

**UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA**

**FACULTAD DE ECONOMÍA Y RELACIONES INTERNACIONALES**

**PROGRAMA DE DOCTORADO EN ESTUDIOS DEL DESARROLLO GLOBAL**



**TESIS:**

**TRANSFERENCIA TECNOLÓGICA DEL SOFTWARE EN LA  
INNOVACIÓN ABIERTA**

**PARA OBTENER EL GRADO DE:**

**DOCTOR EN ESTUDIOS DEL DESARROLLO GLOBAL**

**PRESENTA:**

**OSCAR YUNUE QUIROZ GAXIOLA**

**DIRECTOR DE TESIS:**

**DR. SANTOS LÓPEZ LEYVA**

Tijuana, Baja California, Octubre del 2017

## **DEDICATORIA**

A mis familiares, amigos y maestros.

## **AGRADECIEMINTOS**

En el presente trabajo de tesis agradezco a la Facultad de Economía y Relaciones Internacionales por darme la oportunidad de formar parte del programa de Estudios del Desarrollo Global.

De igual manera a mi director de tesis y mentor, Dr. Santos López Leyva por su nobleza como profesional, por sus consejos, y su ayuda en el camino de la investigación.

También a la Dra. Ana Bárbara Mungaray Moctezuma por haberme apoyado en los momentos difíciles, y a la Dra. Jocelyne Rabelo Ramírez por su visión crítica en muchos aspectos de la vida profesional.

Finalmente, agradezco profundamente al Consejo Nacional de Ciencia y Tecnología por su apoyo, sin el cual este proyecto no hubiese podido ser concretado.

Por todo lo que me han brindado, muchas gracias.

**ÍNDICE GENERAL**

INTRODUCCIÓN.....14

**CAPÍTULO I CONSTRUCCIÓN DEL OBJETO DE ESTUDIO..... 17**

1.1. Planteamiento del problema..... 18

1.2. Formulación del problema de investigación ..... 20

    1.2.1. Interrogante central .....20

    1.2.2. Sistematización del problema de investigación .....20

    1.2.3. Preguntas de investigación.....21

1.3. Justificación del tema..... 21

1.4. Objetivo General..... 22

    1.4.1. Objetivos específicos .....22

1.5. Hipótesis ..... 23

    1.5.1. Hipótesis general.....23

    1.5.2. Hipótesis particulares.....23

    1.5.3. Relación entre la hipótesis general y las particulares .....24

1.6. Alcances..... 24

1.7. Limitantes ..... 25

**CAPÍTULO II FUNDAMENTOS TEÓRICOS DE LA TRANSFERENCIA  
TECNOLÓGICA DEL SOFTWARE EN LA INNOVACIÓN ABIERTA..... 27**

2.1. Introducción al proceso de actualización tecnológica ..... 29

2.2. Antecedentes de la tecnología digital ..... 31

2.3. El software y la industria de los sistemas de información..... 33

    2.3.1. Licencias de software.....36

2.4. La transferencia tecnológica en la industria del software..... 37

2.4.1. Problemas con los periodos de actualización.....	40
2.4.2. Problemas adaptativos de la TSS.....	41
2.5. La innovación abierta.....	42
2.6. Ecosistemas de innovación abierta .....	43
2.6.1. Elementos esenciales en el entorno de innovación abierta .....	44
2.7. Transferencia tecnológica del software en la innovación abierta .....	45
2.7.1. Propiedad intelectual.....	48
2.7.2. Gestión del conocimiento.....	51
2.7.3. Comercialización .....	53
2.7.4. Descripción de los actores involucrados en el ecosistema digital .....	55

### **CAPÍTULO III PERSPECTIVAS METODOLÓGICAS DE LA INVESTIGACIÓN**

.....	<b>61</b>
3.1. Configuración metodológica para el objeto de estudio.....	63
3.2. Diseño metodológico .....	64
3.3. Características de la población.....	65
3.3.1. Criterios de inclusión .....	66
3.4. Construcción de la base de datos .....	67
3.4.1. Extracción de datos .....	67
3.4.2. Función de mapeo .....	68
3.4.3. Función de reducción .....	70
3.5. Análisis estadístico.....	71
3.5.1. Herramientas utilizadas en el análisis .....	72
3.6. Transferencia tecnológica .....	73
3.6.1. Estadísticos descriptivos generales .....	73

3.6.2. Actores predominantes.....	74
3.6.3. Análisis de flujos externos de innovación frente a transacciones locales.....	74
3.6.4. Análisis de variabilidad a los actores.....	75
3.7. Propiedad intelectual.....	76
3.8. Gestión del conocimiento .....	77
3.9. Comercialización .....	79
<b>CAPÍTULO IV RESULTADOS Y DISCUSIÓN.....</b>	<b>81</b>
4.1. Resultados del análisis a la transferencia tecnológica .....	83
4.1.1. Estadísticos descriptivos de muestra.....	83
4.1.2. Estadísticos descriptivos por sector .....	84
4.1.3. Resultados de la hipótesis .....	86
4.1.4. Resultados de la hipótesis por sector .....	87
4.1.5. Costos de gestión .....	88
4.2. Análisis de variación entre sectores.....	89
4.2.1. Detección de valores atípicos.....	89
4.2.2. Transformación a escala de rango de 0 a 1 .....	90
4.2.3. Resumen de la variabilidad en los niveles .....	92
4.2.4. Prueba HSD de Tukey.....	93
4.2.5. Análisis de varianza por pares .....	93
4.3. Resultados del análisis al entorno de la innovación abierta.....	94
4.3.1. Propiedad intelectual.....	95
4.3.2. Gestión del conocimiento.....	96
4.3.3. Comercialización .....	97
4.3.4. Procedimiento de inmersión al mercado.....	98
4.3.5. Procedimiento de posicionamiento .....	100

<b>CAPÍTULO V CONCLUSIONES Y RECOMENDACIONES.....</b>	<b>101</b>
5.1. La transferencia tecnológica del software en la innovación abierta .....	103
5.2. Propiedad intelectual.....	104
5.3. Gestión del conocimiento .....	105
5.4. Comercialización .....	106
5.5. La transferencia tecnológica en el entorno de innovación abierta.....	106
5.6. Hallazgos.....	108
5.7. Recomendaciones .....	109
5.8. Limitantes .....	110
5.9. Futuras líneas de investigación .....	111
REFERENCIAS.....	114
ANEXOS.....	121

**ÍNDICE DE TABLAS**

Tabla 2.1 Eventos relevantes en la historia de la tecnología digital. Con base a (Hua, 2004) ..... 32

Tabla 2.2 Aspectos del software como bien intangible. .... 35

Tabla 2.3 Modelos de transferencia tecnológica..... 39

Tabla 2.4 Descripción de las licencias libres más utilizadas ..... 50

Tabla 3.1 Criterios metodológicos para los elementos de la innovación abierta..... 64

Tabla 3.2 Herramientas y técnicas principales utilizadas en el análisis ..... 73

Tabla 4.1 Descriptivos de los flujos de innovación ..... 84

Tabla 4.2 Resultados de la hipótesis ..... 86

Tabla 4.3 Resultados del análisis de variación general..... 92

Tabla 4.4 Resultados de variabilidad intragrupal ..... 92

Tabla 4.5 Matriz del umbral de HSD..... 90

Tabla 4.6 Informe de las diferencias ordenadas por pares..... 94

Tabla 4.7 Informe de letras de unión ..... 94

Tabla 4.8 Resultados del ajuste logístico de propiedad intelectual (Licencia) en función de  $H_0=1$   
 ..... 95

## ÍNDICE DE FIGURAS

Figura 1.1 El balance entre la transferencia de innovación local y externa asociado a los riesgos para el autor de un proyecto de innovación abierta .....	19
Figura 1.2 Relación entre la hipótesis general y las particulares .....	24
Figura 2.1 Contenido del marco teórico .....	28
Figura 2.2 Evolución del proceso de industrialización. Fuente (Ahumada-Tello et al., 2011). ...	29
Figura 2.3 Línea de tiempo de la tecnología digital su impacto en la sociedad .....	31
Figura 2.4 Pirámide DIKW. Adaptado de (Rowley, 2007, p. 167). .....	34
Figura 2.5 El modelo de innovación abierta habilita el acceso a recursos e ideas externas a la empresa en sus proyectos de investigación y desarrollo.....	43
Figura 2.6 100k de conexiones de internet capturadas en 2013, Fuente: <a href="http://www.olivergroth.dk/site/#5">http://www.olivergroth.dk/site/#5</a> .....	46
Figura 2.7 Comunidades de desarrollo de software.....	47
Figura 2.8 En un proyecto de innovación abierta los actores involucrados cuentan con la posibilidad de adquirir valor e incrementar el acumulado por el colectivo.....	49
Figura 2.9 Proyectos en GitHub, por tipos de licencias abierta.....	51
Figura 2.10 Captura de pantalla de la estructura de los registros de actividad de un proyecto abierto. Fuente propia. Capturado de la herramienta GitKraken .....	52
Figura 2.11 Modelo de Cuádruple Hélice.....	55
Figura 2.12 Influencia de Github en las publicaciones científicas en Scopus.....	60
Figura 3.1 Planteamiento de la transferencia tecnológica en el entorno de la innovación abierta .....	63
Figura 3.2 El diagrama de flujo ilustra la secuencia de pasos utilizados en la metodología .....	65

Figura 3.3 Crecimiento la población de proyectos abiertos en Github.com..... 66

Figura 3.4 Expresiones regulares para rastrar los registros de actividad de los actores (su e-mail).  
..... 68

Figura 3.5 Diagrama de flujo para la función de mapeo de datos. Los registros se rastran por dominio de email y las transacciones por la palabra clave “merge”. Fuente: elaboración propia. .... 69

Figura 3.6 Captura de pantalla del diccionario de datos: {"@dominio"=>sumatoria} ..... 70

Figura 3.7 Procedimiento para la función de reducción, calcula todas las variables y las clasifica por sector. Fuente propia..... 71

Figura 3.8 Entrada y salida del ajuste lineal ..... 75

Figura 3.9 Entrada y salida del ajuste logístico ..... 75

Figura 3.10 La figura muestra las relaciones utilizadas para modelar la estructura los flujos de trabajo y las transacciones, se puede apreciar que las transacciones son el producto de dos flujos [id]..... 78

Figura 3.11 La captura de pantalla contiene el código fuente que se utilizó para replicar la estructura de la red de colaboración..... 79

Figura 4.1 Distribución de valores en los flujos de innovación de la muestra ..... 83

Figura 4.2 Información de los flujos de innovación por sector ..... 85

Figura 4.3 Resultados de la hipótesis por sector..... 87

Figura 4.4 Ajuste lineal de las transacciones locales en función del trabajo externo por sector .. 88

Figura 4.5 Valores atípicos por número de fila. Técnica: Distancias Jackknife..... 90

Figura 4.6 Densidades de los sectores después de aplicar la estandarización ..... 90

Figura 4.7 Variabilidad del cociente de colaboración externa frente a sector. Los 4 círculos a la derecha representan los límites comparados de variabilidad entre los sectores. .... 91

Figura 4.8 La captura de pantalla de Neo4j ilustra los primeros registros en cadena de valor del proyecto (Microsoft/vscode) ..... 97

Figura 4.9 El diagrama entidad relación (RAD) ilustra las interacciones entre los roles a lo largo del proceso de comercialización en la plataforma de la población (Github). .... 99

Figura 5.1 Umbral de decisión para el peso de la hipótesis alternativa frente a sector ..... 104

Figura 5.2 Red de colaboración. .... 105

Figura 5.3 Conceptualización del objeto de estudio en función de los resultados. .... 107

## ABREVIATURAS, SIGLAS Y ACRÓNIMOS

3C	Colaboración = Comunicación + Coordinación + Cooperación.
4H	Cuádruple Hélice.
GC	Gestión del conocimiento.
IA	Innovación abierta.
I+D	Investigación y Desarrollo.
I+D+I	Investigación, Desarrollo, e Innovación.
ISO	Organización Mundial de Normalización.
TT	Transferencia Tecnológica.
TSS	Transferencia Tecnológica del software.
OCDE	Organización para la Cooperación y el Desarrollo Económico.
OMPI	Organización Mundial de Propiedad Intelectual.
PI	Propiedad Intelectual.

## **RESUMEN**

La innovación abierta ha comenzado a popularizarse en el mercado global del software como un modelo de producción inclusivo; esencialmente, este esquema establece que varios actores pueden participar en la misma cadena de valor. Para las organizaciones, el paradigma abierto se presenta como la posibilidad de integrar flujos externos de innovación a su cadena local de producción; y así, agilizar los procesos de investigación y desarrollo (I+D). Sin embargo, el modelo de innovación abierta es un tema reciente y poco explorado, y esta situación se traduce en repercusiones operativas. En cuestiones del entorno, el desarrollo basado en la cooperación es algo novedoso para un mercado global que se caracteriza por ser competitivo. Al mismo tiempo, la gestión de herramientas, recursos, y conocimientos entre los actores locales y externos se ejecuta sobre condiciones vinculadas a la corta historia de la producción de bienes compartidos. Estos eventos han despertado el interés por entender los riesgos y beneficios que esconde el modelo de innovación abierta, así como el proceso de gestión tecnológica que se desarrolla en ese entorno.

En particular, este trabajo de tesis se enfoca en estudiar la cadena de producción en proyectos de innovación abierta para explicar el proceso de gestión tecnológica. Para este propósito, se analizaron los flujos locales y externos en los proyectos; también, se identificaron similitudes y contrastes entre los actores involucrados; y finalmente, se articularon los hallazgos de los pasos anteriores con los elementos del entorno de la innovación abierta. De esta manera, se explican las características generales del proceso de transferencia tecnológica en el entorno de la innovación abierta.

## **PALABRAS CLAVE**

Transferencia tecnológica, Innovación abierta, Métricas de innovación, Software.

## INTRODUCCIÓN

En las últimas décadas, el desarrollo tecnológico se ha propagado rápidamente en los productos y servicios de software, la popularización de estos bienes digitales ha causado un impacto importante en el funcionamiento de las organizaciones y en la sociedad. Sin embargo, para la industria del software esto se traduce en dificultades para adaptar estrategias de actualización tecnológica que satisfagan a un extenso rango de aplicación en periodos de actualización tecnológica cada vez más cortos.

La innovación abierta, por su parte, emerge como un paradigma de construcción de valor alternativo que permite a las organizaciones contrarrestar las dificultades de adaptabilidad y hacer frente a los cortos periodos de actualización a través de la integración de actores externos a su cadena de valor, desde socios comerciales, instituciones, miembros de la comunidad social, o cualquier otro tipo de actor; sin importar su afiliación.

Aunque el modelo de desarrollo abierto provee la posibilidad de acceder a flujos externos de innovación para fortalecer los procesos de I+D, también demanda un marco legal inclusivo; un requisito necesario para conceder el acceso de los actores externos a la cadena de valor, en otras palabras, esto significa conceder los derechos del producto. Por ese lado, la innovación abierta a menudo es referida como una espada de dos filos.

Por consiguiente, una de las cuestiones centrales con la adopción del modelo de innovación abierta es que no se sabe hasta qué punto es recomendable para una organización renunciar a la exclusividad de su producto para acceder los beneficios mencionados.

Por otra parte, el paradigma abierto introduce un escenario en donde la transferencia tecnológica involucra a varios actores dirigidos en la misma dirección, que además, puede desarrollarse concurrentemente.

Teniendo en cuenta que aún no es claro el procedimiento de transferencia tecnológica que se desarrolla entre los actores locales y los externos, resulta necesario describir el comportamiento de los proyectos actuales considerando los riesgos que enmarca el paradigma de innovación abierta para las organizaciones.

El presente trabajo refleja el comportamiento de la gestión de tecnología entre los actores locales y externos y su relación con el entorno de innovación abierta. Proporciona métricas sobre algunos de los proyectos de innovación abierta más populares y puede ayudar a obtener una comprensión global sobre el proceso de actualización tecnológica que involucra el paradigma abierto y, consecuentemente, ayuda a difuminar los riesgos asociados en su implementación.

En particular, este trabajo de tesis se enfoca en estudiar a la cadena de producción en proyectos de innovación abierta para explicar el proceso de gestión tecnológica. El contenido de este documento se divide en cinco capítulos.

El primer capítulo, establece las bases de la tesis, en éste, se especifican los criterios utilizados para estudiar la transferencia tecnológica entre los actores locales y los externos. Se describen los riesgos involucrados en la transferencia flujos, para así, establecer los objetivos y estructurar la hipótesis. En este apartado también se describen las necesidades y la pertinencia de estudiar el comportamiento que sigue la transferencia tecnológica del software en el entorno de la innovación abierta.

En el marco teórico se realiza una revisión sistemática de literatura para establecer los fundamentos teóricos del proceso de actualización tecnológica y se instruye sobre la importancia de las estrategias para hacer frente al cambio tecnológico. También se introduce a los modelos en los que se apoya esta investigación y se fija alcance conceptual del objeto de estudio.

El tercer capítulo aborda las perspectivas metodológicas de la investigación utilizadas para analizar los flujos de trabajo en la cadena de producción inclusiva y su relación con el entorno de la innovación abierta. En este apartado se utilizan los ejes conceptuales incluidos en marco teórico para colocar a la transferencia tecnológica en el entorno de la innovación abierta. En el diseño metodológico, se describen los criterios utilizados para seleccionar a los proyectos de innovación abierta, las herramientas para extraer los registros de actividad sobre los flujos de innovación, y las técnicas de análisis utilizadas para describir el entorno.

El cuarto capítulo, da a conocer las métricas obtenidas en el análisis a la transferencia tecnológica entre los actores locales y los externos involucrados en los proyectos de innovación abierta seleccionados, estos despliegan el comportamiento general de la muestra y el de cada uno de los actores. En lo que se refiere al entorno de innovación abierta, se expresa la importancia de aspectos fundamentales en el proceso de innovación, tales como la propiedad intelectual, la gestión del conocimiento, y la comercialización de los productos desarrollados sobre el entorno legal inclusivo.

En el quinto capítulo titulado “resultados y conclusiones” se explican los comportamientos particulares de la transferencia tecnológica y su relación con el entorno de la innovación abierta. Se proporcionan los hallazgos más relevantes y se sugieren algunas recomendaciones para la implementación del modelo de innovación abierta. Igualmente, se forjan las bases para futuras líneas de investigación.

Finalmente, se facilitan las referencias bibliográficas y los anexos que respaldaron el desarrollo de esta investigación, así como la base de datos y el código fuente utilizado para obtener los resultados.

**CAPÍTULO I**  
**CONSTRUCCIÓN DEL OBJETO DE ESTUDIO**

## 1.1. Planteamiento del problema

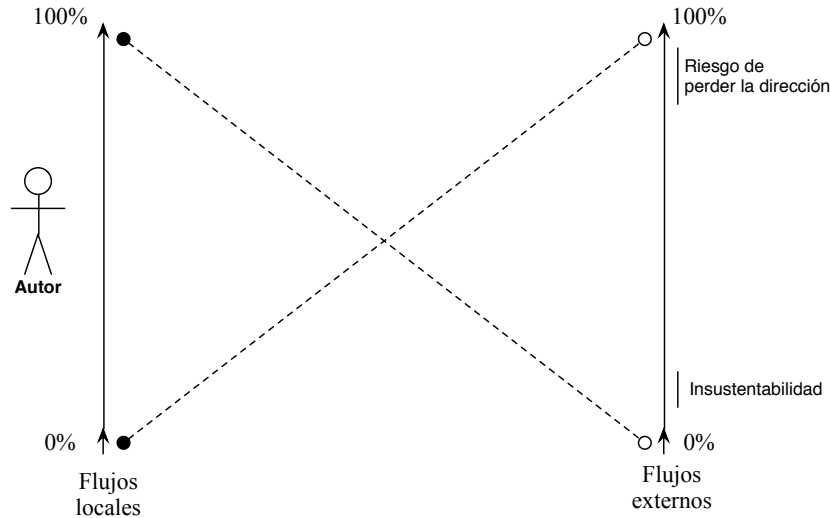
Por décadas la industria del software ha enfrentado problemas de adaptabilidad en las estrategias de actualización tecnológica. Desde inicios de los 90's, se comenzaron a formular estrategias que incorporan aspectos externos al proceso de actualización para hacer frente a las complicaciones adaptativas. No obstante, actualmente no existen modelos o estrategias genéricas para este tipo de procedimientos, ya que los efectos del entorno se siguen considerándose una cuestión problemática en la transferencia tecnológica del software (TTS).

En cuestiones del entorno, la innovación abierta (IA) se desarrolla en un marco legal inclusivo; desde la perspectiva de una organización, el paradigma de producción inclusivo se presenta como una estrategia de construcción de valor alternativa que le permite recolectar recursos externos de innovación; a pesar de los beneficios potenciales, actualmente no se sabe hasta qué punto es recomendable para una organización renunciar a la exclusividad de un producto para acceder a los beneficios de los flujos externos de innovación.

Según lo indican estudios previos, existen dos principales desafíos; por una parte, si los flujos externos son muy bajos, según Mortara, Napp, y Slacik (2009) menores al 10%, entonces significa que el proyecto no ha recaudado los suficientes recursos externos para considerarse rentable; por otra parte, si la organización comienza a verse superada en rendimiento por actores externos, entonces, corre el riesgo de perder la dirección del proyecto (Nyman & Lindman, 2013). Considerando estos riesgos, el balance entre los flujos locales y los externos es un aspecto que merece atención. Para ilustrar mejor esta problemática, en la figura 1.1 se ilustran los riesgos asociados con el balance de flujos.

Figura 1.1 El balance entre la transferencia de innovación local y externa asociado a los riesgos para el autor de un proyecto de innovación abierta

### Importancia del balance de flujos en un proyecto de innovación abierta



Fuente: elaboración propia con base en (Mortara et al., 2009; Nyman & Lindman, 2013).

Además de los riesgos asociados con balance de flujos, recientemente se han identificado otras problemáticas; como por ejemplo, Mortara et al. (2009) advierten que los flujos externos de innovación demandan esfuerzo por parte de las organizaciones para integrarlos a la cadena de valor, y a esto se suman las dificultades de adaptabilidad en las estrategias de actualización que caracterizan a la industria del software (Punter, Krikhaar, & Bril, 2006). Estas cuestiones problemáticas pueden reflejar incertidumbre en la planeación e implementación del modelo de innovación abierta.

En comparación con los modelos de negocios privados, la transferencia tecnológica que se lleva a cabo entre varios actores involucra a otro tipo de procedimientos. En primera instancia, se sabe sobre su alcance global, que el uso y consumo del producto pueden ocurrir al mismo tiempo; sin embargo, el procedimiento de gestión tecnológica que se desarrolla entre los actores locales y externos es un tema reciente y tiene pocos años estudiándose.

## **1.2. Formulación del problema de investigación**

Considerando las problemáticas que la industria del software enfrenta con la transferencia tecnológica y los riesgos asociados con la innovación abierta, en esta investigación se examina a los proyectos abiertos más populares en los cuatro sectores: privado, social, administración, e investigación, con el propósito de contrastar su comportamiento con los riesgos asociados a la implementación del modelo de innovación abierta.

