hu[[4]](#footnote-4):** Algoritmo que genera buena calidad en gráficos (redes). Combina una fuerza dirigida con una técnica de engrosamiento de gráficos (multinivel) para mostrar la complejidad. Las fuerzas de repulsión en un nodo de un grupo de nodos distantes se aproximan por un cálculo de Barnes-Hut, que los trata como un super nodo.

**Modularidad**. Se refiere al agrupamiento de nodos que existe en una red (grafico); además, ayuda a determinar si existen sub-grupos de los grupos que se forman. Esos sub-grupos determinan las dinámicas de todo el grupo. Es un indicador importante para identificar el grado de conexiones entre los sub-grupos, o el grado de aislamiento de los mismos.

Tabla 1. Relación de número de iniciativas por tema.



Tabla 2, Canadá







Tabla 3 Chile





Tabla 4. Colombia





Tabla 5 Japón.





Tabla 6 Corea







Tabla 7 México



Tabla 8 Reino Unido









Tabla 9 España







Tabla 10. Estados Unidos











1. EC-OCDE (2021), STIP Compass: Base de datos internacional sobre políticas de ciencia, tecnología e innovación (STIP), edición 30 de noviembre de 2021, https://stip.oecd.org [↑](#footnote-ref-1)
2. Bastian M., Heymann S., Jacomy M. (2009). ***Gephi: un software de código abierto para explorar y manipular redes*** [↑](#footnote-ref-2)
3. Tomado de https://gephi.org/users/publications/ [↑](#footnote-ref-3)
4. Hu, Y. F. (2005). “Efficient and high quality force-directed graph drawing”. En: *The Mathematica Journa*l, 10 (37-71). [↑](#footnote-ref-4)