El procedimiento para dar solución a las problemáticas planteadas anteriormente se presenta a continuación:

### **1.2.1. Interrogante central**

¿De qué manera se lleva a cabo la transferencia tecnológica del software en el entorno de la innovación abierta, y cuáles son las implicaciones para los actores involucrados en los proyectos de innovación abierta?

### **1.2.2. Sistematización del problema de investigación**

Tomando en cuenta que recaudación de flujos externos es el propósito central para el modelo de innovación abierta, y que el balance de flujos se asocia a riesgos de sustentabilidad, en esta investigación se ha colocado al balance de flujos como base para estudiar la transferencia tecnológica entre los actores locales y los externos involucrados en el entorno de la IA.

En lo que respecta a los problemas de adaptabilidad, el estudio al balance de flujos se aplica para cuatro diferentes actores: privado, social, administración, e investigación. Para ser más específicos, las variables son las siguientes:

Variable dependiente (VD). Trabajo acumulado del autor del proyecto de innovación abierta.

VARIABLES INDEPENDIENTES. Contribuciones acumuladas de los actores:

- Variable independiente 1 (Vi1). Contribuciones del sector privado.
- Variable independiente 2 (Vi2). Contribuciones del sector social.
- Variable independiente 3 (Vi3). Contribuciones del sector de administración.
- Variable independiente 4 (Vi4). Contribuciones del sector de investigación.

### **1.2.3. Preguntas de investigación**

Derivadas de la interrogante central se presentan las siguientes preguntas de investigación:

1. ¿Cuál es el comportamiento que sigue la transferencia tecnológica del software en el entorno de innovación abierta?
2. ¿Cuáles son los elementos fundamentales de la innovación abierta, y cómo repercuten en la transferencia de tecnología entre los flujos locales de innovación y los externos?
3. ¿Cómo se refleja el comportamiento de los sectores en los proyectos de innovación abierta?

### **1.3. Justificación del tema**

El interés científico de esta investigación consiste en proporcionar métricas sobre la transferencia de tecnología en proyectos de innovación abierta, y describir sus características y su relación con los elementos del entorno de innovación abierta.

Proporcionar información sobre el reciente paradigma de innovación abierta es importante porque puede ayudar a desarrollar nuevas estrategias para los procedimientos de construcción de valor en el contexto de los bienes compartidos. Similarmente, estudiar la gestión de tecnología y su relación con el entorno de innovación abierta puede ayudar a obtener una comprensión global

sobre el proceso de actualización tecnológica que involucra el paradigma abierto, y consecuentemente, difuminar los riesgos asociados en su implementación.

Actualmente, en los proyectos abiertos se proporciona la información completa sobre actividades y transacciones de conocimiento entre los actores involucrados. El acceso a esta información habilita la oportunidad de abrir investigaciones sobre el comportamiento social en estos entornos de trabajo (Kalliamvakou et al., 2016). En este caso, la transferencia de tecnológica.

Considerando que la innovación a menudo es un concepto difícil de cuantificar, y que los proyectos de información abierta registran cada agregado sobre el producto, resulta pertinente utilizar esa información para estudiar la transferencia de tecnología que se desenvuelve en el entorno de la innovación abierta.

## **1.4. Objetivo General**

Estudiar a los proyectos más populares de innovación abierta para explicar el comportamiento que sigue la transferencia tecnológica del software en el entorno de la innovación abierta, y proporcionar métricas sobre la conducta de los actores involucrados.

### **1.4.1. Objetivos específicos**

- Analizar y describir la transferencia tecnológica entre los actores locales y los externos.
- Analizar y describir los elementos del entorno de la innovación abierta.
- Calcular el comportamiento de los actores involucrados en los proyectos de innovación abierta.
- Situar y explicar la transferencia tecnológica en el entorno de la innovación abierta.

## 1.5. Hipótesis

En esta investigación se utiliza el método deductivo para estudiar el comportamiento que sigue la transferencia tecnológica del software en el entorno de la innovación abierta, partiendo de la hipótesis general para llegar a una conclusión de tipo particular.

Asimismo, se maneja el criterio de “actor predominante” para representar el balance entre los flujos locales de innovación y los externos; específicamente, para identificar al actor que refleja un mayor rendimiento en la cadena de producción abierta. Considerando los riesgos que representa el balance de flujos para las organizaciones, se establece como supuesto que el autor de un proyecto abierto es el acreedor a ese rol. Concretamente, la hipótesis general es la siguiente:

### 1.5.1. Hipótesis general

$H_0$ . En el marco de la innovación abierta, las contribuciones del autor de un proyecto abierto son mayores que el total de las contribuciones proporcionadas por cada uno de los actores externos.

### 1.5.2. Hipótesis particulares

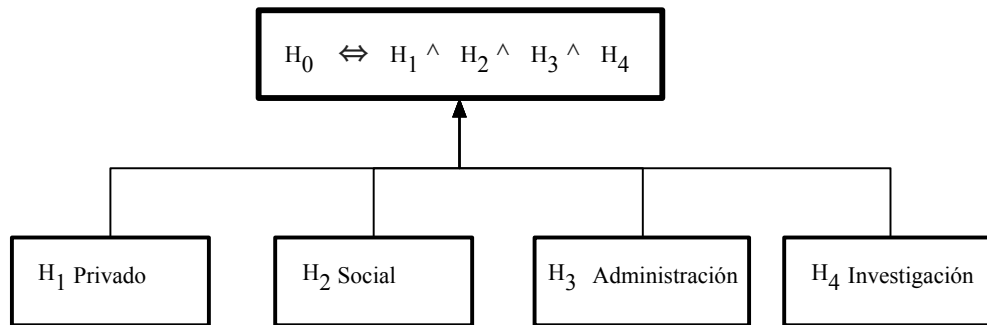
La hipótesis general se toma como base para abordar el proceso de transferencia tecnológica, y relacionarlo con el entorno de la innovación abierta. Por lo tanto, el supuesto anterior ( $H_0$ ) se asume para cada uno de los actores involucrados en el entorno de innovación abierta. La formulación del supuesto para cada actor se presenta a continuación:

- $H_1$ .  $VD \in [Vi1] \Rightarrow VD > (Vi1 - VD) | Vi2 | Vi3 | Vi4$
- $H_2$ .  $VD \in [Vi2] \Rightarrow VD > Vi1|(Vi2 - VD)| Vi3 | Vi4$
- $H_3$ .  $VD \in [Vi3] \Rightarrow VD > Vi1| Vi2 |(Vi3 - VD)| Vi4$
- $H_4$ .  $VD \in [Vi4] \Rightarrow VD > Vi1| Vi2 | Vi3 |(Vi4 - VD)$

### 1.5.3. Relación entre la hipótesis general y las particulares

A partir de las especificaciones anteriores, la hipótesis general se considera verdadera si y solo si supera el margen de error de  $\alpha = 0.05$ ; en caso de lo contrario, da lugar a una conclusión de tipo particular. Para ilustrar mejor, en la figura 1.2 se puede observar la estructura de la hipótesis general y su relación con las particulares:

*Figura 1.2 Relación entre la hipótesis general y las particulares*



*Fuente: elaboración propia.*

### 1.6. Alcances

- Debido a la naturaleza inclusiva de la innovación abierta y a que los proyectos de innovación abierta en la industria del software se desarrollan a lo largo de internet, el objeto de estudio se extiende a un alcance global.
- El alcance de la información proporcionada sobre los proyectos abiertos es exhaustiva y exclusiva, en otras palabras, no hay registros perdidos o replicados, y cada tipo de dato pertenece solo a una categoría. Esto se debe a que las proyectos estudiados proveen la información completa sobre los registros de actividad de cada uno de los actores y los cálculos realizados en este trabajo se implementan a partir de estos criterios.

- De acuerdo con Kalliamvakou et al. (2016) muchos de los proyectos de la población de proyectos abiertos de software son de uso personal, por lo tanto, esta investigación comprende únicamente a proyectos abiertos de organizaciones.
- La transferencia tecnológica es un concepto que aplica para diferentes contextos, pero considerando los objetivos planteados, en esta investigación el estudio a la gestión tecnológica se focaliza sobre el proceso de construcción de valor.

### 1.7. Limitantes

- En las métricas sobre el comportamiento de los actores no se proporciona la localización geográfica de los actores, aunque es posible calcularla como anteriormente lo realizaron (Yu, Yin, Wang, & Wang, 2014; Graham, De Sabbata, & Zook, 2015). Esto se debe a que la construcción de la base de datos es una de las partes más extensas de esta investigación, ya que incluye el cálculo del comportamiento de cada uno de los actores en los proyectos abiertos. Por consiguiente, no se ha considerado viable incluir el cómputo de la geolocalización.
- La población se limita únicamente a proyectos abiertos albergados en la plataforma de alojamiento *GitHub*<sup>1</sup>. Existen otras plataformas similares como *Atlassian*, *Bitbucket* y *Sourceforge*; sin embargo, se considera a *Github* la mejor alternativa porque recientemente se ha comenzado a utilizar para proyectos del sector de investigación y procedimientos regulatorios en el de administración, tal como indican (Longo & Kelley, 2015).

---

<sup>1</sup> <https://github.com>

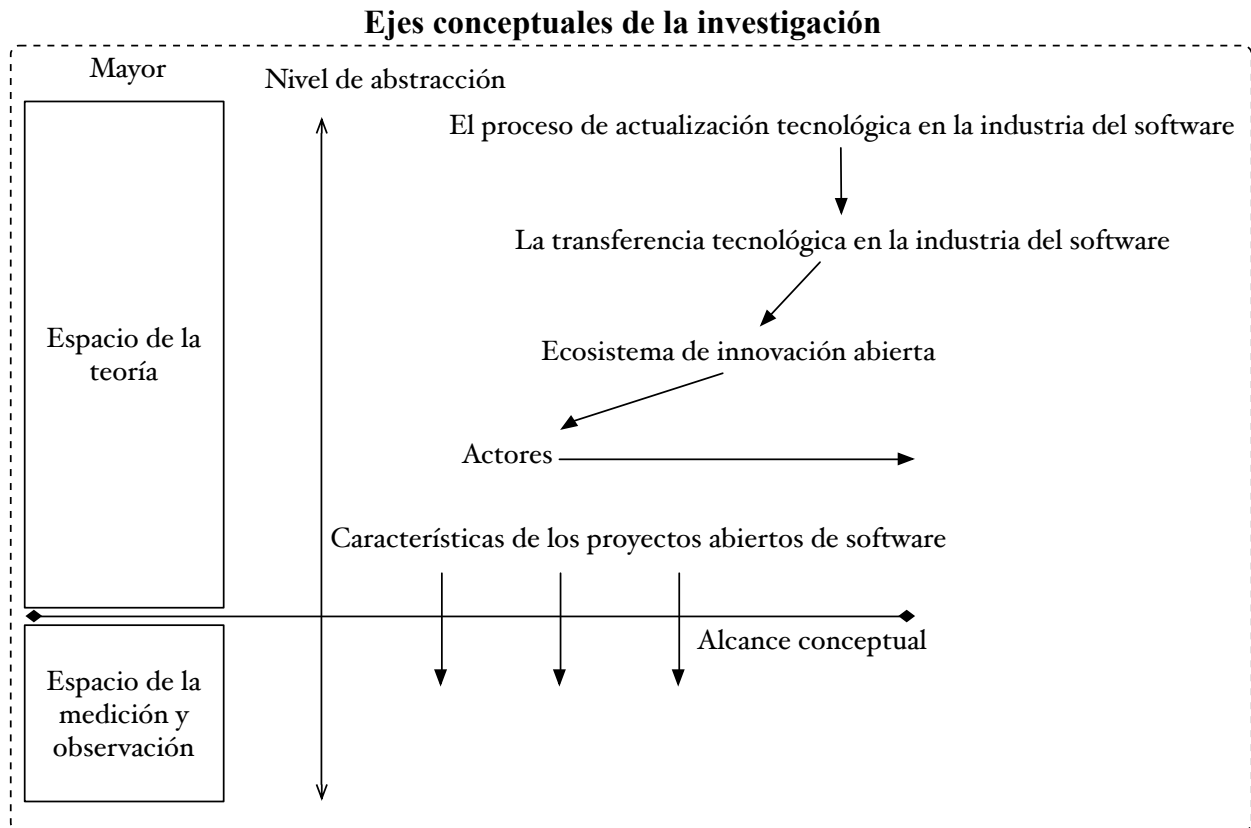
- Anteriormente, Biazzini et al. (2014) y Kalliamvakou et al. (2016) advirtieron que el cálculo de los registros de actividad exige altos costos de procesamiento y computación, por estos motivos la muestra se delimitó a 100 proyectos abiertos, 25 para cada sector.

## **CAPÍTULO II**

# **FUNDAMENTOS TEÓRICOS DE LA TRANSFERENCIA TECNOLÓGICA DEL SOFTWARE EN EL ENTORNO DE LA INNOVACIÓN ABIERTA**

En este apartado se proporcionan los fundamentos teóricos del desarrollo tecnológico, se explica la importancia de la transferencia tecnológica en la industria del software, y se introducen los orígenes de la innovación abierta. Dado que uno de los objetivos centrales de este trabajo es situar a la transferencia tecnológica en el entorno de la innovación abierta, también se establecen los elementos fundamentales de la innovación abierta como ejes conceptuales para formar el contexto del objeto de estudio. Esta última parte es muy importante porque introduce a los modelos en los que se apoya esta investigación y fija alcance conceptual del objeto de estudio. Para ilustrar mejor, en la figura 2.1 se ilustra el contenido de este capítulo.

*Figura 2.1 Contenido del marco teórico*



*Fuente: elaboración propia*

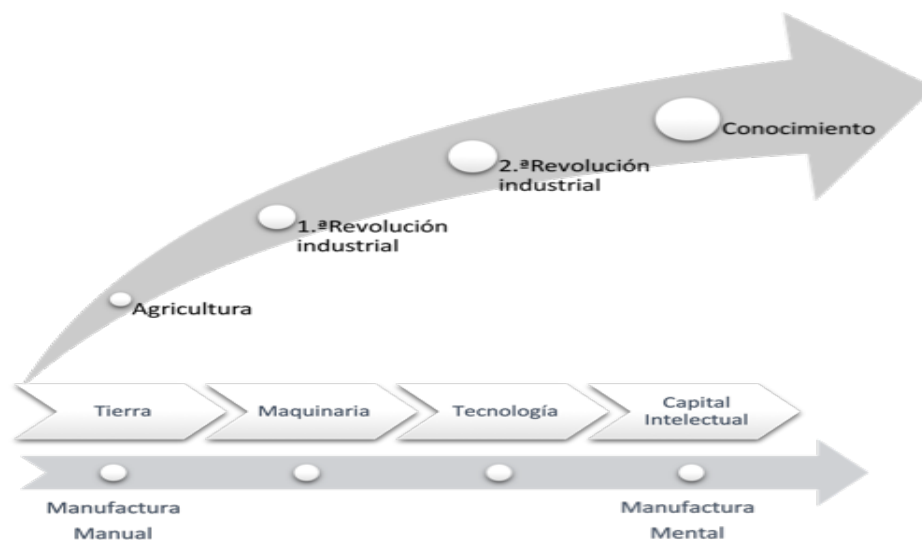
## 2.1. Introducción al proceso de actualización tecnológica

En cada época de la civilización, se puede distinguir un vínculo entre la sociedad y el entorno tecnológico del periodo. Asimismo, a medida que el progreso tecnológico sigue su camino, tanto la sociedad como las organizaciones enfrentan necesidades de adaptación una y otra vez. No obstante, aunque el progreso tecnológico se desenvuelve en torno a los nuevos descubrimientos, su trayectoria no es algo completamente impredecible.

La literatura científica generalmente coloca a la primera revolución industrial como el punto de partida de la sociedad contemporánea, ya que a partir de entonces la civilización comenzó a trascender de una economía basada en la agricultura a una basada en el desarrollo industrial. A esa trayectoria de desarrollo tecnológico generalmente se le refiere como el proceso de industrialización.

Como se puede apreciar en la figura (2.2), el desarrollo tecnológico refleja una tendencia hacia la optimización del conocimiento aplicado.

*Figura 2.2 Evolución del proceso de industrialización*



*Fuente: adaptado de (Ahumada-Tello et al., 2011, p. 562)*

En 1947 Schumpeter precisó una de las explicaciones más reconocidas sobre los efectos del proceso de industrialización utilizando el concepto de “Destrucción Creativa” para referirse al comportamiento evolutivo del desarrollo tecnológico, Schumpeter atribuye el desarrollo del proceso a un sistema de competencia que es necesariamente dinámico, constantemente interactuando y transformándose.

Asimismo, la destrucción creativa sucede cada vez que llega al mercado un nuevo producto con mejores características de funcionalidad, eficiencia, precio, etc. Sobre estas condiciones, el nuevo producto toma el lugar del actual, y así, se establece un proceso continuo de actualización. Sin embargo, de acuerdo con el economista, el proceso tiene sus efectos a largo plazo:

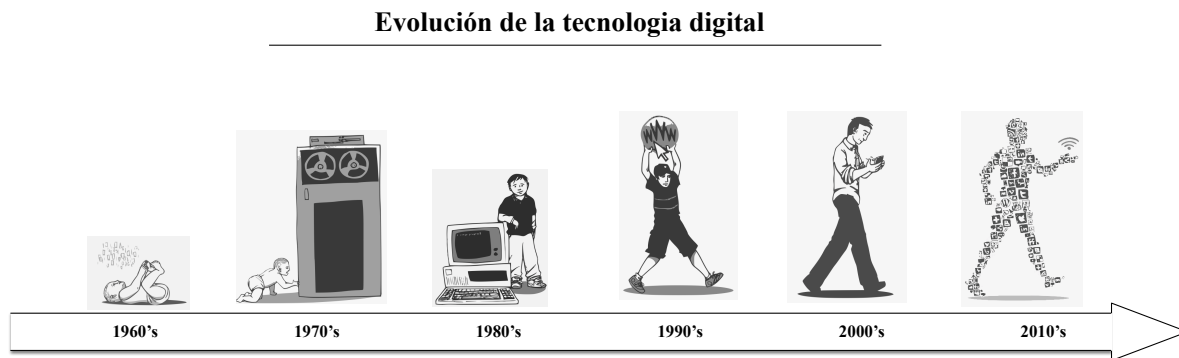
*“En primer lugar, puesto que se trata de un proceso en el que cada elemento tarda un tiempo considerable en revelar sus verdaderas características y efectos finales (...) debemos juzgar su desempeño con el tiempo, a medida que se despliega a través de décadas o siglos. Un sistema, cualquiera que sea el sistema económico o de otro tipo que en cada momento en el tiempo utilice plenamente sus posibilidades en el mejor de los casos, puede ser, a largo plazo, inferior a un sistema que no lo hace en ningún momento dado, hacerlo puede ser una condición para el nivel o la velocidad del rendimiento a largo plazo.”*(Schumpeter, 1947, p.83).

En el siglo XXI, el entorno tecnológico en el que la sociedad contemporánea se desenvuelve ha debilitado las barreras geográficas y temporales para facilitar la comunicación instantánea. Asimismo, la necesidad de adaptación para hacer frente a un proceso de actualización tecnológica cada vez más exigente ha impactado drásticamente el comportamiento dentro y fuera de las organizaciones (Sassen, 2000).

## 2.2. Antecedentes de la tecnología digital

Con la llegada de la tecnología digital a mediados del siglo XX, comenzaron a desarrollarse nuevas modalidades para explotar las posibilidades del formato digital. Como se ilustra en la figura (2.3), en pocos años se comenzaron a observar los efectos:

*Figura 2.3 Línea de tiempo de la tecnología digital su impacto en la sociedad*



*Fuente: Adaptado de (Wilms, 2012, Enero 10)*

En las últimas décadas, el progreso tecnológico ha avanzado radicalmente en el esfuerzo por extender los lazos comerciales e interconectar a las organizaciones con agentes externos (Mahajan, 2006). La tecnología digital, por su parte, introdujo nuevos paradigmas en la producción de bienes y prestación servicios digitales; por ejemplo, el trabajo a distancia, aparecen nuevos modelos de negocio, y se difunde la posibilidad de distribuir la cadena de producción en diferentes regiones (Friedman, 2007).

Para ilustrar mejor, la tabla (2.1) resume algunos de los eventos más relevantes en lo que se refiere al progreso tecnológico de la tecnología digital.

*Tabla 2.1 Eventos relevantes en la historia de la tecnología digital*

<i>Periodo</i>	<i>Eventos</i>
<i>1960's</i>	Surgen las primeras computadoras.
<i>1970's</i>	Se construyen los primeros laboratorios informáticos.
	La tecnología digital se comienza a utilizar en la industria para resolver cálculos complejos y como medio de almacenamiento.
<i>1980's</i>	Con los microcomponentes, el hardware reduce el tamaño y mejora en rendimiento.
	Salen al mercado las primeras computadoras personales.
	Aparecen las primeras licencias de software libre.
	La industria de software se consolida.
<i>1990's</i>	Aparece el primer navegador de internet.
	El internet comienza a propagarse.
	Se populariza el comercio electrónico.
<i>2010's</i>	Llega la tecnología móvil.
	Las compañías reaccionan al cambio tecnológico.
	Los medios sociales comienzan a popularizarse.
	Surgen nuevos modelo de negocio.

*Fuente: elaboración propia con base en (Hua, 2004)*

### 2.3. El software y la industria de los sistemas de información

En términos generales, el software es el componente lógico de los sistemas de información, y funciona a través de código con instrucciones que le indican a un dispositivo de cómputo específicamente lo que tiene que hacer.

De acuerdo con McCullagh (2000), el término “software” fue usado por primera vez por John Tukey en 1957 para referirse al intercambio de datos entre dispositivos electrónicos. Con el paso del tiempo, el concepto ha extendido su alcance a toda información procesada por los sistemas informáticos (Chambers, 2008), es decir, tanto aplicaciones como datos. Algunos autores como Aguilar (2016) amplifican el concepto del software más allá del código fuente incorporando *meta-datos*<sup>2</sup>, documentación, e incluso la información de los usuarios, en otras palabras, al paquete completo de elementos intangibles.

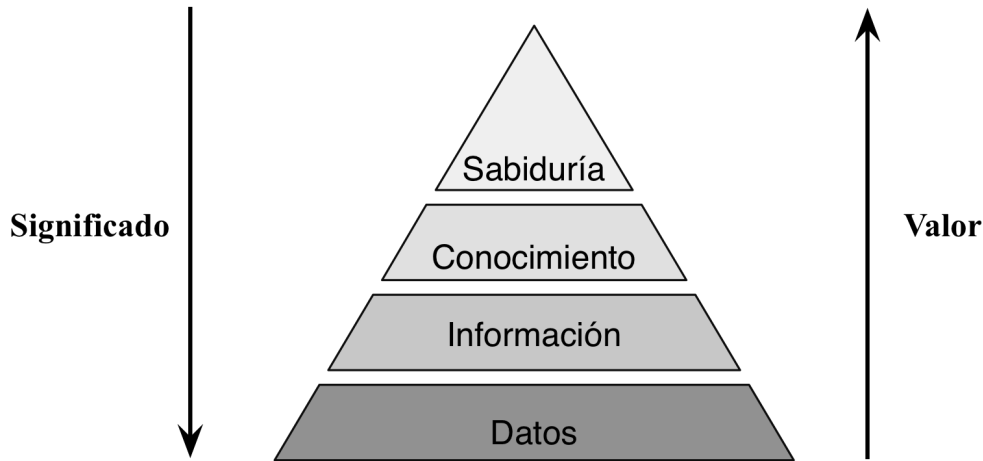
A simple vista el alcance del software puede parecer trivial, pero algunos autores indican que existe una línea que separa a cada uno de sus elementos. Numerosos teóricos en Informática y Sistemas de Información analizan su naturaleza de acuerdo a la jerarquía de datos, información, conocimiento y sabiduría (DIKW, por sus siglas en inglés).

La pirámide DIKW (ver figura 2.4) explica una secuencia jerárquica para el alcance del software. Concretamente, el orden de la jerarquía DIKW indica que los datos representan registros no estructurados, una vez estructurados se convierten en información, a través del procesamiento analítico se produce el conocimiento, hasta llegar al último escalón con la aplicación del conocimiento especializado, también conocido como sabiduría (Fricke, 2008).

---

<sup>2</sup> Un *metadato* proporciona datos descriptivos de los datos. Ej. Si en un archivo de texto los datos representan el contenido, los metadatos corresponderían a la fecha de creación del archivo, extensión, autor, tamaño del archivo, etc.

*Figura 2.4 Pirámide DIKW*



*Fuente: adaptado de (Rowley, 2007, p. 167).*

De acuerdo con Rowley (2007), existe un tipo de sistema de información asociado a cada elemento de la jerarquía DIKW, comenzando con los sistemas de transacciones de datos, sistemas de información para gestión, sistemas de soporte para toma de decisiones, alcanzando el nivel más complejo de automatización con los sistemas expertos (inteligencia de clientes, algoritmos de aprendizaje automático, inteligencia artificial, etc.).

En cuestiones comerciales, la Organización para la Cooperación y el Desarrollo Económico (OCDE) indicó en un reporte sobre la innovación en el sector del software que los sistemas de información y comunicación pertenecen al sector de bienes y servicios basados en conocimiento; una nueva modalidad en la construcción de valor que introdujo nuevas características comerciales, en palabras de la organización: “(...) como un producto digital, el software ofrece un costo extremadamente bajo de duplicación. Por lo tanto, los costos marginales de producción pueden ser un mal punto de partida para un ejercicio destinado a determinar el valor económico de una unidad adicional de software” (Douglas & Piotr, 2009, p. 42). En otras palabras, es complicado establecer el precio para un producto digital basándose en el costo de reproducción.

En lo que se refiere al aspecto intangible de los bienes digitales, (Andersson et al., 2011) señalan otras diferencias comerciales, desde ventajas como su facilidad en transporte y los bajos costos de replicación; así como las complicaciones en su regulación y la estimación de costos de producción. En la tabla 2.2 se describen algunos de esos aspectos.

*Tabla 2.2 Aspectos del software como bien intangible.*

<i>Aspecto</i>	<i>Descripción</i>
<i>Intangible</i>	Este aspecto habilita una serie de diferencias en las actividades de producción, distribución, y comercialización del bien; y consecuentemente, demanda nuevas formas de medir, establecer costos y estrategias de comercialización para este tipo de bienes.
<i>Transporte</i>	El formato digital genera bajos costos de transporte, ya que un bien digital puede distribuirse a través de dispositivos de almacenamiento físicos, pero también puede utilizar canales digitales como medio de transporte (Internet u otras redes de telecomunicaciones). Tampoco se requieren cuidados especiales para mantener la salud del producto.
<i>Acceso concurrente</i>	Los bienes digitales pueden ser utilizados concurrentemente por múltiples usuarios.
<i>Bajo costo de replicación</i>	Los costos de replicación pueden ser tan bajos que no se contabilicen en el proceso de producción. Una vez que un producto es completado, se pueden crear un sinnúmero de copias.
<i>Evolutivo</i>	El desarrollo de un producto digital puede continuar una vez que ha sido distribuido. Esto se puede llevar a cabo a través de actualizaciones periódicas o con la incorporación de nuevas funcionalidades.

*Fuente: elaboración propia con base en (Philipson, 2004; Douglas & Piotr, 2009)*

En el contexto global, la industria del software provocó un efecto al que Friedman (2007) describe como “competencia plana”, el concepto explica escenarios en donde varias empresas pueden competir remotamente por un mismo trabajo, ya sea por especialización o reducción de

costos. Actualmente, tanto las características de los bienes digitales como la competencia plana destacan la necesidad de adaptación para hacer frente a los aspectos mencionados.

### **2.3.1. Licencias de software**

Una licencia permite establecer las condiciones de utilización para los productos. Tal como (Katz, 2012, p. 43) indica, las licencias son instrumentos jurídicos que regulan la producción, el uso y la comercialización del software. En lo que se refiere a su alcance, (Lowry, 1995, p. 63) explica que los términos y condiciones en las licencias de software por lo general incluyen limitaciones de responsabilidad, y garantías de protección en caso de infracciones a los derechos propiedad intelectual (PI) sobre el bien.

Llegado a este punto, es importante señalar la diferencia entre una licencia y una patente. De acuerdo con la Organización Mundial de Propiedad Intelectual (OMPI) una patente aplica el derecho legal para “excluir” a otros actores de usar o comercializar un bien, mientras que una licencia emplea el derecho legal a “conceder” el uso, modificación, y distribución sobre el bien (WIPO, 2015). Aunque la perspectiva legal de las patentes y las licencias son inversas; pueden parecer similares en la ejecución cuando el producto se registra con fines privados.

En la actualidad, las patentes de software no se han podido consolidar. De acuerdo con (Graham, 2013), esto se debe a que la naturaleza intangible del software genera ambigüedad en las especificaciones del producto, y esa ambigüedad complica el proceso de evaluación y validación en el registro de la patente.

Debido a las complicaciones con las patentes, en la industria del software generalmente se utilizan licencias para enmendar las problemáticas con la propiedad intelectual (PI) del producto. Actualmente existen infinidad de licencias para registrar un producto de software, pero se pueden categorizar en dos tipos, propietario y abiertas.

*Licencias de propietario.* A menudo una licencia propietario está sujeta a un costo y generalmente se utiliza por las empresas (Lowry, 1995). Esencialmente, la licencia de propietario permite ejecutar el software en una sola máquina y realizar copias sólo como respaldo. Por ejemplo, la EULA (por sus siglas en inglés, *End-User License Agreement*), una de las más comunes, prohíbe al usuario compartir el software con cualquier otra persona, también, incluye una vigencia para los términos y condiciones establecidos sobre el producto (Gomulkiewicz, 2013).

*Licencias libres.* Las licencias libres conceden la utilización, modificación, y re-distribución de los productos de software (Lowry, 1995). Aunque en algunos casos pueden ser utilizados para fines privados, cada tipo de licencia establece condiciones específicas en las concesiones sobre el producto. La primera licencia libre llegó en 1984 (Stallman, 2007), y desde ese punto, se establecieron las primeras condiciones legales para conceder el uso, modificación, y re-distribución de un producto de software.

## **2.4. La transferencia tecnológica en la industria del software**

Generalmente, las nuevas tecnologías llegan acompañadas con la necesidad de evaluar su viabilidad. Tanto en las características de las nuevas herramientas como en los conocimientos para integrarlas adecuadamente; después de todo, es importante saber si vale la pena invertir esfuerzo en el proceso de actualización.

De acuerdo con Specer (2006), el desarrollo tecnológico se enfoca, principalmente, en optimizar actividades repetitivas, y eso da origen a que la planeación en los procesos de actualización tecnológica a menudo se concentren en la optimización de ese comportamiento recurrente.

Asimismo, las estrategias para hacer frente la actualización tecnológica pueden aplicar en varios contextos, por ejemplo:

La Organización Mundial de Propiedad Intelectual expresa que la transferencia tecnológica es un concepto que comprende a todas las estrategias involucradas en la gestión de nuevas tecnologías, y se ejecuta a través de mecanismos de vinculación entre los sectores económicos para evaluar e integrar los avances tecnológicos eficientemente (WIPO,2013).

Por otro lado, el Centro de Desarrollo Tecnológico del Instituto Tecnológico de Massachusetts (MIT, 2005) expone que la transferencia de tecnología consiste en la gestión de nuevos conocimientos y descubrimientos al público en general, y se puede llevar a cabo a través de publicaciones, intercambios en conferencias, y programas de vinculación con la industria.

En lo que respecta a la comercialización de los nuevos productos tecnológicos, Ameka y Dhewanto (2013) agregan que la transferencia tecnológica puede aplicar en dos sentidos. Como proveedor, la transferencia tecnológica (TT) consiste en llevar las nuevas tecnologías al mercado. Mientras que como cliente la necesidad se centra de transferir la innovación basada en la tecnología a los procedimientos actuales.

En los últimos treinta años, se han formulado diversos modelos para facilitar la transferencia tecnológica en diferentes contextos. Considerando las dificultades a las que se enfrentan los actores involucrados en el proceso de actualización, las estrategias para la actualización tecnológica han sido abordadas desde diferentes perspectivas. La tabla 2.3 describe algunos de los modelos más utilizados.

Tabla 2.3 Modelos de transferencia tecnológica

<b>Modelo de TT</b>	<b>Descripción</b>
<i>Lineal</i>	En este modelo, la transferencia tecnológica de una universidad a una empresa es entendida como un proceso conformado por una secuencia lineal de etapas. El modelo comienza con un descubrimiento científico y termina con un producto comercializado.
<i>Dinámico</i>	El modelo dinámico es similar al lineal pero incluye el análisis de los factores internos que pueden afectar el éxito de la transferencia de conocimiento científico-tecnológico. A pesar de ser una propuesta más integral con respecto al modelo lineal, éste no contempla el análisis de los factores externos al proceso de transferencia, entre ellos el papel del Estado.
<i>Catch Up</i>	También conocido como modelo de acercamiento tecnológico, consiste en la imitación de tecnología creada por un tercero, este esquema ha sido empleado exitosamente por China, Corea, y Japón. En pocas palabras, su funcionamiento se puede resumir en “copiar y mejorar”. Sin embargo, su complejidad radica en la habilidad del implementador para adaptar un proceso dinámico de actualización de conocimientos, así como en las actividades de mejora.
<i>Triple Hélice</i>	Este modelo resalta la triada Empresa-Universidad-Estado. Silicon Valley en EE.UU. y Cambridge en el Reino Unido son considerados los ejemplos más representativos del rol desempeñado por los tres actores. Básicamente, este modelo se enfoca en la concentración de pequeñas empresas de alta tecnología. Sus principales actividades son asesoría técnica y tecnológica con el fin de fundar una cultura de negocios e investigación.
<i>Cuádruple Hélice</i>	Superficialmente, se parece al modelo Triple Hélice, con dos diferencias distinguibles: 1) integra al sector social como una nueva hélice, y 2) el rol del Estado cambia por “Administración” para extender su alcance territorial. Sin embargo, Carayannis (2009) explica que más que un cambio de roles, el modelo Cuádruple Hélice (4H) se orienta en la promoción de un enfoque democrático a la innovación. Desde esa perspectiva, los actores son vistos como actores clave en la formulación de políticas y prácticas socialmente responsables.

Fuente: elaboración propia con base en (Audretsch, Lehmann, Link, & Starnecker, 2012; Moonan, 2013)

En lo que respecta a la industria del software, la transferencia de tecnológica poco a poco se ha consolidado como una cuestión problemática. Solinger, Engberts, y Ning (1994, p. 102) atribuyen la falta de teorías y metodologías para productos de software a que el proceso de desarrollo tecnológico suele ser aún menos formal que en otros contextos.

Adicionalmente, el entorno digital incorpora otras dificultades. Debido a que los productos y servicios de software a menudo operan a través de la Internet su comercialización puede operar en un contexto global (Nachira, Nicolai, Dini, Le-Louarn, & Rivera-León, 2007). Asimismo, Zelkowitz, Wallace, & Binkley (1998) advierten que la TTS puede exigir un enfoque personalizado de tal manera que esta pueda funcionar en diferentes lugares al mismo tiempo, ya que se puede ver afectada por aspectos culturales o la regulación doméstica de los Estados.

#### **2.4.1. Problemas con los periodos de actualización**

En la década de los 80's, las investigaciones de TTS a menudo se enfocaban en la estandarización de procedimientos; sin embargo, no se obtuvieron buenos resultados en la formalización de estrategias de actualización, y aunque persisten los patrones de diseño, técnicas y teorías; actualmente no existe tal cosa como un modelo genérico.

Para ilustrar mejor, uno de los estudios más robustos se desarrolló en 1985 por Redwine y Riddle, en su investigación analizaron los periodos de madurez y propagación del software. En aquel entonces, encontraron que se necesitaba alrededor de 15 a 20 años para que una tecnología emergente alcanzara un estado en el cual podía ser popularizada y difundida. En el peor de los escenarios, la transición tardó 23 años para pasar de un estado de formulación a un punto en el que la divulgación podría ser considerada; el mejor caso tomó 11, y el tiempo promedio fue de 17 años.

En años más recientes, la organización de análisis corporativo *NeoChange* (2012) implementó una encuesta a 300 empresas en el rubro de tecnologías de información (IT, por sus siglas en inglés) para evaluar el procedimiento de TTS. En contraste con los 17 años de Redwine y Riddle, los resultados de *NeoChange* explican un tiempo estimado de 2 años para el proceso de evaluación y adopción de nuevas tecnologías, además agregan que el 17% de los proyectos fallaron.

Los ejemplos anteriores reflejan el panorama sobre la concepción de los tiempos de adopción en la década de los 80 y como se concibe en esta década. Ya que un proceso de transición que en el mejor de los casos tome 11 años, difícilmente sería aplicable en la actualidad.

#### **2.4.2. Problemas adaptativos de la TSS**

De acuerdo con Zelkowitz (1998), las organizaciones a menudo utilizan estrategias empíricas o personalizadas para hacer frente a la actualización de herramientas tecnológicas, sin embargo, advierte que aunque las estrategias pueden servir como guía, no garantizan el éxito en su implementación, ya que los resultados dependen en gran medida del contexto y el entorno en el que se aplican.

Asimismo, no nada más el tiempo en los periodos de actualización ha cambiado desde la década de los 80, el enfoque por automatizar el proceso de actualización tecnológica ha disminuido drásticamente desde 1992 con el cierre del departamento de TTS en la Administración Nacional de la Aeronáutica y del Espacio, más conocida como NASA (por sus siglas en inglés), según indica (Zelkowitz, 1998). La situación no había cambiado para el 2010, en un reporte sobre la innovación en la industria del software de la OECD se indica que “no existe una medida clara y universal de software” (Douglas & Piotr, 2009, p. 8). A raíz de los problemas de adaptabilidad, las mayoría de las investigaciones comenzaron a contemplar factores externos, como por ejemplo, la vinculación inter-organizacional, conflictos culturales (Zelkowitz, Wallace, & Binkley, 1998), y la formulación de estrategias para hacer frente al cambio tecnológico adaptadas a cada tipo de organización (Schuh & Aghassi, 2013).

De acuerdo con investigadores de ese periodo, el problema de adaptabilidad era algo que ya se había contemplado anteriormente. Pau (1988) y Pfleeger (1998), señalan que las cuestiones adaptativas poco a poco fueron integrando factores externos en los modelos de TTS, desde la

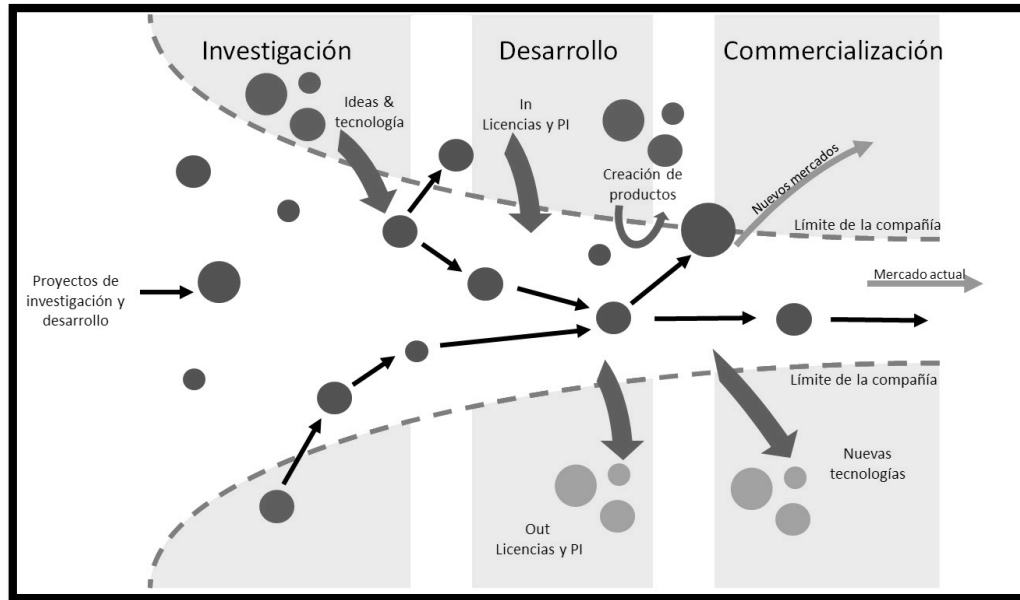
incorporación de los riesgos, aspectos de propiedad intelectual, características del producto y otros elementos pueden repercutir en las fases de iniciación, adopción y aplicación del nuevo producto (Ameka & Dhewanto, 2013).

## **2.5. La innovación abierta**

El concepto de innovación abierta fue definido por Chesbrough et. al. como "el uso de entradas externas de conocimiento para acelerar la innovación" (2006, p. 2). De acuerdo con Solleiro y Teran-Bustamante (2012, p. 80) el término conociste en "la práctica de mirar más allá de la propia organización. Esa visión puede incluir a los proveedores, clientes, instituciones, centros de investigación, e incluso a competidores".

Como se ilustra en la figura 2.5, el modelo de innovación abierta comprende la integración entre los recursos de la organización y flujos de innovación externa, dirigidos en una misma dirección. Los beneficios centrales que se plantean para las organizaciones son integrar flujos externos de innovación a su cadena de producción y expandir el alcance del producto a nuevos mercados (Chesbrough & Appleyard, 2007).

Figura 2.5 El modelo de innovación abierta habilita el acceso a recursos e ideas externas a la empresa en sus proyectos de investigación y desarrollo



Fuente: (Solleiro & Terán-Bustamante, 2012)

Existen varias razones que impulsan a su utilización; por una parte, las empresas están buscando nuevas maneras de optimizar sus capacidades para adaptarse a los periodos cortos de actualización tecnológica (Mortara et al., 2009); al mismo tiempo, el acceso a los flujos externos de innovación les permite reducir costos en sus procedimientos de investigación y desarrollo (I+D) para promover la innovación (WIPO, 2015).

## 2.6. Ecosistemas de innovación abierta

De acuerdo con (Carayanis, 2001), los ecosistemas de innovación son una estrategia que se comenzó a utilizar para hacer frente a los desafíos del cambio tecnológico, en particular, para mantener la competitividad frente a los ciclos de vida más cortos, y a los costos en I+D asociados a creación de nuevas generaciones de tecnología.

El concepto de los “ecosistemas de innovación” se comenzó a utilizar a principios de los 90’s y "al igual que su contraparte biológica, se mueve gradualmente de una colección aleatoria de elementos a una comunidad más estructurada" (Moore 1993, p.76).

En 2008, un reporte de la OCDE titulado “Innovación abierta en las redes globales” indica que “las empresas adoptan cada vez más ecosistemas de innovación entre países. En estas redes de innovación, las empresas se vinculan con personas, instituciones (universidades, agencias gubernamentales, etc.) y otras empresas de diferentes países para resolver problemas y aprovechar nuevas ideas” (OECD, 2008, p.9).

Ocho años después, la Organización Mundial de la Propiedad Intelectual (OMPI) reportó en el “Índice global de desarrollo 2015” que las políticas de innovación para el desarrollo poco a poco han ido cambiando de impulsar la I+D local a la creación de ecosistemas de innovación. También, añade la OMPI, que a través de la construcción de ecosistemas se busca generar las condiciones para que las empresas además de competir, compartan ideas y cooperen de manera abierta, todo esto, con el propósito de facilitar sus procesos de I+D (WIPO, 2015, p. 106).

El término se ha popularizado desde el 2005 junto con el modelo de innovación abierta, desde entonces numerosas empresas se han vuelto conscientes de la tecnología de externos para mantener su competitividad en el mercado global (Abulrub & Lee, 2012, p. 131)

### **2.6.1. Elementos esenciales en el entorno de innovación abierta**

Aunque pueden existir infinidad de elementos involucrados en los ecosistemas de innovación abierta, los expertos en el tema generalmente consideran elementos fundamentales de la innovación abierta a la propiedad intelectual (PI), la gestión del conocimiento (GC), la transferencia tecnológica (TT), y la comercialización.

Solleiro y Terán-Bustamante publicaron un extenso trabajo al respecto en 2012, de acuerdo con los investigadores existen varios elementos que inciden en el procedimiento de innovación, desde la conformación del paquete tecnológico hasta su comercialización.

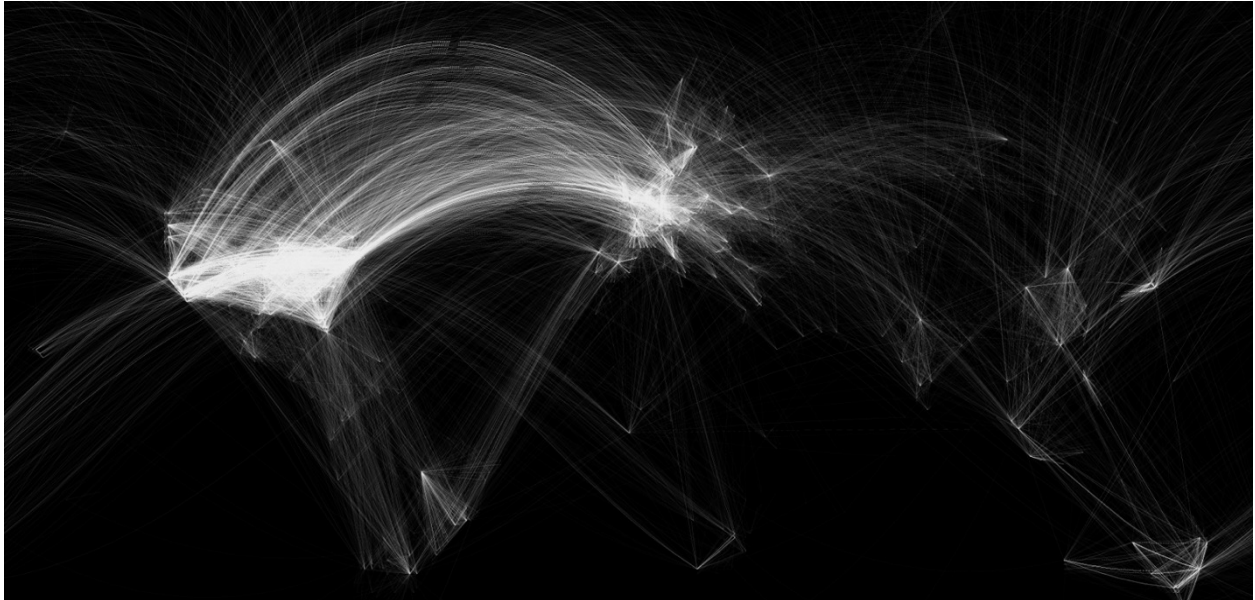
De acuerdo Chesbrough (2016), la innovación abierta opera sobre un entorno legal inclusivo, por ese lado, (1) la PI es considerada un elemento primordial para conceder el acceso a los flujos externos de innovación, este aspecto fue señalado por Lowry en 1995 incluso antes de que se formulara el concepto de la IA.

Adicionalmente, Solleiro y Terán-Bustamante consideran fundamentales a (2) la gestión del conocimiento que se lleva a cabo en la (3) transferencia de tecnología, ya que de estas dependen los procesos de creación, captación, transferencia y aplicación de conocimiento (2012, p. 61). Finalmente, (4) la comercialización, es señalada por Chesbrough (2016) y Vanhaverbeke (2009) una cuestión vital para los proyectos de IA porque potencializa la red de colaboradores externos, o en otras palabras, la recaudación de flujos externos de innovación.

## **2.7. Transferencia tecnológica del software en la innovación abierta**

En los últimos años, la difusión de la Internet habilitó una manera novedosa para apoyar la transferencia de tecnología a distancia. Actualmente, la red internacional es la anfitriona de espacios de trabajo remotos en donde las empresas y actores externos adoptan nuevos enfoques para sus estrategias y procesos de innovación. En la figura (2.6) se ilustra estado de interconectividad hasta el 2013.

*Figura 2.6 100k de conexiones de la internet capturadas en 2013*

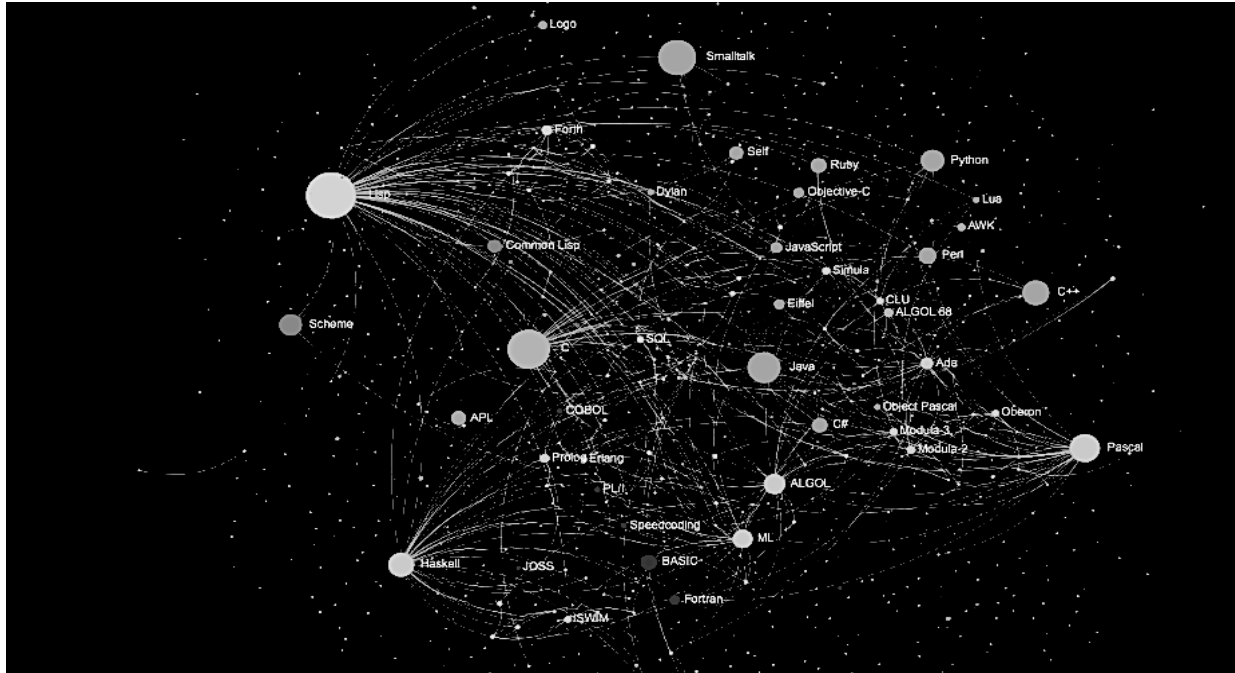


*Fuente: <http://www.olivergroth.dk/site/#5>*

Años después de su propagación en los 90's, comenzaron a desarrollarse nuevas herramientas para dar soporte a las necesidades del trabajo a distancia, ya que los grupos de desarrollo a menudo se distribuyen geográficamente y no pueden basar su coordinación en estrategias presenciales (Sassen, 2000).

Desde entonces, tanto empresas como la comunidad social utilizan herramientas de gestión vía web para rastrear y seguir las actividades de un gran número de externos, independientemente de la ubicación o afiliación (Dabbish, 2011, p. 1). Como se ilustra en la figura (2.7), actualmente existen diversas comunidades de desarrollo de software que a menudo también se relacionan entre ellas.

Figura 2.7 Comunidades de desarrollo de software



Fuente: Elaboración propia. Banco de datos: <https://developers.google.com/freebase/data>

En la última década, los proyectos de software libre comenzaron a desarrollarse en espacios de trabajo compartidos tales como *Bitbucket*, *Sourceforge*, y *Github* entre otros (Nyman & Lindman, 2013, p. 10). Estos repositorios proporcionan herramientas a los desarrolladores para comunicarse entre sí y gestionar su trabajo.

Cabe mencionar que las anteriores son empresas, y aunque sus fines son particulares, solo cargan un precio cuando el proyecto se registra con licencia privada. Debido a la ausencia de costo, (Kalliamvakou et al., 2016) señalan que los repositorios de trabajo abiertos se han convertido en una de las principales fuentes de acceso a productos de software libre.

Recientemente, se han desarrollado varios estudios en lo que se refiere al proceso de construcción de valor. (Thung, Bissyandé, Lo, & Jiang, 2016b) lo describen como un proceso de

colaboración concurrente, (Yu et al., 2014) se enfocan a explicar como los involucrados aprovechan el medio tecnológico para superar las restricciones ligadas al plano físico, y (Biazzini, Monperrus, & Baudry, 2014) exponen que la cadena de valor se construye en orden secuencial. Sin embargo, los repositorios de trabajo tienen poco menos de una década, y la información sobre el comportamiento que se desenvuelve en ese entorno tecnológico actualmente es limitada.

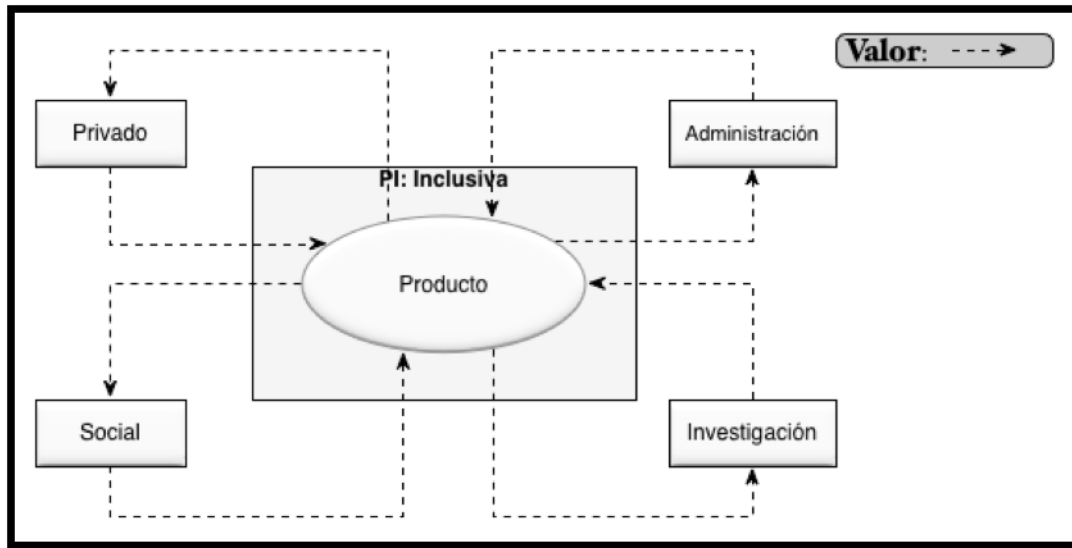
Los repositorios de trabajo compartido se comenzaron a utilizar en 2008, y aunque sus inicios se utilizaban principalmente para proyectos del sector privado y el social; recientemente, se han comenzado a utilizar en investigaciones, y procedimientos burocráticos (Longo & Kelley, 2015).

### **2.7.1. Propiedad intelectual**

La propiedad intelectual regularmente es considerada una de las cuestiones más problemáticas en la implementación de la innovación abierta (Mortara et al., 2009). De acuerdo con (Hall, 2014, p. 3), la controversia se centra en que el tipo de licencia puede impedir el desarrollo de innovaciones presentadas por fuentes externas.

En primera instancia, se sabe que cualquier licencia abierta «inclusiva» concede la libertad de uso, modificación, y re-distribución (West & Gallagher, 2006), aplicando una regla estricta, “cualquier re-distribución debe llevar los mismos términos de licencia” (de Laat, 2007, p. 166). De acuerdo con (Stallman, 2007), el propósito es regresar las modificaciones a la fuente y mantenerlas en un área común (ver figura 2.8).

Figura 2.8 En un proyecto de innovación abierta los actores involucrados cuentan con la posibilidad de adquirir valor e incrementar el acumulado por el colectivo



Fuente: Elaboración propia

Sin embargo, aunque las licencias libres parten de la misma base legal, cada una aplica concesiones sobre el producto de manera diferente. De acuerdo con (Nyman & Lindman, 2013), las problemáticas que surgen a raíz del tipo de licencia se deben a que cada configuración legal sirve a diferentes modelos de negocio, y las concesiones especificadas en una licencia asisten a ese fin. Asimismo, los términos en las concesiones inclusivas se han diversificado a para adaptarse a diferentes entornos de desarrollo tecnológico. Las algunas de las más utilizadas se describen en la tabla (2.4).

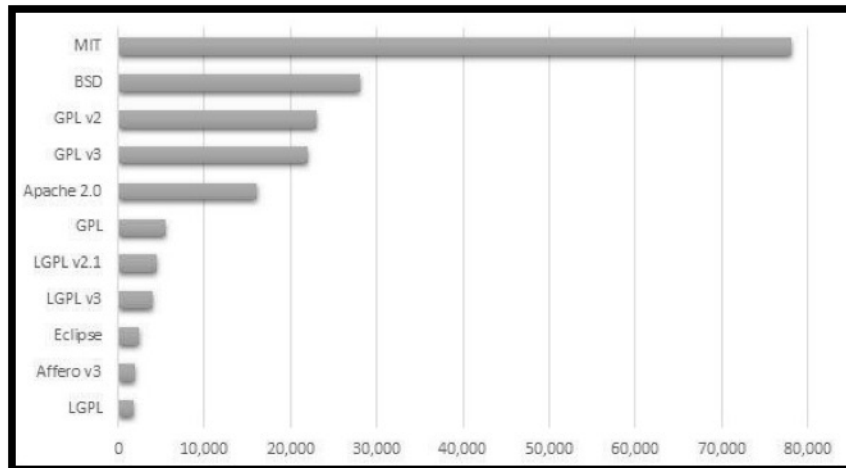
*Tabla 2.4 Descripción de las licencias libres más utilizadas*

Licencia	Descripción
Copyleft	Permite usar, modificar, y distribuir la licencia, pero exige compartir el código fuente.
Distribución de Software de Berkeley (BSD).	Esta licencia admite combinar el software propietario y liberarlo bajo una licencia propietaria, en este procedimiento se debe conservar el mismo tipo de licencia. Adicionalmente, se pueden agregar especificaciones para el uso del nombre del autor.
Publica GNU (GPL)	Admite usar, modificar, y distribuir el software de forma gratuita o por una tarifa. Asimismo, se tiene que distribuir el código fuente, y si el software se combina con otro software, ese software también tendrá que ser re distribuido con los lineamientos GPL
Publica Lesser (LGPL)	Es parecida a la GPL, pero además el software se considera distribuido; esencialmente, concede la utilización del software como servicio.
Publica Affero (AGPL)	Es parecida a la GPL, pero además el software se considera distribuido; esencialmente, concede la utilización del software como servicio.
Instituto de Tecnología de Massachusetts (MIT)	Actualmente es la más utilizada; permite usar, modificar y distribuir copias del software, incluyendo fines privados.
Apache	Permite usar, modificar y distribuir copias del software, adicionalmente, permite agregar una declaración personalizada de copyright a los cambios que realicen sobre el código fuente
Licencia Publica Mozilla (MPL)	Permite usar, modificar, distribuir y vender el software, siempre que también le proporcione al destinatario el código fuente. Con esta licencia, también se puede sub-licenciar el trabajo modificado, siempre y cuando no restrinja los derechos sobre el código fuente.

*Fuente: Elaboración propia con base en (Stallman, 2007; Gomulkiewicz, 2013)*

En la figura 2.9, se ilustra una comparativa de las licencias libres de acuerdo a su utilización en los proyectos abiertos registrados *Github*.

Figura 2.9 Proyectos en GitHub, por tipos de licencias abierta.



Fuente: [http://www.theregister.co.uk/2013/04/18/github\\_licensing\\_study](http://www.theregister.co.uk/2013/04/18/github_licensing_study)

## 2.7.2. Gestión del conocimiento

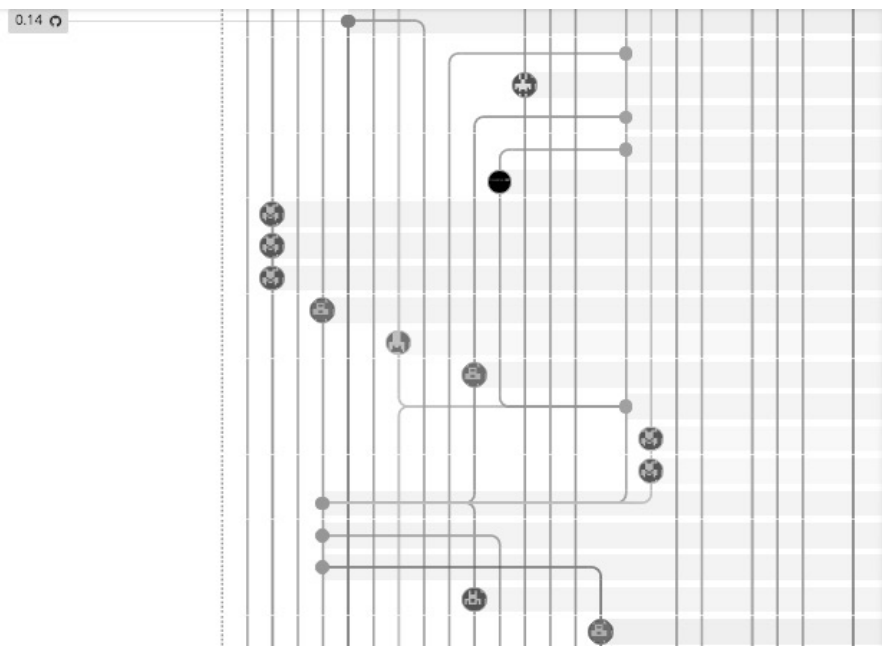
La coordinación entre equipos distribuidos se considera un reto debido a las diferencias de tiempo, geográficas, y culturales (Manteli, van Vliet, & Hooff, 2012). A pesar de eso, (Kalliamvakou, Damian, Blincoe, Singer, & German, 2015a) argumentan que la gestión del conocimiento se centra en mantener los esfuerzos interdependientes coordinados hacia un objetivo común, y a partir de ahí, mitigar los desafíos que los proyectos distribuidos enfrentan en la colaboración, tales como la coordinación y las rupturas de comunicación.

Varias investigaciones previas indican que la gestión del conocimiento que se lleva a cabo en proceso de colaboración involucra al menos tres elementos: comunicación, coordinación y cooperación (3C) (Manteli et al., 2012; Kalliamvakou, Damian, Blincoe, Singer, & German, 2015; Kock, 2007). En 2012 Manteli et al. investigaron el funcionamiento de esos elementos en los proyectos abiertos, y explica que en los tres casos se llevan a cabo a través de enlaces virtuales, los investigadores agregan que estas actividades se ejecutan a través de procedimientos ya integrados en el servicio de alojamiento.

Tres años después Kalliamvakou y sus colaboradores estudiaron los mismos criterios desde otra preceptiva, según los expertos no son solo los mecanismos sino también las implicaciones de la parte social, de acuerdo su investigación, por ejemplo, la transparencia en la triangulación de las 3C promueve el intercambio de conocimiento entre los colaboradores (Kalliamvakou et al., 2015 ).

La figura (2.10) ilustra la estructura de un repositorio de trabajo compartido. A simple vista se puede observar una estructura ordenada que fluye en paralelo y se ramifica horizontalmente, pero en el contexto de la transferencia tecnológica representa gráficamente a los flujos de trabajo.

*Figura 2.10 Captura de pantalla de la estructura de registros de actividad de un proyecto abierto*



*Fuente propia. Capturado de la herramienta GitKraken*

Recientemente, varios investigadores han estudiado la estructura de los flujos de trabajo distribuidos utilizando técnicas de análisis de red social. Por ejemplo, (Biazzini et al., 2014) y (Thung et al., 2016) describen al conjunto de flujos locales y externos como una red de colaboración, ahí, cada nodo representa un agregado sobre el bien, mientras que las intersecciones entre líneas representan la integración de flujos externos. En otro estudio similar, (Shmueli,

Altshuler, & Pentland, 2014, p. 335) explican que la forma en la que se estructuran los registros de trabajo «en paralelo» permite soportar una cantidad ilimitada de participantes, mientras que la localización geográfica de los actores es compensada por la dinámica temporal del medio tecnológico.

A grandes rasgos, los estudios anteriores describen la estructura de los flujos que integran a un producto compartido. Sin embargo, otras cuestiones no se han podido vislumbrar, tales como los procedimientos, mecanismos de tolerancia a fallos, y las implicaciones de la parte social que sustentan la colaboración a escala.

### **2.7.3. Comercialización**

Actualmente existen diferentes interrogantes sobre el proceso de comercialización que se ejecuta en torno a las concesiones del uso, modificación, y re-distribución. Particularmente, por la ausencia del valor monetario en cada una de estas. En este aspecto, los investigadores suelen explicar los procedimientos de inmersión y posicionamiento involucrados en esta fase a partir de sus diferencias con la producción de bienes exclusivos.

Por ejemplo, (Rifkin, 2014, p. 43) explica que los efectos de las concesiones de las licencias libres se exteriorizan desde la etapa de producción al eliminarse el sobreprecio directo sobre la concepción del bien, después, cuando el producto sale a competir con otros bienes compartidos las diferencias se exhiben otra vez con la eliminación del aspecto monetario en las transacciones cliente-proveedor.

En el mismo sentido, Kalliamvakou y Vanhaverbeke expresan que hay cambios en la interacción entre clientes y proveedores, de acuerdo con (Vanhaverbeke, 2009) uno de los más distintivos sucede con la integración de los clientes a la cadena de producción, más tarde,

(Kalliamvakou, Damian, Blincoe, Singer, & German, 2015b) agregan que en los repositorios compartidos la producción y consumo puede ocurrir al mismo tiempo.

En lo que respecta al modelo de innovación abierta, la comercialización es señalada por Chesbrough como una de las cuestiones más importantes cuando propone el modelo en 2003, de acuerdo con el investigador de la Escuela de Negocios de UC Berkeley el propósito de la comercialización no radica en los beneficios directos sobre el bien sino en la recaudación de flujos externos de innovación. En el mismo contexto, (Vanhaverbeke, 2009) argumenta que a raíz de la eliminación de costos, el posicionamiento del producto pasa a ser una cuestión vital para el seguimiento del proyecto ya que “potencializa la red de colaboradores externos” (p. 154).

Debido a las concesiones legales del entorno abierto, hoy en día el posicionamiento es considerado un aspecto fundamental en la etapa de comercialización. Asimismo, el mecanismo de posicionamiento que se utiliza en las plataformas de colaboración ya ha sido investigado anteriormente.

Yu et al. y sus colegas revisaron este aspecto en 2014, en su investigación explican que el posicionamiento se automatiza a través de algoritmos que calculan las posición de los productos basándose en su popularidad, generalmente *PageRank* y *SimRank*, según indican (Easley y Kleinberg, 2010).

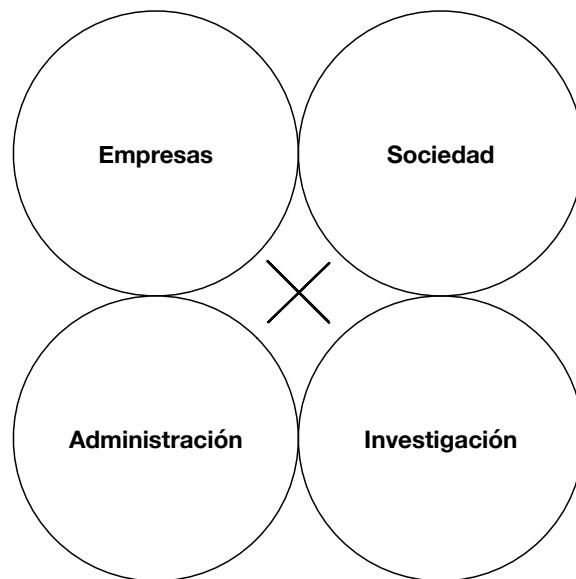
Como ya se ha mencionado, actualmente existen infinidad de interrogantes alrededor del proceso de construcción de valor en el entorno legal inclusivo. Adicionalmente, (Rifkin, 2014, p. 93) advierte que con la llegada del “hardware libre” lo que se sabe sobre la producción y consumo de los bienes compartidos tendrá que adaptarse a los productos físicos. Hasta ahora, el paradigma de desarrollo abierto sigue una tendencia de crecimiento rápida desde el 2008 y se sigue

diversificando a aplicaciones sin relación con el código fuente. Pero hasta este punto es difícil saber cual es el rumbo que tomara la producción de bienes compartidos en el futuro.

#### 2.7.4. Descripción de los actores involucrados en el ecosistema digital

Debido a que la TSS a menudo es considerada un “problema adaptativo”, en este apartado se proporciona un breve perfil teórico de los actores involucrados. La figura (2.11) personifica gráficamente el modelo 4H de Elías Carayannis. De acuerdo con el modelo, cada hélice representa a diferentes actores involucrados en los procesos de innovación. En lo que se refiere al entorno de innovación abierta, cada uno cuenta con sus características y sus roles específicos en el proceso de TTS.

*Figura 2.11 Modelo de Cuádruple Hélice*



*Adaptado de (Carayannis & Campbell, 2014)*

Por cuestiones prácticas, a lo largo de este documento se hace referencia a las 4 hélices como sectores: privado para las empresas, social, de administración y de investigación respectivamente.

### 2.7.4.1. Sector privado

Las actividades del sector privado se relacionan directamente con la producción de bienes y prestación de servicios (Leydesdorff, 2012). En la industria del software, las empresas comenzaron a utilizar el sistema de producción sobre bases legales inclusivas hace aproximadamente quince años, poco antes de que Chesbrough formulara el modelo de innovación abierta.

Desde entonces, se ha convertido en uno de los protagonistas en lo que se refiere a la producción de bienes compartidos, aunque existen programas para pequeñas y medianas empresas por lo general los proyectos abiertos pertenecen a corporaciones del rubro de las telecomunicaciones.

Algunas de las empresas más representativas en la implementación de proyectos abiertos en la actualidad son: Adobe<sup>3</sup> en servicios de diseño digital, Uber<sup>4</sup> en servicios de movilización, Amazon<sup>5</sup> (*AWS*) en plataformas de servicios web, Netflix<sup>6</sup> en la industria del entretenimiento, Microsoft<sup>7</sup> en el desarrollo de sistemas y servicios computacionales, y Android<sup>8</sup> en tecnología para dispositivos móviles.

### 2.7.4.2. Sector Social

Cuando Chesbrough acuñó el concepto de “innovación abierta” en 2003, explicó, desde su perspectiva, que las condiciones en la que se desarrollan los bienes abiertos puede ser vista como una manifestación del comportamiento comunitario, en donde los individuos se reúnen para

---

<sup>3</sup> <https://github.com/adobe>

<sup>4</sup> <https://github.com/uber>

<sup>5</sup> <https://github.com/aws>

<sup>6</sup> <https://github.com/Netflix>

<sup>7</sup> <https://github.com/Microsoft>

<sup>8</sup> <https://github.com/android>

atender intereses, gustos y preferencias. Pero en el caso de desarrollo de software, el comportamiento de la comunidad social comenzó a reflejarse en el mercado.

De acuerdo con Friedman (2007), los primeros avistamientos del sector social en la producción de software comenzaron a principios de los 80s, poco después de llegada de las computadoras personales.

A menudo el sector social es conceptualizado como un conjunto de organizaciones que no pertenecen al público ni al privado (Méndez et al., 1993). Sin embargo a partir de la llegada de la primera licencia libre en 1984, el sector social comenzó a desarrollar su propia identidad legal.

Actualmente, el sector social utiliza el marco legal inclusivo para desarrollar sistemas operativos móviles como Mozilla<sup>9</sup>, también se desarrollan los sistemas de *criptomonedas* Bitcoin<sup>10</sup> y Ethereum<sup>11</sup>, diversas herramientas de desarrollo de software como Node<sup>12</sup> y Rails<sup>13</sup>, además de los bienes digitales que ya se venían desarrollando desde antes, como es el caso del sistema operativo Linux<sup>14</sup> y los productos de la Fundación de Software Apache<sup>15</sup>.

### 2.7.4.3. Sector de administración

El sector de administración, como lo plantean Carayannis y Campbell (2014), se puede extender a un contexto internacional, tales como tratados internacionales, oficinas de transferencia tecnológica, o también, puede aplicar en un entorno doméstico a través de instituciones.

---

<sup>9</sup> <https://github.com/mozilla-mobile>

<sup>10</sup> <https://github.com/bitcoin>

<sup>11</sup> <https://github.com/ethereum/>

<sup>12</sup> <https://github.com/nodejs>

<sup>13</sup> <https://github.com/rails>

<sup>14</sup> <https://github.com/torvalds/linux>

<sup>15</sup> <https://github.com/apache>

De acuerdo con Arnkil, Jarvensivu, Koski, y Piirainen (2011) el objetivo central para esta división es entender y regular las actividades de los demás actores. En lo que respecta a los proyectos de innovación abierta, Longo y Kelley (2015) estudiaron recientemente este apartado, de acuerdo con su investigación el propósito de los proyectos en este sector consiste en involucrar a la sociedad en programas regulatorios.

Actualmente, los proyectos más representativos del sector de administración comprenden algunos casos federales como la Oficina de Impuestos de Australia<sup>16</sup>, el servicio digital gubernamental en Inglaterra<sup>17</sup>, peticiones ciudadanas en Washington<sup>18</sup>, el ayuntamiento de Barcelona<sup>19</sup>, además de programas de beneficencia internacionales como *Hospital Run*<sup>20</sup>.

Arnkil et al. (2011) señalan que el objetivo de la actividad de innovación es, ante todo, establecer los mecanismos regulatorios para que las organizaciones puedan funcionar adecuadamente, y en consecuencia, ofrecer mejores productos y servicios a los ciudadanos. En ese mismo sentido, los proyectos de innovación abierta formulados por el sector de administración suelen centrarse en el desarrollo de las organizaciones y los servicios públicos para propiciar una tendencia a la innovación en procedimientos de regulación (Carayannis & Campbell, 2011).

#### **2.7.4.4.Sector de investigación**

A este sector, se le atribuyen las mejores capacidades para generar conocimiento ya que opera sobre una base científica (Bendis & Craciunoiu, 2002). A través de la tecnología digital, se

---

<sup>16</sup> <https://github.com/atogov>

<sup>17</sup> <https://github.com/alphagov>

<sup>18</sup> <https://github.com/whitehouse>

<sup>19</sup> <https://github.com/AjuntamentdeBarcelona>

<sup>20</sup> <https://github.com/HospitalRun>

habilitaron nuevas posibilidades para impulsar el desarrollo de herramientas para dar soporte a la creación y difusión de conocimiento.

Asimismo, la relevancia y expectativas de la comunidad de investigación aumenta conforme la actualización de la tecnología digital se vuelve cada vez más rápida. De acuerdo con Zelkowitz (1998), la responsabilidad para generar conocimiento y reducir la incertidumbre sobre la naturaleza de la transferencia tecnológica en el mercado actual recaen sobre este sector.

Kurman (2013) indica que aunque existen ejemplos de universidades involucradas en la producción este sector no cuenta con la misma madurez en actividades productivas que las empresas. De acuerdo con Hallerstedte (2013), esto se debe a que la prioridad de las universidades y los centros de investigación se focalizan en generar y difundir conocimiento antes que la producción de bienes y servicios .

Actualmente, el sector de investigación utiliza los recursos externos de innovación principalmente para impulsar proyectos relacionados la creación y difusión de conocimiento. Por ejemplo, el Centro de Desarrollo Internacional de la Universidad de Harvard<sup>21</sup> está desarrollando herramientas y bancos de datos para estudiar al sistema financiero, también se encuentran programas de aprendizaje a distancia como EdX<sup>22</sup>, programas de agencias de investigación que buscan integrar a la comunidad social en el desarrollo de tecnología como es el caso de la NASA<sup>23</sup>, además de herramientas para investigaciones de energías renovables<sup>24</sup>, geo-procesamiento y

---

<sup>21</sup> <https://github.com/cid-harvard>

<sup>22</sup> <https://github.com/edx>

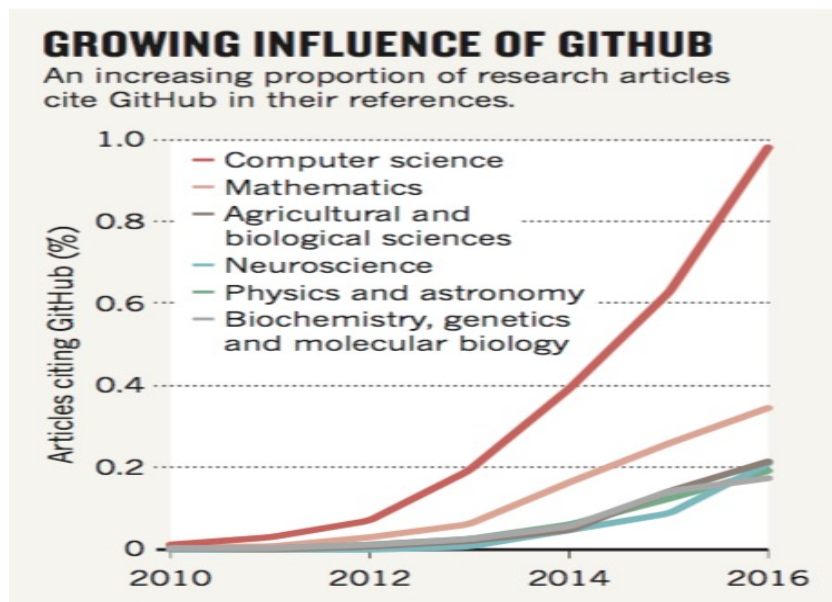
<sup>23</sup> <https://github.com/nasa>

<sup>24</sup> <https://github.com/NREL>

cálculo de datos climáticos, e infinidad de bancos de datos, como es el caso del Centro Nacional de Información Biotecnológica<sup>25</sup>.

Debido a que repositorios abiertos son cada vez más utilizados por investigadores para compartir, mantener y actualizar los conjuntos de datos científicos y el código, esos los recursos se han comenzado a utilizar también para la difusión, en la figura (2.12) se puede observar una tendencia ascendente desde el 2010 en la base de datos (BD )de citas *Scopus*.

Figura 2.12 Influencia de Github en las publicaciones científicas en Scopus



Fuente: (Perkel, 2016, p. 1)

<sup>25</sup> <https://github.com/ncbi>

**CAPÍTULO III**  
**PERSPECTIVAS METODOLÓGICAS DE LA**  
**INVESTIGACIÓN**

En este capítulo se describe el procedimiento metodológico utilizado para explicar la transferencia tecnológica del software en el marco de la innovación abierta. Para este fin, se utilizó un enfoque mixto, con una parte cuantitativa centrada en el análisis a los elementos del objeto de estudio; cabe señalar, que aunque se obtuvo la información completa sobre los flujos de innovación, el estudio es exploratorio, por lo tanto se recurrió a una parte cualitativa para establecer el contexto del entorno. Para ser más específicos, se utilizaron en orden de secuencia los ejes conceptuales planteados en la revisión de literatura para situar los elementos del entorno de la innovación abierta.

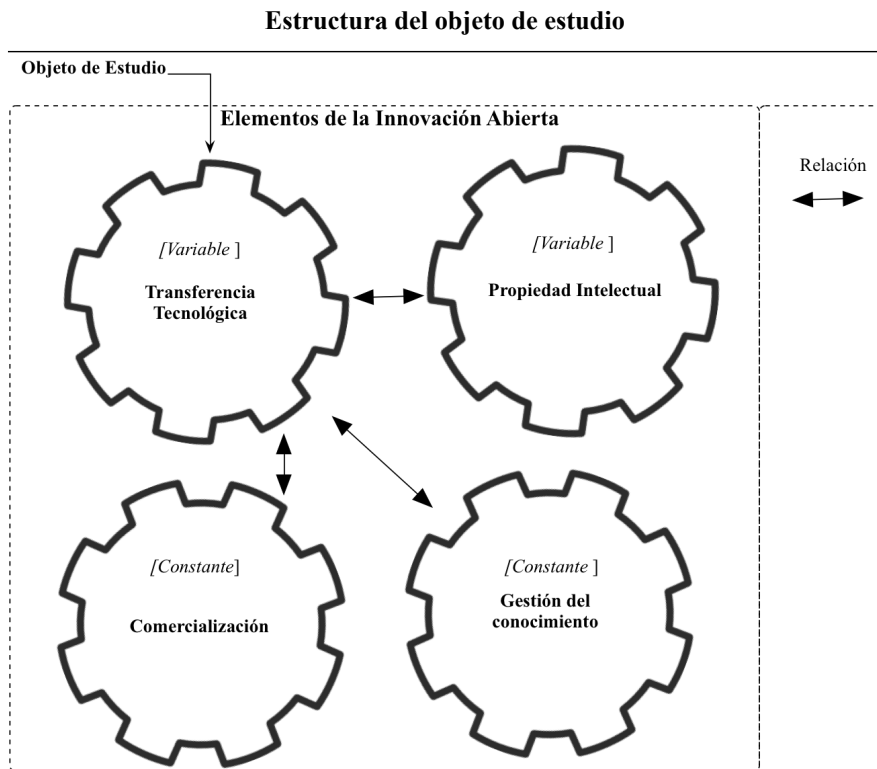
Inicialmente, fueron seleccionados 100 proyectos de innovación abierta en la industria del software para extraer sus registros de actividad y construir la base de datos; posteriormente, se analizaron los datos extraídos para describir el balance entre los actores locales y los externos, tales como estadísticos descriptivos, su relación con la presencia del actor predominante, los costos para gestionar recursos externos y la variabilidad en los cuatro sectores que componen la muestra.

En lo que se refiere a establecer la transferencia tecnológica en el marco de la innovación abierta, se incorporó un análisis complementario a los elementos considerados fundamentales en el proceso de innovación: propiedad intelectual, gestión del conocimiento y comercialización. Para ser más específicos, se relacionan los resultados de la transferencia tecnología con el tipo de licencia para explicar la propiedad intelectual; después, se analizó la estructura de la gestión de conocimiento, y el procedimiento de comercialización.

### 3.1. Configuración metodológica para el objeto de estudio

Asumiendo que el entorno de innovación abierta se compone de cuatro elementos fundamentales: transferencia tecnológica, propiedad intelectual, gestión del conocimiento y comercialización. Primero se analizó la transferencia tecnológica, y después se estudió la relación con los elementos restantes de la innovación abierta (ver figura 3.1). Todo esto, con el propósito de explicar la TTS en el entorno de innovación abierta.

Figura 3.1 Planteamiento de la transferencia tecnológica en el entorno de la innovación abierta



Fuente: elaboración propia.

En el capítulo anterior, se proporcionaron los fundamentos teóricos de cada uno de los elementos de la innovación abierta. La relación entre los flujos y el tipo de licencia (Hall, 2014), la estructura de los flujos de innovación, y los mecanismos de automatización para el

procedimiento de inmersión al mercado y el posicionamiento de los productos. En la tabla 3.1 se reúnen los criterios de análisis para la TTS y su relación con los elementos del entorno .

*Tabla 3.1 Criterios metodológicos para los elementos de la innovación abierta*

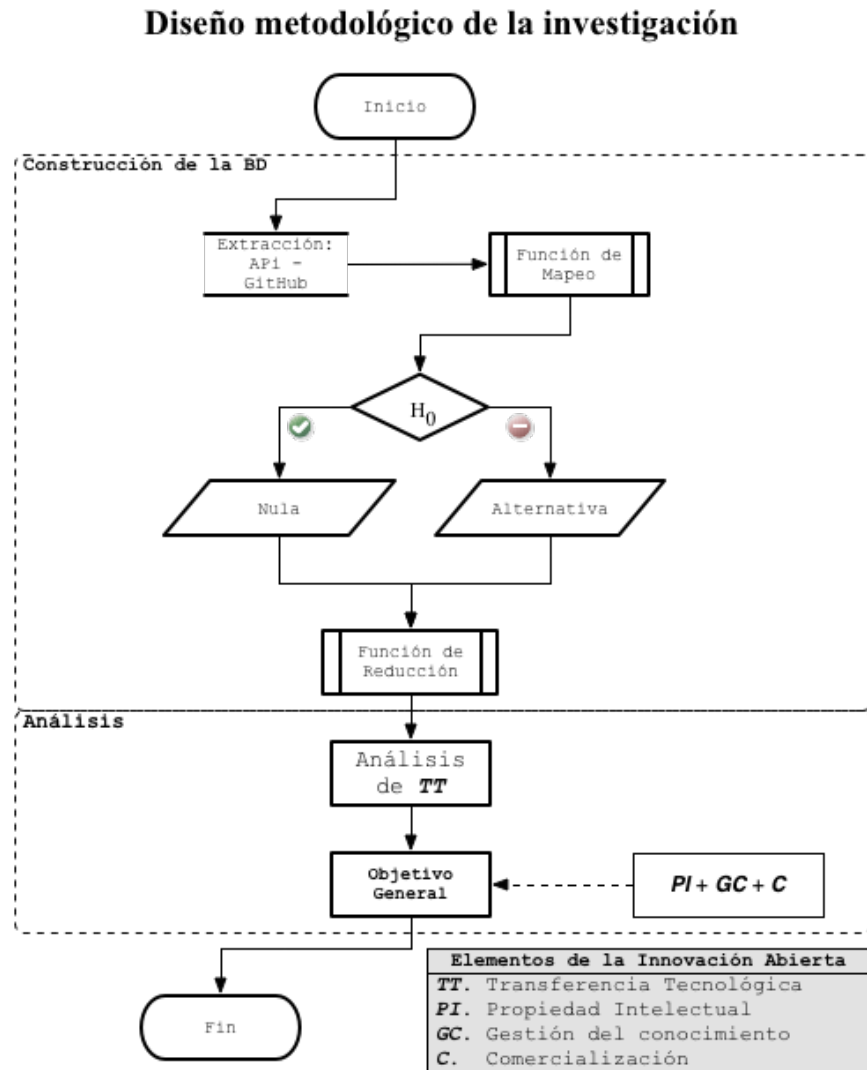
<b>CRITERIO</b>	<b>TIPO</b>	<b>TÉCNICA DE ANÁLISIS</b>	<b>HERRAMIENTA</b>
<i>Transferencia tecnológica</i>	Variable	ANOVA de un factor	R, JMP
		Regresión lineal	R, JMP
<i>Propiedad intelectual</i>	Variable	Regresión logística	R
<i>Gestión del conocimiento</i>	Constante	Análisis de red social	Neo4j
<i>Comercialización</i>	Constante	Diagrama de entidad relación (RAD)	Omnigraffle

Fuente: elaboración propia

### **3.2. Diseño metodológico**

En esta investigación, se estudiaron 25 de los proyectos más populares de innovación abierta en cada sector (100 en total). Todos seleccionados de la población de *Github*. Como se puede apreciar en el diagrama de flujo de la figura 3.2; primero, se extrajeron los registros de la bitácora de actividades; después, se implementó un análisis de texto para construir la base de datos; posteriormente, el análisis a la transferencia tecnológica; y por último, se agrega un análisis complementario para relacionar la TTS con los otros elementos del entorno de innovación abierta.

Figura 3.2 El diagrama de flujo ilustra la secuencia de pasos utilizados en la metodología



Fuente: elaboración propia.

### 3.3. Características de la población

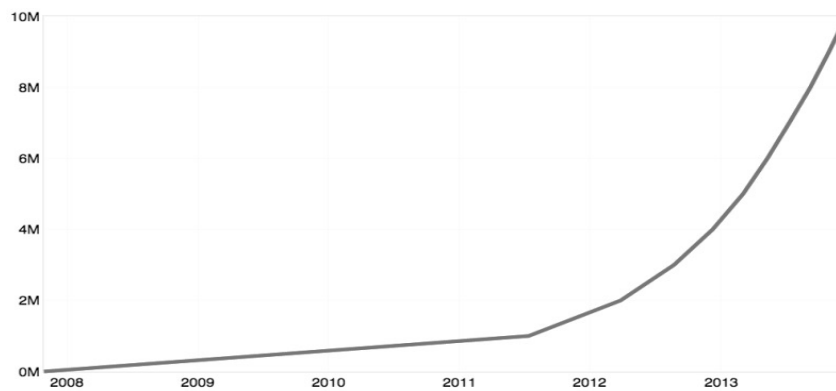
*GitHub*<sup>26</sup> es un servicio de alojamiento web que hospeda proyectos de colaboración electrónica, utiliza un mecanismo de versionamiento distribuido para permitir el acceso concurrente al

<sup>26</sup> <http://github.com>

producto (Longo, 2015). Al mismo tiempo, proporciona una interfaz web que automatiza el proceso de colaboración para grupos de trabajo remotos.

Kalliamvakou (2016) indica que desde su apertura en 2008 hasta el 2013 almacenaba poco más de 10 millones de repositorios (ver figura 3.3). Según informa *GitHub*, para agosto del 2017 la cifra incrementó a poco más de 19 millones de proyectos, y prestaba sus servicios a aproximadamente a 331,000 organizaciones.

*Figura 3.3 Crecimiento la población de proyectos abiertos en Github.com*



Ahí, la información de los proyectos abiertos es presentada a través de un sistema de archivos que registra altas, bajas, y modificaciones sobre los archivos que componen al producto. Al mismo tiempo, las actividades del grupo de trabajo son almacenadas en un registro “log” que, básicamente, funciona como una bitácora de actividades (West & Gallagher, 2006). De esta manera, se puede acceder al total de registros de cada colaborador.

### **3.3.1. Criterios de inclusión**

Aunque *GitHub* actualmente hospeda más 19 millones de proyectos abiertos tampoco significa que en todos se desarrolle la innovación abierta. Manteli et al. (2012) explican que muchos de esos proyectos son de uso personal o simplemente tienen poca actividad. Por lo tanto, al igual que

(Kalliamvakou et al., 2016) y (Biazzini et al., 2014), en este trabajo se optó por delimitar la población, los criterios de inclusión son los siguientes:

- Tipo de licencia: Libre.
- 25 proyectos por sector, 100 en total.
- Los proyectos se seleccionan del ranking de popularidad en *PageRank*<sup>27</sup>. Este criterio fue utilizado anteriormente por (Thung, Bissyande, Lo, & Lingxiao Jiang, 2016a).
- No se consideraron proyectos anteriores a surgimiento de *Github* «2008».
- 1 proyecto por organización.
- No se contemplan las “re-distribuciones”, de lo contrario los registros de trabajo realizados por el autor del proyecto previo entrarían en conflicto con los registros realizados por el autor actual.
- El muestreo se realizó el 24 de mayo del 2017, todos los cálculos corresponden desde la apertura de *Github* hasta esa fecha.

### 3.4. Construcción de la base de datos

#### 3.4.1. Extracción de datos

Para la extracción de datos, se utilizó la interfaz de programación de aplicaciones<sup>28</sup> (API) de *Github*, a través de ese módulo de transacciones, se extrajeron los registros de actividad de los proyectos de innovación abierta en formato texto.

Posteriormente, se utilizó la estrategia “mapear y reducir<sup>29</sup>” para la construcción de la base de datos. Básicamente, el mapear y reducir consiste en dos funciones: una de “mapear” para

---

<sup>27</sup> *PangRank* es un algoritmo de posicionamiento desarrollado por google.com

<sup>28</sup> <https://developer.github.com/v3/>

<sup>29</sup> Mapear y reducir transforma una lista de pares ordenados (clave, valor) en una matriz de valores, utilizada generalmente para tratar con datos no estructurados (Roebuck, 2011).

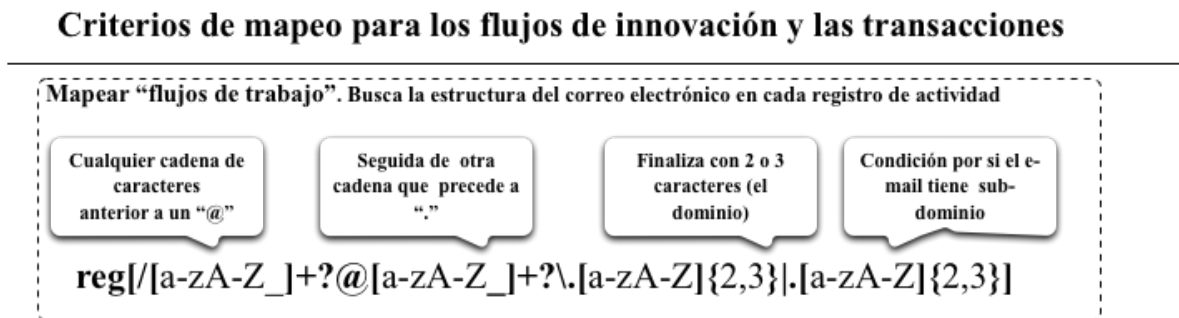
identificar y almacenar los datos y otra de “reducción” para presentarlos en columnas «variables», y filas «observaciones» (Roebuck, 2011).

Inicialmente, la bitácora de actividades proporciona los datos en formato texto. Este tipo de fuente de datos difiere en varios aspectos con una base de datos convencional, en un texto, los datos de interés están distribuidos en cadenas de caracteres separadas por espacios y saltos de línea. Por esta cuestión, se utilizó una función para rastrear los datos sobre los flujos de trabajo y generar un diccionario de datos<sup>30</sup> En este caso, el diccionario de datos es indispensable ya almacena los resultados del mapeo en un formato estructurado.

### 3.4.2. Función de mapeo

En el registro de actividades de los proyectos abiertos, el correo electrónico funciona como una firma electrónica que vincula a los registros de trabajo con sus autores (Thung et al., 2016). En la función de mapeo, se utilizaron como referencia los dominios del correo electrónico para identificar el total de registros de cada actor, en la figura (3.4) se ilustra el criterio de mapeo de datos, mientras que la figura (3.5) describe el procedimiento de esta función.

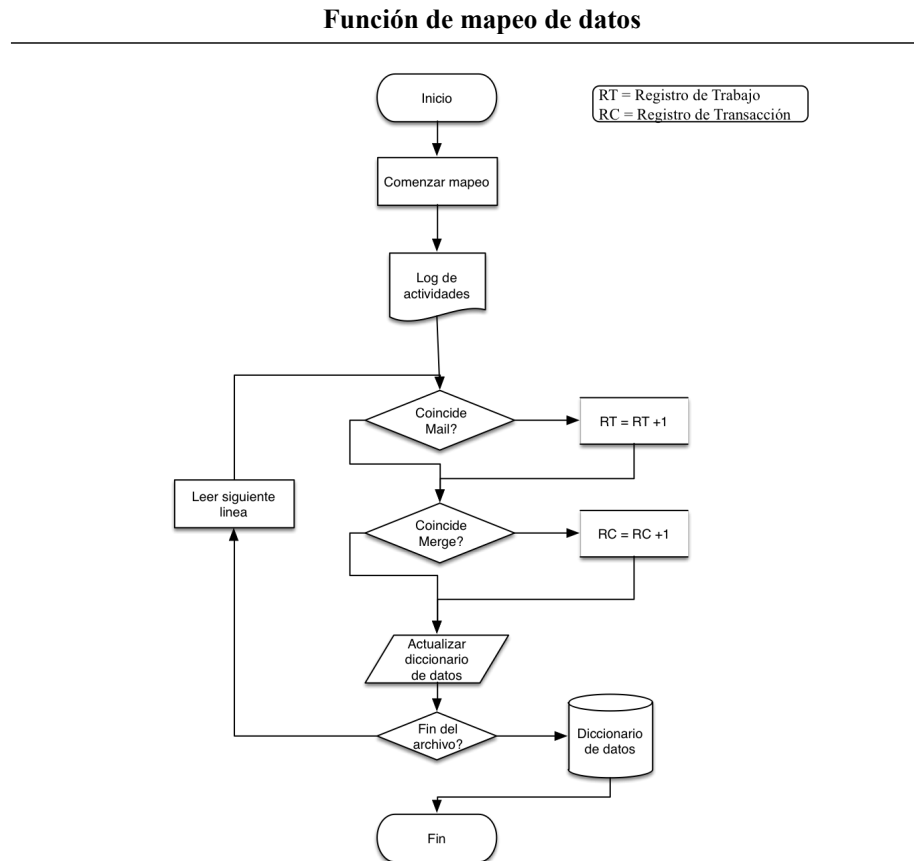
Figura 3.4 Expresiones regulares para rastrear los registros de actividad de los actores (su e-mail)



*Fuente: elaboración propia.*

<sup>30</sup> Un diccionario de datos estructura una serie de datos en un arreglo de pares ordenados (clave, valor) (Metz, 2012).

Figura 3.5 Diagrama de flujo para la función de mapeo de datos. En esta, los registros se rastrean por dominio de email y las transacciones por la palabra clave “merge”

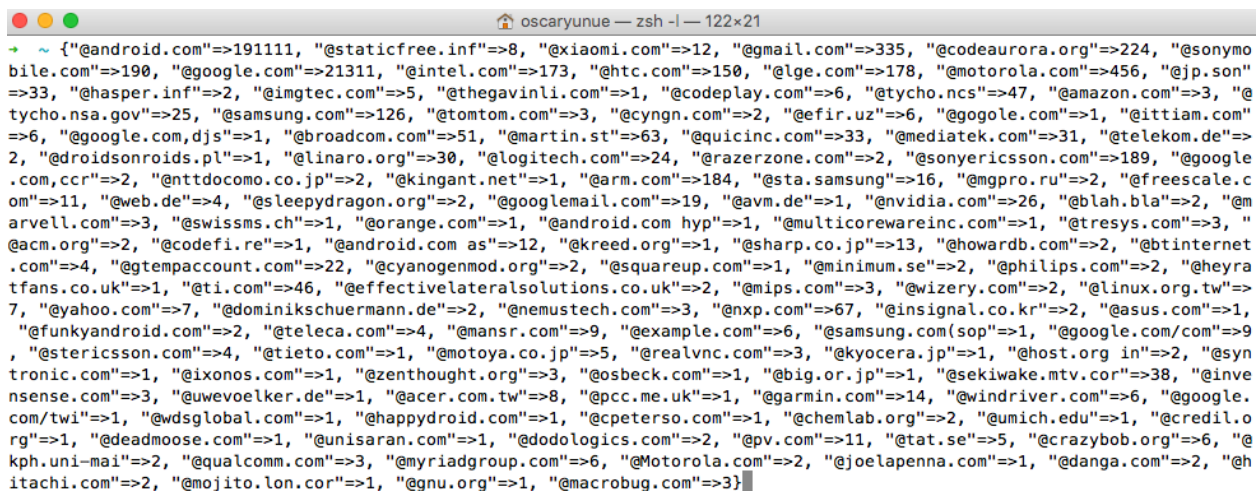


Fuente: elaboración propia

La función de mapeo fue desarrollada en el lenguaje de programación *Ruby* versión 2.2.1, y la información sobre su funcionamiento es la siguiente:

1. *Entrada*. La bitácora de actividades en formato texto.
2. *Proceso*. La función lee «una por una» las líneas del texto buscando correos de e-mail para extraer el dominio. Posteriormente, acumula las incidencias y las categoriza. Finalmente, se aplica el mismo procedimiento para calcular las transacciones.
3. *Salida*. Diccionario de datos {“clave”=> valor}.

Figura 3.6 Captura de pantalla del diccionario de datos: {"@dominio"=&gt;sumatoria}



```

+ ~ {"@android.com"=>191111, "@staticfree.inf"=>8, "@xiaomi.com"=>12, "@gmail.com"=>335, "@codeaurora.org"=>224, "@sonymo
bile.com"=>190, "@google.com"=>21311, "@intel.com"=>173, "@htc.com"=>150, "@lge.com"=>178, "@motorola.com"=>456, "@jp.son
"=>33, "@hasper.inf"=>2, "@imgtec.com"=>5, "@thegavinli.com"=>1, "@codeplay.com"=>6, "@tycho.ncs"=>47, "@amazon.com"=>3, "@
tycho.nsa.gov"=>25, "@samsung.com"=>126, "@tomtom.com"=>3, "@cyngn.com"=>2, "@efir.uz"=>6, "@gogole.com"=>1, "@ittiam.com"
=>6, "@google.com,djs"=>1, "@broadcom.com"=>51, "@martin.st"=>63, "@quicinc.com"=>33, "@mediatek.com"=>31, "@telekom.de"=>
2, "@droidsonroids.pl"=>1, "@linaro.org"=>30, "@logitech.com"=>24, "@razerzone.com"=>2, "@sonyericsson.com"=>189, "@google
.com,ccr"=>2, "@nttdocomo.co.jp"=>2, "@kingant.net"=>1, "@arm.com"=>184, "@sta.samsung"=>16, "@mgpro.ru"=>2, "@freescale.c
om"=>11, "@web.de"=>4, "@sleepydragon.org"=>2, "@googlemail.com"=>19, "@avm.de"=>1, "@nvidia.com"=>26, "@blah.bla"=>2, "@m
arvell.com"=>3, "@swissms.ch"=>1, "@orange.com"=>1, "@android.com hyp"=>1, "@multicorewareinc.com"=>1, "@tresys.com"=>3, "
@acm.org"=>2, "@codefi.re"=>1, "@android.com as"=>12, "@kreed.org"=>1, "@sharp.co.jp"=>13, "@howardb.com"=>2, "@btinternet
.com"=>4, "@gtempaccount.com"=>22, "@cyanogenmod.org"=>2, "@squareup.com"=>1, "@minimum.se"=>2, "@philips.com"=>2, "@heyra
tfans.co.uk"=>1, "@ti.com"=>46, "@effectivelateralsolutions.co.uk"=>2, "@mips.com"=>3, "@wizery.com"=>2, "@linux.org.tw"=>
7, "@yahoo.com"=>7, "@dominikschuermann.de"=>2, "@nemustech.com"=>3, "@nxp.com"=>67, "@insignal.co.kr"=>2, "@asus.com"=>1,
"@funkyandroid.com"=>2, "@teleca.com"=>4, "@mansr.com"=>9, "@example.com"=>6, "@samsung.com(sop"=>1, "@google.com/com"=>9
, "@stericsson.com"=>4, "@tieto.com"=>1, "@motoya.co.jp"=>5, "@realvnc.com"=>3, "@kyocera.jp"=>1, "@host.org in"=>2, "@syn
tronic.com"=>1, "@ixonos.com"=>1, "@zenthought.org"=>3, "@osbeck.com"=>1, "@big.or.jp"=>1, "@sekiwake.mtv.cor"=>38, "@inve
nsense.com"=>3, "@uwevoelker.de"=>1, "@acer.com.tw"=>8, "@pcc.me.uk"=>1, "@garmin.com"=>14, "@windriver.com"=>6, "@google
.com/twi"=>1, "@wsglobal.com"=>1, "@happydroid.com"=>1, "@cpeterso.com"=>1, "@chemlab.org"=>2, "@umich.edu"=>1, "@credil.o
rg"=>1, "@deadmoose.com"=>1, "@unisaran.com"=>1, "@dodologics.com"=>2, "@pv.com"=>11, "@tat.se"=>5, "@crazybob.org"=>6, "@
kph.uni-mai"=>2, "@qualcomm.com"=>3, "@myriadgroup.com"=>6, "@Motorola.com"=>2, "@joelapenna.com"=>1, "@danga.com"=>2, "@h
itachi.com"=>2, "@mojito.lon.cor"=>1, "@gnu.org"=>1, "@macrobug.com"=>3}

```

Fuente: propia, capturado de la terminal del sistema operativo OSX.

El Anexo A, proporciona más información de la función de mapeo y el acceso a su código fuente.

### 3.4.3. Función de reducción

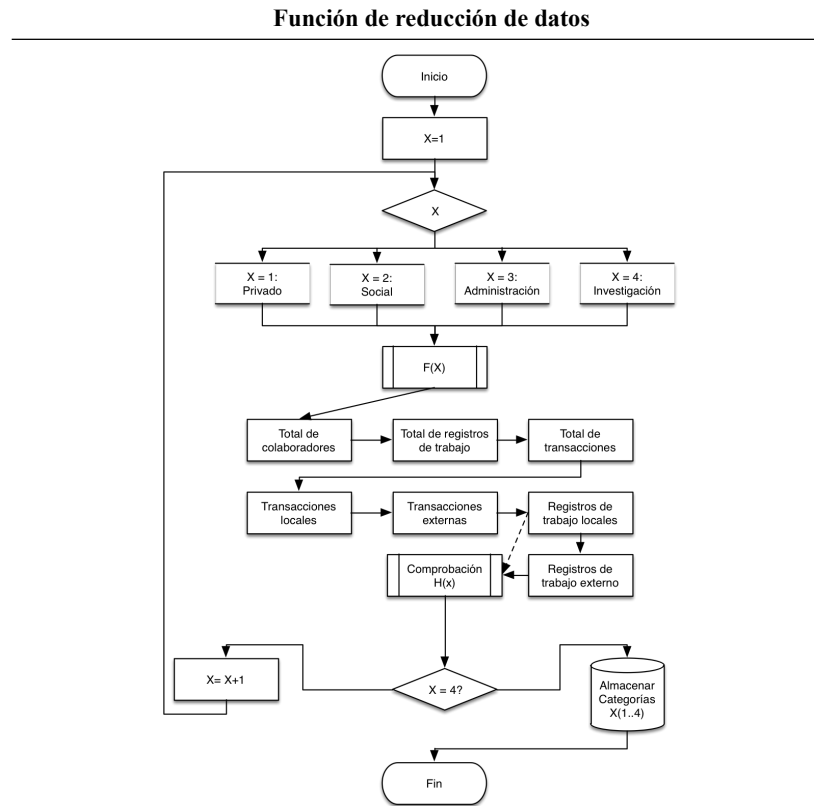
Con este procedimiento se construyó una instancia de base de datos a partir de la información almacenada en el diccionario de datos (previamente generado por la función de mapeo). La función de reducción de datos (ver figura 3.7) fue desarrollada en el lenguaje de programación R versión

3.4.1, y su funcionamiento se describe en la siguiente lista:

- *Entrada.* El diccionario de datos proporcionado por la función de mapeo.
- *Proceso.* En este procedimiento se utiliza el diccionario de datos para calcular cada una de las variables por sector, y después, las ordena en una tabla. Cabe señalar, que también contiene el cálculo de la hipótesis, esto se hizo con la intención de construir todas las columnas (variables) antes de la etapa de análisis.
- *Salida.* Una instancia<sup>31</sup> de la base de datos.

<sup>31</sup> En el contexto de programación, una instancia refiere a un objeto temporal que se construye a partir de un modelo o clase.

Figura 3.7 Procedimiento para la función de reducción, calcula todas las variables y las clasifica por sector.



Fuente: elaboración propia.

El Anexo B provee más información acerca de la función de reducción y el acceso a su código fuente, mientras que el Anexo D contiene los resultados de la función de mapeo, en otras palabras, la base de datos con los proyectos de cada sector.

### 3.5. Análisis estadístico

En esta parte se describen las herramientas y técnicas de análisis utilizadas en el presente trabajo. El análisis a la transferencia tecnológica comprende valores descriptivos generales y por sector, el balance entre los flujos de innovación locales y externos, la estimación de costos de gestión y el análisis de variabilidad entre sectores. Posteriormente, se describen las técnicas de análisis

utilizadas para analizar y describir los otros elementos de la innovación abierta: propiedad intelectual, gestión del conocimiento, y comercialización.

### 3.5.1. Herramientas utilizadas en el análisis

En este trabajo se utilizaron a varias herramientas para la obtención de datos y su posterior análisis.

Las instrucciones utilizadas en la construcción de la base de datos se pueden consultar en el Anexo F, mientras que las principales herramientas utilizadas en la etapa del análisis se describen en la siguiente lista:

- *Lenguaje de programación R*. Es un lenguaje de programación de código abierto, provee un entorno de software para generar estadísticos, gráficos, reportes, e implementar modelos de aprendizaje automático (Chang, 2012).
- *El lenguaje de programación Ruby*. Es un lenguaje de programación orientado a objetos, con una gramática compleja pero expresiva y una biblioteca de clases centrales (Metz, 2012). Este lenguaje se suele utilizar para modelar estructuras de datos, ya que permite definir la estructura de datos en las clases.
- *Neo4j*. Es un manejador de base de datos orientado a grafos, a diferencia de una base de datos relacional, las BD orientadas a grafos se construyen a partir de relaciones entre las observaciones, así, las consultas se pueden concentrar sobre las interacciones (Robinson, Webber, & Eifrem, 2015). La estructura de Neo4j es de propósito general y permite modelar todo tipo de escenarios.
- *La herramienta de modelado grafico Omnigraffle*. Es una herramienta para modelado vectorial de diagramas. Permite a los usuarios ajustar prácticamente cualquier propiedad de una forma, línea u otro elemento gráfico en un lienzo digital. *Omnigraffle* posee una comunidad que ha expandido el uso de esas herramientas a todo tiempo de aplicaciones (Olsen, 2010).

La tabla 3.2 describe las herramientas que se utilizaron en cada parte del análisis:

Tabla 3.2 Herramientas y técnicas principales utilizadas en el análisis

Criterio	Tipo	Técnica de análisis	Herramienta
Transferencia tecnológica	Variable	ANOVA de un factor	R
		Regresión lineal	R
Propiedad intelectual	Variable	Regresión logística	R
Gestión del conocimiento	Constante	Modelado de red social	Neo4j
Comercialización	Constante	Modelado de entidades relaciones (RAD)	Omnigraffle

Fuente: elaboración propia.

### 3.6. Transferencia tecnológica

#### 3.6.1. Estadísticos descriptivos generales

El propósito del análisis descriptivo es proporcionar información sobre el comportamiento de los flujos de innovación en un contexto general. En la tabla 3.3 se describen las métricas utilizadas en este paso.

Tabla 3.3 Criterios de medición para los flujos de innovación

Métricas de flujos de innovación	Descripción
Total de colaboradores	Cantidad de colaboradores registrados en un proyecto abierto
Total de registros de trabajo	Total de flujos de trabajo (global)
Total de transacciones	Total de interacciones entre actores (global)
Registros de trabajo locales	Sustrae al autor del total de registros de trabajo; representa el trabajo del actor local
Registros de trabajo externos	Sustrae a los actores externos de total de registros de trabajo, representa a los flujos externos de innovación
Transacciones locales	Sustrae al autor del total de transacciones; representa la integración de flujos externos a la cadena local de valor
Transacciones externas	Sustrae a los actores externos del total de transacciones; representa las interacciones en las que no participa el actor local

Nodos de transacción	Actores con al menos un registro de actividad destinado a la integración de flujos externos
Hipótesis	Resultados del balance entre flujos de innovación locales y los externos

*Nota. Las métricas corresponden a las variables construidas por la función de reducción. Fuente propia.*

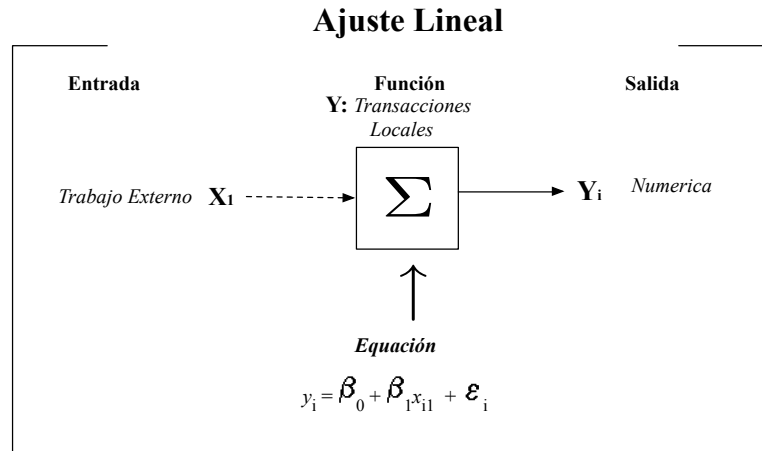
### 3.6.2. Actores predominantes

El balance entre flujos locales y externos es una de las cuestiones centrales de este trabajo ya que se vincula con el papel protagónico en la cadena de valor abierta. Para este aspecto, se utilizó la función de reducción (figura 3.7) para incluir este parámetro a la base de datos. Su procedimiento consiste en extraer las contribuciones de actor local y comparar con los actores externos. En el cálculo se utilizó el número “1” para indicar que el supuesto es verdadero y “0” para los casos alternativos.

### 3.6.3. Análisis de flujos externos de innovación frente a transacciones locales

Anteriormente, se ha planteado por Mortara, Minshall, Napp, y Slacik (2010) que los flujos externos de innovación exigen trabajo para integrarlos a la cadena de valor. Para calcular este aspecto, se aplicó un ajuste lineal (ver figura 3.8) para calcular los flujos externos de innovación en función de las transacciones locales.

Figura 3.8 Entrada y salida del ajuste lineal



Fuente: elaboración propia.

Tal como explica Weisberg (2013), la regresión lineal simple es un método estadístico que permite resumir y estudiar las relaciones entre dos variables continuas. Básicamente, con esta técnica de análisis se indica la relación entre los flujos externos de innovación en función de las transacciones locales para integrarlas a la cadena de valor.

El propósito de este paso es describir la relación del ajuste y contrastar los 4 sectores; no se contempló ningún tipo de ajuste sobre el modelo base.

### 3.6.4. Análisis de variabilidad a los actores

Una de las cuestiones centrales de este trabajo es describir la transferencia tecnológica en el ecosistema de innovación abierta, y de acuerdo con los fundamentos teóricos el entorno incluye el comportamiento de los actores. Para este aspecto, se calculó la variabilidad del cociente del total de colaboradores entre el trabajo externo con respecto al sector.

El ANOVA de un factor es una técnica de análisis que permite calcular la variabilidad de un factor (variable categórica) con respecto a variables continuas. A través de este modo se puede estudiar la variabilidad en los niveles y entre los niveles del factor (Cardinal y Aitken, 2013).

Asimismo, se calculó la variabilidad entre los sectores con la técnica ANOVA de un factor, el factor se compone de 4 niveles, y cada nivel corresponde a uno de los sectores (privado, social, administración, e investigación). Con esta técnica de análisis se calcularon los coeficientes de variabilidad en cada sector y entre los sectores.

Los criterios utilizados en esta parte del análisis fueron los siguientes:

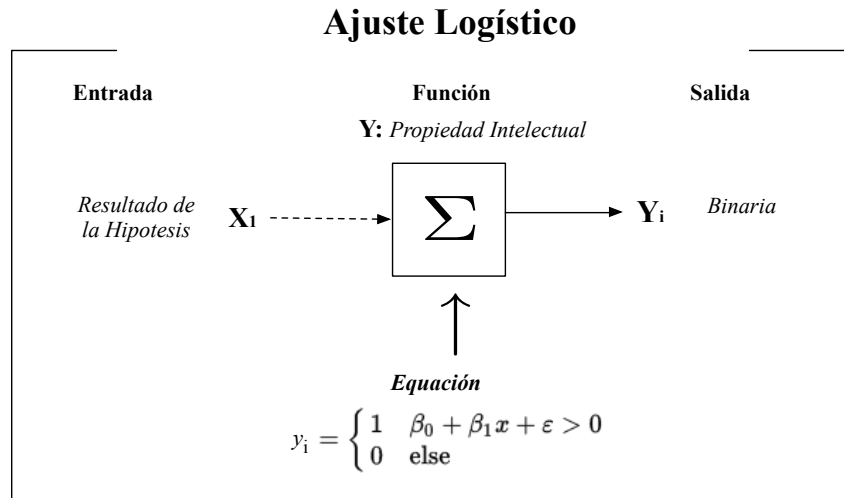
- Las poblaciones tienen la misma varianza, es decir, la suposición se basa en la homogeneidad en la variabilidad de los niveles como base al aplicar la comparación.
- Las poblaciones se distribuyen uniformemente. Se aplicó la estandarización de escala de rango (0 a 1) sobre los valores de la muestra para aseverar el aspecto comparable entre los actores.
- Cada valor es muestreado independientemente. Este aspecto se validó con los criterios de inclusión, específicamente, se excluyeron las re-distribuciones para evitar dependencias entre los proyectos seleccionados.

### **3.7. Propiedad intelectual**

El propósito de este paso consiste en relacionar la transferencia tecnológica con la propiedad intelectual. Para cuantificar esta relación, se analizó el tipo de licencia en función de la hipótesis nula (la presencia del actor predominante). Para este propósito, se aplicó una regresión logística (ver figura 3.9). Tal como (O'Connell, 2006) explica, el ajuste lógico es un método estadístico para

analizar un conjunto de datos en el que hay una o más variables independientes asociadas a un respuesta lógica, en este caso, se utilizó “1” para indicar verdadero y “0” para indicar falso.

Figura 3.9 Entrada y salida del ajuste logístico



Fuente: elaboración propia

### 3.8. Gestión del conocimiento

Este parámetro fue declarado como constante en el planteamiento metodológico. De acuerdo con la literatura consultada, las actividades en los proyectos se encuentran automatizadas en la plataforma de colaboración, o en otras palabras, el procedimiento de gestión de conocimiento es el mismo para cualquiera de los proyectos alojados en la población.

De acuerdo el planteamiento teórico, la estructura de los proyectos abiertos sostiene a actores distribuidos geográficamente (Thung et al., 2016), los flujos de trabajo se desarrollan en paralelo (Biazzini, Monperrus, & Baudry, 2014), y su estructura se puede modelar con herramientas de análisis de red social.

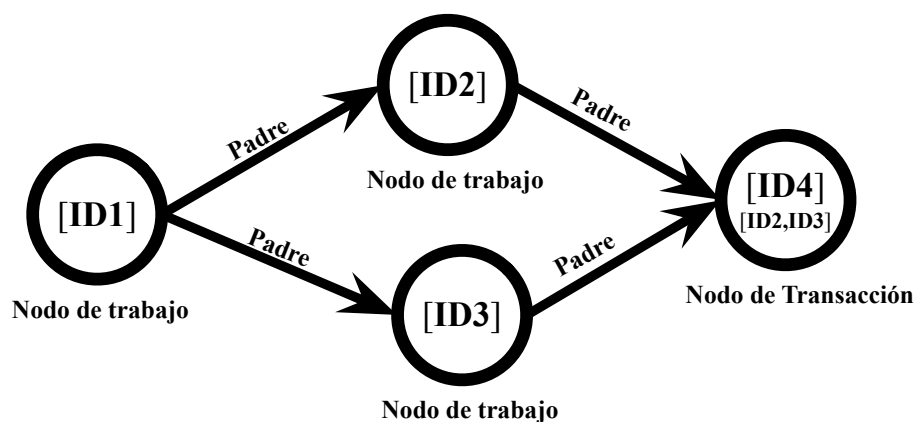
Partiendo de las especificaciones anteriores, se utilizó la versión comunitaria de la plataforma de desarrollo *Neo4j* para comprobar la estructura de colaboración. Robinson et al.

(2015) explican que las BD orientadas a grafos se construyen definiendo la relación entre los registros. Asimismo, estudios relacionados como (Zafarani, Abbasi, & Liu, 2009) y (Biazzini et al., 2014) indican que las interacciones entre los flujos de trabajo definen los enlaces para nodos conectados «su acoplamiento».

La figura (3.10) ejemplifica los criterios utilizados para modelar los flujos de trabajo y las transacciones.

Figura 3.10 La figura muestra las relaciones utilizadas para modelar la estructura los flujos de trabajo y las transacciones, se puede apreciar que las transacciones son el producto de dos flujos [id].

### Criterios de acoplamiento para flujos de trabajo y transacciones



Fuente: elaboración propia

Con base a los criterios anteriores, en la plataforma de análisis Neo4<sup>32</sup> la estructura de una red se define con el lenguaje de consultas Cypher<sup>33</sup>. En la figura (3.12) se puede observar un fragmento del código utilizado para establecer la relación que conecta a los registros de actividad.

<sup>32</sup> <https://neo4j.com/>

<sup>33</sup> <https://neo4j.com/developer/cypher/>

Figura 3.11 Fragmento de código fuente utilizado para replicar la estructura de la red

### Código para replicar la estructura de la red de colaboración

```

```cypher
BEGIN
create constraint on (c:Commit) assert c.sha1 is unique;
COMMIT
BEGIN
CREATE (:Commit {author_email:'foo@bar.com',date_iso_8601:'2014-05-22 20:53:05 +0200',
parents:['b6393fc9d5c065fd42644caad600a9b7ac911ae2'],refs:['HEAD', 'origin/master', 'master', 'in-index'],
sha1:'934cac9fe6cd0188be642b3e609b529edaad527',subject:'Some commit message',timestamp:'1400784785'});

CREATE (:Commit {author_email:'bar@foo.com',date_iso_8601:'2014-05-22 13:22:10 +0200',
parents:['7765539ff17310f2c736ee7f0a8fc5e05180e262', '2d3abe010c36214b71c9bbcaa9f6063947068de'],
sha1:'b6393fc9d5c065fd42644caad600a9b7ac911ae2',subject:'Merge pull request #2445 from foo/bar',timestamp:'1400757730'});
...
MATCH (parent:Commit {sha1:"934cac9fe6cd0188be642b3e609b529edaad527"}),
      (child:Commit {sha1:"b6393fc9d5c065fd42644caad600a9b7ac911ae2"})
CREATE (parent)-[:PARENT]-(child);
...
COMMIT
```

```

Fuente: elaboración propia.

En caso de requerirlo, el Anexo F contiene el código fuente utilizado para replicar la estructura de la red de colaboración.

### 3.9. Comercialización

La comercialización, al igual que la gestión del conocimiento fue declarada como un parámetro constante en el planteamiento metodológico de esta investigación. Anteriormente, se ha planteado que la comercialización comprende una serie de transacciones entre clientes y proveedores; también, que en los productos abiertos el aspecto monetario no existe en esas interacciones. Con base a estas especificaciones, en esta etapa se analizó el procedimiento de comercialización para describir los pasos establecidos por la plataforma de colaboración; particularmente, la etapa de inmersión y el posicionamiento.

El diagrama de entidad relación (RAD) es una técnica para modelar las interacciones entre los diferentes roles que interactúan en un procedimiento. En este caso, se utilizó para describir el procedimiento de inmersión y las interacciones entre proveedores y clientes. De acuerdo con Ali et al. (2017), el modelado de un RAD se enfoca en definir sistemas iterativos, en este caso, el

proceso de inmersión al mercado de los productos. El procedimiento de inmersión fue modelado tal como se indica en la guía proporcionada por (Pipinellis, 2015).

En lo que respecta al posicionamiento, se ha revisado la fórmula de posicionamiento de los productos en *PageRank* para describir los criterios considerados por la plataforma de colaboración para calcular el posicionamiento de los productos.

**CAPÍTULO IV**  
**RESULTADOS Y DISCUSIÓN**

En este capítulo se revelan los hallazgos encontrados en el análisis a los proyectos de innovación abierta. Siguiendo con la estructura de la metodología mixta, los resultados se presentan en orden secuencial. Todo esto, con el propósito de conservar la estructura de los ejes conceptuales del ecosistema de innovación abierta en la industria del software.

Primero, se proporcionan los resultados del análisis a los flujos de colaboración para describir la transferencia tecnológica. Esta parte contiene cuatro elementos: 1) se proporcionan estadísticos descriptivos de la muestra con el propósito de establecer el contexto general sobre el comportamiento de las variables y los actores, 2) se describe la presencia del actor predominante, 3) se presenta el cálculo de los efectos de los flujos externos de trabajo en la cadena local de producción (costos de gestión), y 4) se describe el comportamiento de los actores con base al análisis de varianza.

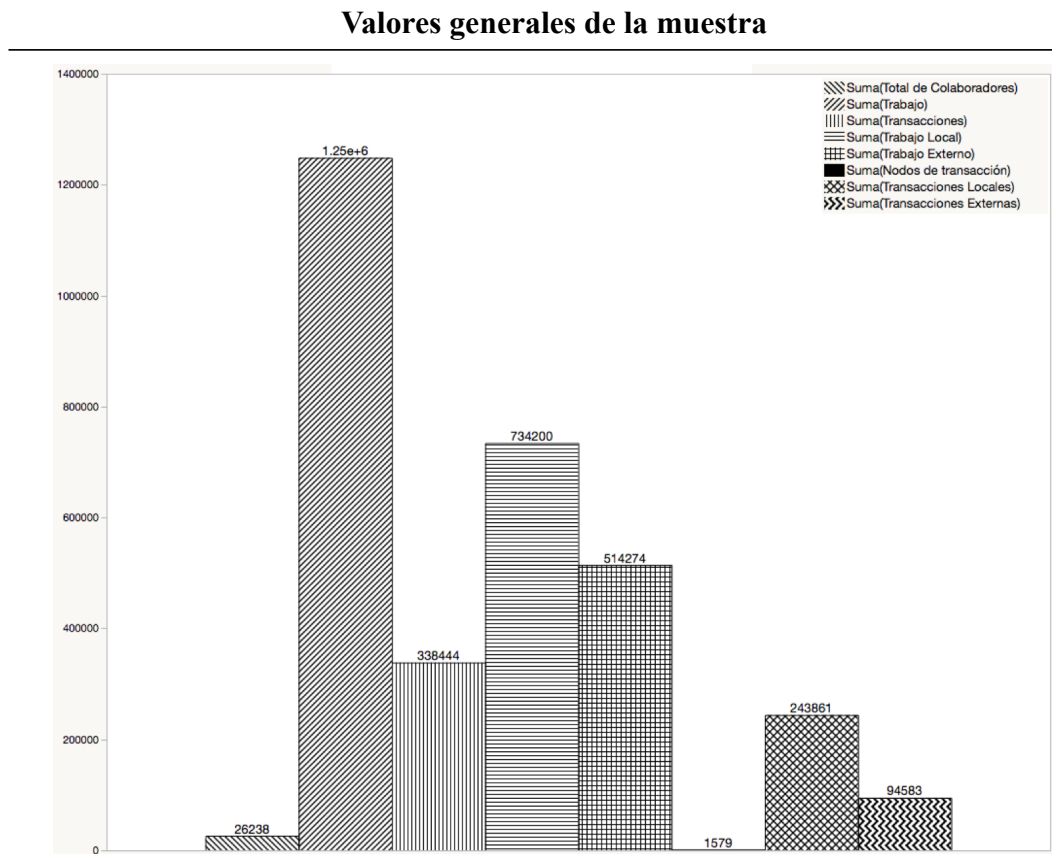
Una vez explicados los aspectos de la transferencia tecnológica se procede a complementar los resultados de la transferencia tecnológica con los aspectos de la innovación. Para este fin, se incorporan los efectos de la licencia para explicar la propiedad intelectual; después, se describe la estructura de las interacciones entre los actores para la gestión de conocimientos; y finalmente, se describe el procedimiento de comercialización para los productos abiertos.

## 4.1. Resultados del análisis a la transferencia tecnológica

### 4.1.1. Estadísticos descriptivos de muestra

Después de la recolección y clasificación de los registros de actividad, se obtuvo que en los 100 proyectos de innovación abierta que conforman la muestra participan un total de 26,238 individuos y acumulan 1,248,474 registros de trabajo. El total de trabajo se divide en 734,200 para los flujos locales y 517,274 flujos externos de innovación, lo que equivale a 1.42 registros locales por cada externo. Las proporciones de las variables se ilustran en el gráfico de barras de la figura 4.1.

Figura 4.1 Distribución de valores en los flujos de innovación de la muestra



Fuente: elaboración propia.

En lo que respecta a las transacciones, se detectaron un total de 1,579 nodos de transacción; es decir, aquellos actores con al menos un registro de actividad destinado a integrar el trabajo de

actores externos. Al sustraer a los 100 autores (actores locales) suman un total de 1,479 nodos de transacción externos. Al mismo tiempo, el total de transacciones suman 338,444; 243,861 para los flujos locales, y 94,583 para los flujos externos. Esto equivale a un 72% en costos de gestión para los actores locales y un 28% que se distribuye entre los actores externos, equivalente a 2.57 registros de transacciones locales por cada externa.

*Tabla 4.1 Estadísticos descriptivos de los flujos de innovación*

| Columna         | N   | Grados de libertad | Media   | Desviación estándar | Suma    | Mínimo | Máximo |
|-----------------|-----|--------------------|---------|---------------------|---------|--------|--------|
| Colaboradores   | 100 | 99                 | 262.380 | 440.811             | 26238   | 3      | 3307   |
| Trabajo         | 100 | 99                 | 12484.7 | 31150.4             | 1248474 | 169    | 215512 |
| Trabajo Local   | 100 | 99                 | 7342.00 | 22541.3             | 734200  | 99     | 191111 |
| Trabajo Externo | 100 | 99                 | 5142.74 | 11696.7             | 514274  | 11     | 83899  |

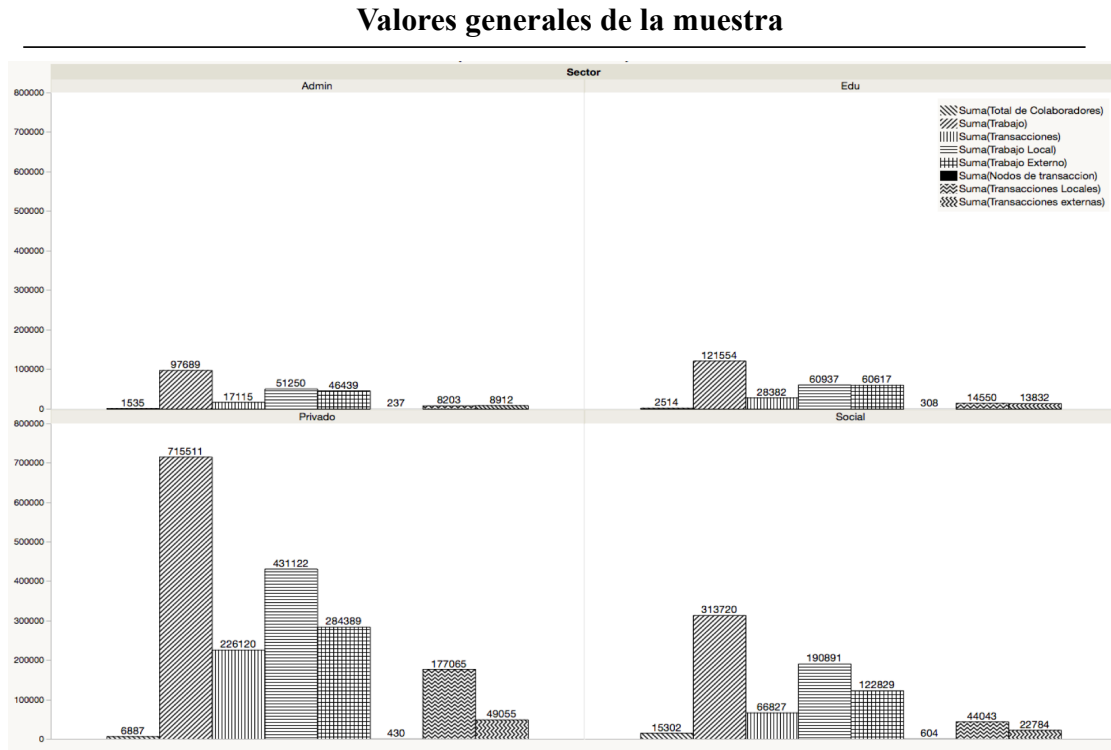
*Fuente: elaboración propia.*

Los estadísticos descriptivos de la tabla 4.1 reflejan un margen extenso en los límites de los valores observados. En cuestiones metodológicas, esto sugiere dos cosas; primero, prestar atención en la distribución interna de cada actor; y después, poner en perspectiva el comportamiento interno con el externo, es decir, la variación entre los sectores para detectar que tan uniforme es la distribución local en cada sector, y después, comparar el grupo.

#### **4.1.2. Estadísticos descriptivos por sector**

Una vez descritos los valores generales de la muestra, ahora se procede a describir el comportamiento local de los actores.

Figura 4.2 Información de los flujos de innovación por sector



Fuente: elaboración propia.

La figura 4.2 revela una distribución de valores relativamente uniforme entre los actores. Claramente, el sector privado y el social son los más activos; en particular, el sector social acumula una cantidad mayor de colaboradores, mientras que el sector privado lo supera con más del doble de registros de trabajo. Por otro lado, el sector educativo y el de administración comparten similitudes en el balance de sus valores que los diferencian de privado y el social. Asimismo, existe un balance equitativo entre en los flujos externos y locales en el sector de administración y el de investigación.

Tabla 4.2 Estadísticos descriptivos por sector

|                                       | <i>Privado</i> | <i>Social</i> | <i>Admin</i> | <i>Edu</i> |
|---------------------------------------|----------------|---------------|--------------|------------|
| <i>Flujo de trabajo local</i>         | 60.25%         | 60.85%        | 52.45%       | 50.13%     |
| <i>Flujo de trabajo externo</i>       | 39.75%         | 39.15%        | 47.55%       | 49.87%     |
| <i>Frecuencia I/e (trabajo)</i>       | 1.53           | 1.55          | 1.1          | 1.01       |
| <i>Transacciones locales</i>          | 78.30%         | 65.90%        | 47.93%       | 51.26%     |
| <i>Transacciones Externas</i>         | 21.70%         | 34.91%        | 52.07%       | 48.74%     |
| <i>Frecuencia I/e (transacciones)</i> | 3.6            | 1.93          | 0.92         | 1.05       |

Fuente: elaboración propia.

Las métricas para los sectores de la tabla 4.2 revelan un comportamiento heterogéneo en lo que se refiere a la distribución de los valores, principalmente, en el balance de flujos locales y externo. Por ejemplo, el sector privado registra un cociente de 1.53 flujos de trabajo y 3.6 de transacciones, mientras que en de administración el balance es inverso 1.1 y 0.92 respectivamente.

#### 4.1.3. Resultados de la hipótesis

En el 81 de los 100 casos estudiados se encontró que el autor del proyecto de innovación abierta también es el actor que más contribuye a la cadena de producción. Al haber especificado un  $\alpha = 0.05$  el 81% se toma la hipótesis alternativa. La tabla 4.3 contiene el resto de métricas sobre los resultados de la hipótesis general.

Tabla 4.3 Resultados de la hipótesis general

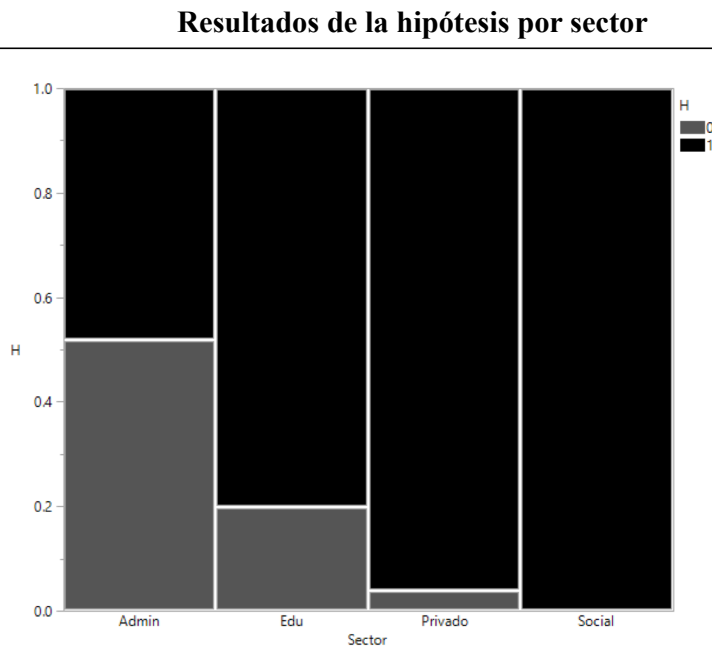
|  |           |
|--|-----------|
| Media  | 81        |
| Desviación estándar                          | 0.3942772 |
| Error estándar de la media                   | 0.0394277 |
| Extremo superior del IC al 90% para la media | 0.8754654 |
| Extremo inferior del IC al 90% para la media | 0.7445346 |
| N  | 100       |
| N cero                                       | 19        |

Fuente: elaboración propia.

#### 4.1.4. Resultados de la hipótesis por sector

Aunque la hipótesis resultó alternativa, en los 19 casos localizados se puede apreciar una tendencia que varía considerablemente de un sector a otro (ver figura 4.3); para ser más específicos, la mayoría se encontraron en el sector de administración con 13, seguido del educativo con 5, y 1 en el sector privado.

Figura 4.3 Resultados de la hipótesis por sector



Fuente: elaboración propia.

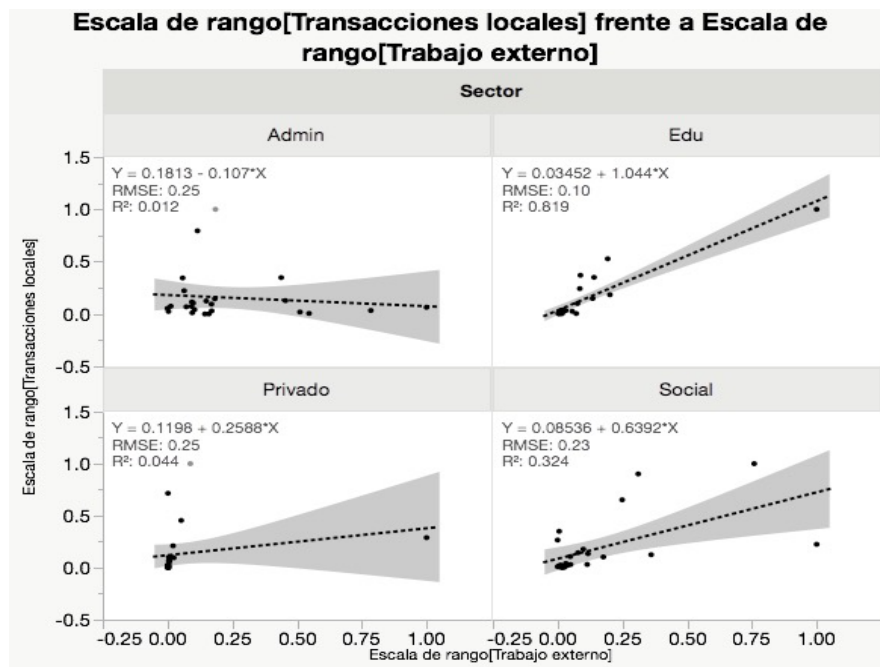
Al poner en perspectiva los resultados de la hipótesis por sector (figura 4.3) con el valor acumulado de las variables por sector (figura 4.2), se puede apreciar una relación inversa entre la proporción de las variables y la presencia del actor predominante. A simple vista, y a juzgar por esta relación, se puede suponer que la posibilidad de la hipótesis nula disminuye conforme aumentan los registros de trabajo. No obstante, en la figura (4.2) también se puede observar una diferencia entre el trabajo externo y las transacciones locales de los actores, el estudio de esta

relación es fundamental en el proceso de producción, ya que el trabajo externo tarde o temprano tiene que integrarse al producto.

#### 4.1.5. Costos de gestión

Continuando con el análisis, ahora se describen los costos vinculados a la gestión de conocimiento entre el actor local y los externos para cada sector. En particular, el propósito de este paso es describir el peso de los flujos externos sobre el esfuerzo de integración. El cálculo de este aspecto se ilustra en la figura 4.4, con el ajuste lineal de los costos de gestión en función de los flujos externos de trabajo para cada actor.

Figura 4.4 Ajuste lineal de las transacciones locales en función del trabajo externo por sector



Fuente: Elaboración propia.

En primera instancia, el único ajuste negativo se localiza en el sector de administración. Un error en el ajuste ( $R^2$ ) de 0.044 para el sector privado, 0.044 en el sector social, 0.012 en el de administración, y en el educativo 0.819. Estos resultados indican que el comportamiento de los

proyectos dentro de cada sector es heterogéneo; sin embargo, los tres casos positivos indican que el peso de los flujos externos se traduce en costos de gestión sobre los autores, excepto por el sector de administración.

A juzgar por la dispersión de las observaciones, es probable que los costos de gestión involucren variables endógenas en cada proyecto de innovación abierta. La tabla 4.4 resume la información del ajuste. Cabe mencionar que el análisis a los costos de gestión se basa en la posición del el ajuste lineal; el resto de coeficientes sobre la regresión lineal se proporcionan en el anexo E.

*Tabla 4.4 Resultados del ajuste lineal*

| <i>Trabajo Externo</i> | <i>Ordenada al origen de Y</i> | <i>Pendiente</i> | <i>R<sup>2</sup></i> |
|------------------------|--------------------------------|------------------|----------------------|
| <i>Social</i>          | 0.0853                         | 0.639            | 0.324                |
| <i>Privado</i>         | 0.1198                         | 0.258            | 0.033                |
| <i>Admin</i>           | 0.1813                         | -0.107           | 0.012                |
| <i>Edu</i>             | 0.0345                         | 1.044            | 0.819                |

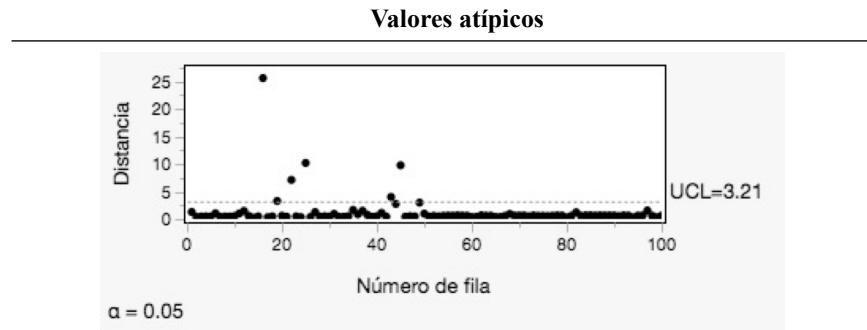
*Fuente: Elaboración propia*

## 4.2. Análisis de variación entre sectores

### 4.2.1. Detección de valores atípicos

En la muestra seleccionada se identificaron proyectos de innovación abierta que poseen valores de colaboración lo suficientemente altos para sesgar el análisis de variación. En este caso, –se utilizaron las distancias de Jackknife, en la figura 4.5 se indica como una línea de corte que separa a las observaciones con valores altos del resto. Con base a este criterio, se aplicó una eliminación sistemática antes de proceder al análisis. Básicamente, se encontraron 5 proyectos que superan el UCL, y tres que lo intersectan. Las observaciones señaladas fueron excluidas del análisis de variación.

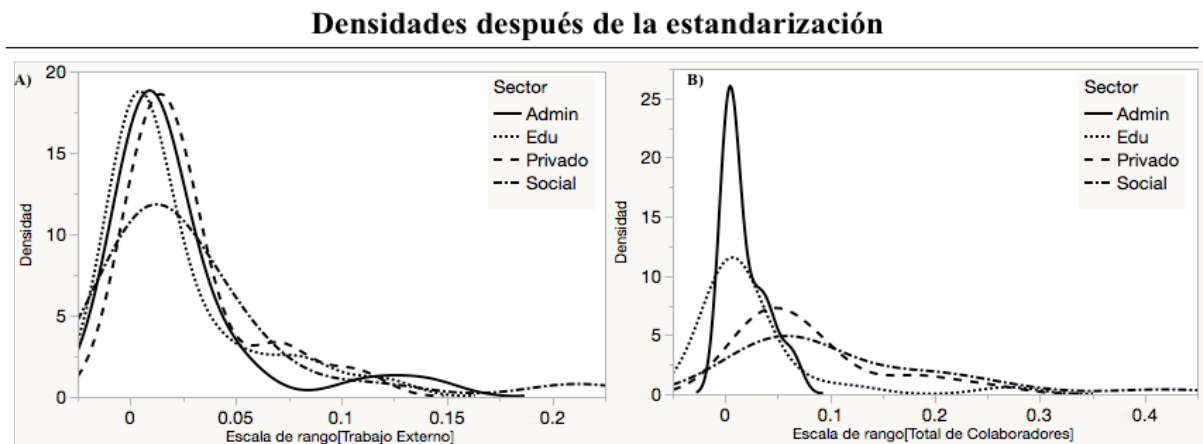
Figura 4.5 Valores atípicos por número de fila. Técnica: Distancias Jackknife



### 4.2.2. Transformación a escala de rango de 0 a 1

Anteriormente, la información descriptiva reveló una distribución de valores heterogénea entre actores, por lo tanto se estandarizaron en escala de rango de 0 a 1, un paso necesario para aplicar el análisis de variabilidad de los proyectos dentro de sus respectivos sectores. Después de aplicar la estandarización, la distribución de frecuencias de los flujos externos se mantiene en un espectro comparable. Los polígonos de frecuencias suavizados en la figura 4.6 ilustran las densidades de los mismos una vez aplicada la estandarización.

Figura 4.6 Densidades de los sectores después de aplicar la estandarización



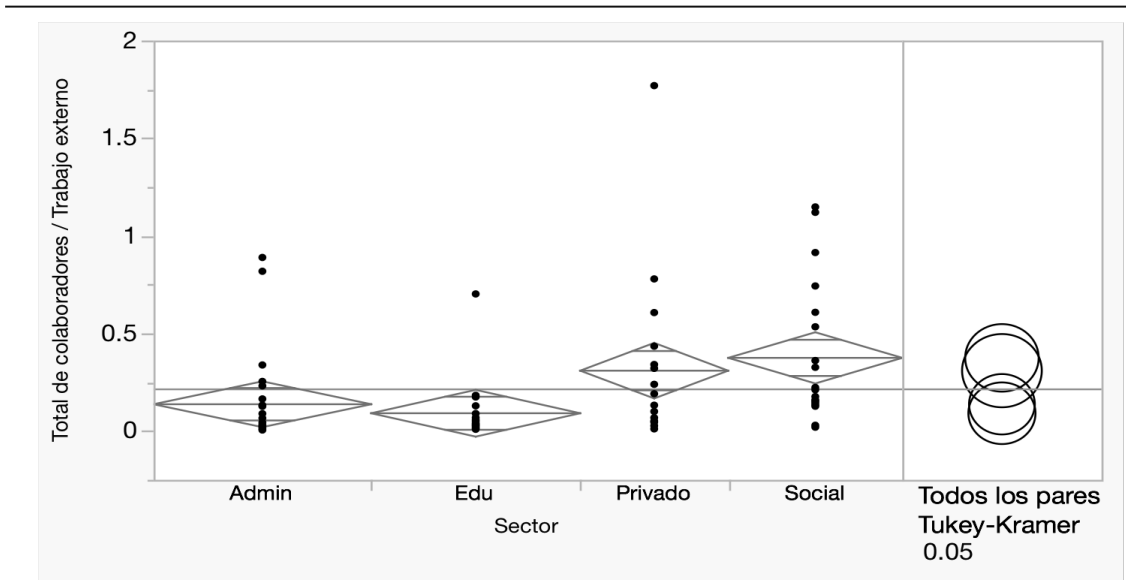
Fuente: elaboración propia. Capturado de JMP.

#### 4.2.2.1. ANOVA de un factor

Una vez estandarizado el cociente de colaboración externa, en esta parte se procede a explicar el total de variabilidad por sector (los niveles) para, posteriormente, ponerlo en perspectiva con el total de variabilidad general (el factor). Después se proporcionan los resultados de la variabilidad por pares. Asimismo, la gráfica de la figura 4.7 ilustra la variabilidad en los sectores, y entre los sectores.

Figura 4.7 Variabilidad del cociente de colaboración externa frente a sector. Los 4 círculos a la derecha representan los límites comparados de variabilidad entre los sectores.

#### Actividad de flujos externos frente a sector



Fuente: elaboración propia. Capturado de JMP.

En esta parte, se comienza por verificar si hay al menos una media diferente, contrastando la variabilidad de los niveles en torno a la media del factor. El análisis de variabilidad a los flujos externos de innovación proporcionado en la tabla 4.5 indica una suma de error cuadráticos (SSE) de 7.13, y 1.18 en la suma de cuadrados del tratamiento (SSTR) asociados a los niveles del factor, en total acumulan 8.31 para la suma de cuadrados totales (SST).

Tabla 4.5 Resultados del análisis de variación general

| Criterio | Suma de cuadrados | Media de cuadrados | Razón F | p > F   |
|----------|-------------------|--------------------|---------|---------|
| Niveles  | 1.1809926         | 0.393664           | 4.5266  | 0.0055* |
| Error    | 7.1312486         | 0.086966           |         |         |
| C. Total | 8.3122413         |                    |         |         |

Fuente: elaboración propia

En la tabla 4.5, también se puede observar que a partir del SSTR y SSE, se calculó una media de cuadrados MSTR de 0.39, y MSE de 0.086 respectivamente, que se traducen en una razón F (MSTR/MSE) de 4.52. El resultado de razón F indica que al menos un sector es significativamente distinto en torno a la media de la muestra, con una probabilidad de 0.0055. Considerando que entre más se aleja la razón F de 1 implica un incremento entre la variabilidad inter-grupal con respecto a la intra-grupal.

#### 4.2.3. Resumen de la variabilidad en los niveles

En el paso anterior, se obtuvo que al menos un nivel es significativamente diferente, ahora se procede a describir la variabilidad cada grupo. En la tabla 4.6 se indican límites relativamente estrechos para los extremos calculados en el sector de administración y el educativo. El error estándar refleja una estimación de variabilidad relativamente uniforme en estos dos sectores equivalente a un margen de error entre 0.0589 a 0.06, respectivamente; más arriba, se encuentra el sector social con 0.065, y 0.071 para el privado.

Tabla 4.6 Resultados de variabilidad intragrupal

| Nivel   | Media    | Error estándar | Extremo inferior del IC al 95% | Extremo superior del IC al 95% |
|---------|----------|----------------|--------------------------------|--------------------------------|
| Admin   | 0.137076 | 0.05898        | 0.0197                         | 0.25441                        |
| Edu     | 0.090997 | 0.06020        | -0.0288                        | 0.21075                        |
| Privado | 0.309084 | 0.07152        | 0.1668                         | 0.45137                        |
| Social  | 0.374667 | 0.06594        | 0.2435                         | 0.50585                        |

Fuente: elaboración propia

#### 4.2.4. Prueba HSD de Tukey

Después de haber rechazado la hipótesis nula y descrito la variabilidad entre los niveles del factor, ahora se procede con la prueba de comparaciones múltiples para la hipótesis alternativa. La tabla 4.7 muestra los valores críticos de la prueba HSD (*Honestly Significant Difference*) de Tukey, los valores se calcularon con base a un cuantil de confianza “q” con de 2.62 que se extiende a una probabilidad desde  $\alpha = 0.05$  (el límite superior) hasta el .01 (el límite inferior). La representación gráfica de este paso se proporcionó previamente en la figura 4.7.

Al mismo tiempo, los resultados proporcionados en la tabla 4.7 indican que el sector privado comparte similitudes con los demás actores, mientras que el sector social es significativamente diferente del sector educativo y el de administración. Asimismo, la variabilidad de colaboración externa entre el educativo y el de administración es similar.

Tabla 4.7 Matriz del umbral de HSD

|         | Social   | Privado  | Admin    | Edu      |
|---------|----------|----------|----------|----------|
| Social  | -0.23575 | -0.18035 | 0.04320  | 0.05795  |
| Privado | -0.18035 | -0.25571 | -0.03300 | -0.01824 |
| Admin   | 0.04320  | -0.03300 | -0.21521 | -0.20046 |
| Edu     | 0.05795  | -0.01824 | -0.20046 | -0.21521 |

Nota. Los valores positivos indican pares de medias que son significativamente distintas

#### 4.2.5. Análisis de varianza por pares

Hasta este punto se ha detectado que la variabilidad de colaboración externa tiene un efecto significativo con respecto al sector al que los proyectos pertenecen, en este paso se procede a describir la comparación las probabilidades de la comparación por pares. La tabla 4.8 contiene el valor p para para los niveles calculados en el paso anterior. Con un  $\alpha = 0.05$ , se pueden identificar dos grupos similares, el social y el privado con una diferencia de 0.0655 y un valor p de 0.89, y el de administración con el de investigación con una diferencia de 0.014 y un valor p de 0.997. Por

otro lado, las diferencias entre el sector social y el segundo grupo resultan estadísticamente significativas, con una probabilidad de 0.0078 y 0.0129 respectivamente.

*Tabla 4.8 Informe de las diferencias ordenadas por pares*

| Nivel   | - Nivel | Diferencia | Error estándar de la diferencia | Límite de control inferior | Límite de control superior | Valor p |
|---------|---------|------------|---------------------------------|----------------------------|----------------------------|---------|
| Social  | Edu     | 0.2836697  | 0.0860467                       | 0.057954                   | 0.5093855                  | 0.0078* |
| Social  | Admin   | 0.2689162  | 0.0860467                       | 0.043200                   | 0.4946320                  | 0.0129* |
| Privado | Edu     | 0.2180865  | 0.0900929                       | -0.018243                  | 0.4544161                  | 0.0812  |
| Privado | Admin   | 0.2033329  | 0.0900929                       | -0.032997                  | 0.4396625                  | 0.1169  |
| Social  | Privado | 0.0655832  | 0.0937540                       | -0.180350                  | 0.3115167                  | 0.8970  |
| Admin   | Edu     | 0.0147536  | 0.0820423                       | -0.200458                  | 0.2299651                  | 0.9979  |

*Fuente: elaboración propia*

Con base al análisis de variabilidad se identificaron 3 sub conjuntos, en la tabla 4.9 se presentan como A, AB, y B. Ahí, se puede apreciar que diferencia entre A y B es poco más de triple. En la prueba HSD de Tukey se detectaron similitudes entre el sector de administración y el de investigación, la comparación múltiple refleja una diferencia entre el sector social y el conjunto anterior de 0.0129 y 0.0078 respectivamente, mientras que el sector privado no refleja diferencias significativas con ninguno de los otros niveles, de acuerdo con la media del nivel, el sector privado se posiciona entre el social, y el primer conjunto.

*Tabla 4.9 Informe de letras de unión*

| Nivel          | Media          |
|----------------|----------------|
| Social         | A 0.37466713   |
| Privado        | A B 0.30908388 |
| Administración | B 0.10575097   |
| Investigación  | B 0.09099740   |

*Nota. Los niveles que no están conectados se consideran significativamente distintos*

### 4.3. Resultados del análisis al entorno de la innovación abierta

En esta segunda etapa, el análisis se enfoca en los otros tres aspectos de la innovación considerados como elementos fundamentales: propiedad intelectual, gestión del conocimiento y comercialización. Cabe mencionar, que el análisis estadístico solo aplica para la propiedad intelectual, en el planteamiento se ha explicado que el comportamiento de los otros dos elementos

es constante para cualquier proyecto abierto, o al menos en lo que respecta a la población de *Github*.

### 4.3.1. Propiedad intelectual

En este paso, se describe el resultado de la hipótesis en función del tipo de licencia. El modelo de la tabla 4.10 toma como referencia la hipótesis nula para establecer los parámetros de respuesta. La diferencias entre los valores de la variable dependiente y los estimados reflejan una desviación residual de 76.208 y una desviación nula de 97.245, ambas se mantienen por debajo de los grados de libertad. Asimismo, el AIC (criterio de información de *Akaike*), extiende el cociente de la verosimilitud con respecto al número de parámetros (los tipos de licencia) hasta 106.21. Por lo tanto, los coeficientes indican que no hay una buena bondad de ajuste pero tampoco particularmente baja, con un equivalente de precisión estimada de 62%.

Tabla 4.10 Resultados del ajuste logístico de propiedad intelectual (Licencia) en función de  $H_0=1$

|             | Estimación | Error E.  | Valor z | Pr(> z ) | Cod. de Sig. |
|-------------|------------|-----------|---------|----------|--------------|
| (Intercept) | -1.0986    | 0.8165    | -1.346  | 0.17846  |              |
| AGLP 3      | 19.6647    | 6522.6386 | 0.003   | 0.99759  |              |
| Apache 2    | 3.2958     | 1.1055    | 1.982   | 0.00287  | **           |
| ARR         | 2.7081     | 1.3663    | 1.982   | 0.04747  | *            |
| BSD         | 1.7918     | 1.472     | 1.217   | 0.2235   |              |
| BSD 3       | 2.8904     | 1.354     | 1.354   | 0.03279  | *            |
| CC          | 2.0149     | 1.472     | 1.724   | 0.08479  | .            |
| GLP 2       | 19.6647    | 3261.3194 | 1.217   | 0.2235   |              |
| GLP 3       | 1.7918     | 1.472     | 0.004   | 0.9966   |              |
| MIT         | 19.6647    | 0.9747    | 3.194   | 0.0014   | **           |
| Mozilla PL2 | 3.1135     | 6522.6386 | 0.003   | 0.99759  |              |
| SIL OFL     | 19.6647    | 6522.6386 | 0.003   | 0.99759  |              |

Código de significancia: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Desviación nula: 97.245 con 99 grados de libertad

Desviación residual: 76.208 con 85 grados de libertad

AIC: 106.21

Valor estimado de precisión: 0.62

Número de iteraciones de puntuación de Fisher: 17

En la tabla 4.10 se pueden observar cinco tipos de licencia que explican significativamente la hipótesis nula. La MIT es una de las que mejor explica esta relación con 0.0014, seguida de la

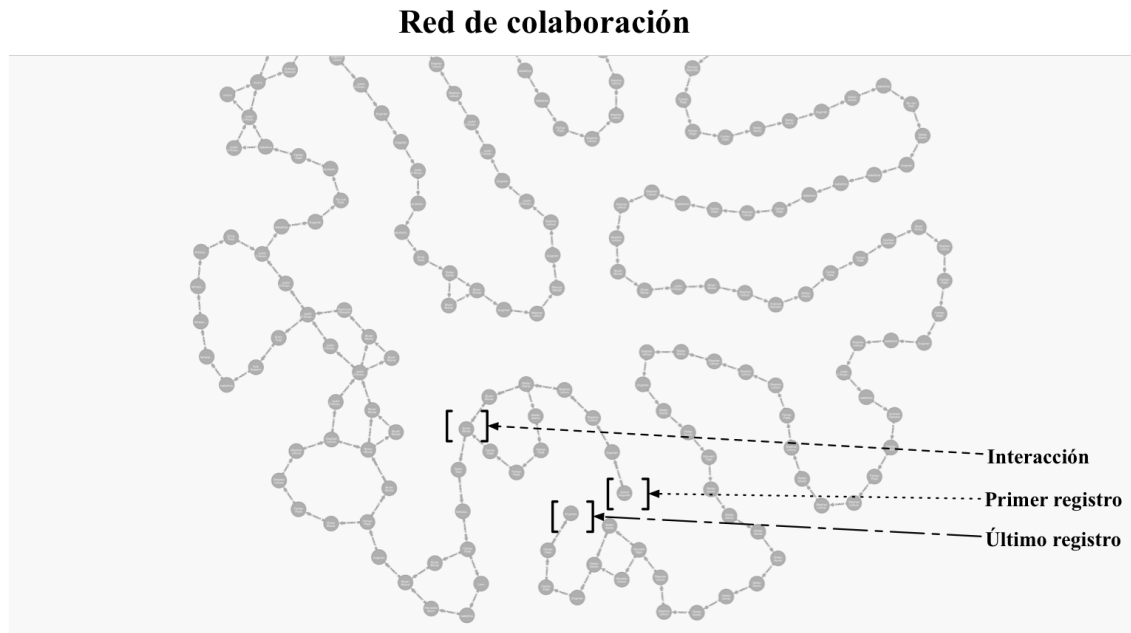
Apache 2 con 0.00287, cabe mencionar estas licencias suelen ser las más utilizadas en el sector social en donde no se identificaron casos de hipótesis alternativa, de ahí le siguen la ARR y GLP 2 con 0.047 y 0.084 respectivamente.

#### **4.3.2. Gestión del conocimiento**

El concebir la idea de la transferencia tecnológica en una red global de colaboración, implica que los actores pueden estar distribuidos (en el caso de la muestra a lo largo de la internet), y que el proyecto abierto permite integrar el trabajo de esos actores a la cadena de valor.

Como se indica en la literatura consultada, los registros de trabajo se almacenan en un orden secuencial, esto se comprobó con el análisis a la estructura de registros. Considerando esa estructura, cada uno de los nodos (los registros) se vinculan con el registro que le precede desde que se registra el proyecto hasta la última versión disponible del producto. De esta manera, el trabajo de los diferentes actores se va desarrollando en paralelo evitando que entren en conflicto entre ellos (de Laat, 2007).

Figura 4.8 La captura de pantalla de Neo4j ilustra los primeros registros en cadena de valor del proyecto (Microsoft/vscode<sup>34</sup>).



*Fuente: elaboración propia. Capturado de Neo4j.*

En la figura 4.8 el identificador “primer registro” representa la creación del producto, “último registro” la versión disponible del producto, mientras que la integración de flujos externos se indica como “intersección”. Al mismo tiempo, se ilustra como las intersecciones entre los flujos locales de innovación y los externos forman una red de colaboración (Biazzini, Monperrus, & Baudry, 2014).

### 4.3.3. Comercialización

De acuerdo con la revisión de literatura, en la industria del software, los proyectos de innovación abierta suelen manejarse principalmente por popularidad, y el posicionamiento toma importancia debido a que potencializa el incremento de la red de colaboradores externos. A raíz de de-la

---

<sup>34</sup> <https://github.com/Microsoft/vscode>

ausencia de transacciones monetarias, su comercialización se basa principalmente en el posicionamiento del producto, mientras que las transacciones de clientes y proveedores se centran en el intercambio de conocimiento. En esta parte se describen las interacciones entre los roles involucrados en la producción del producto, y después el procedimiento de posicionamiento.

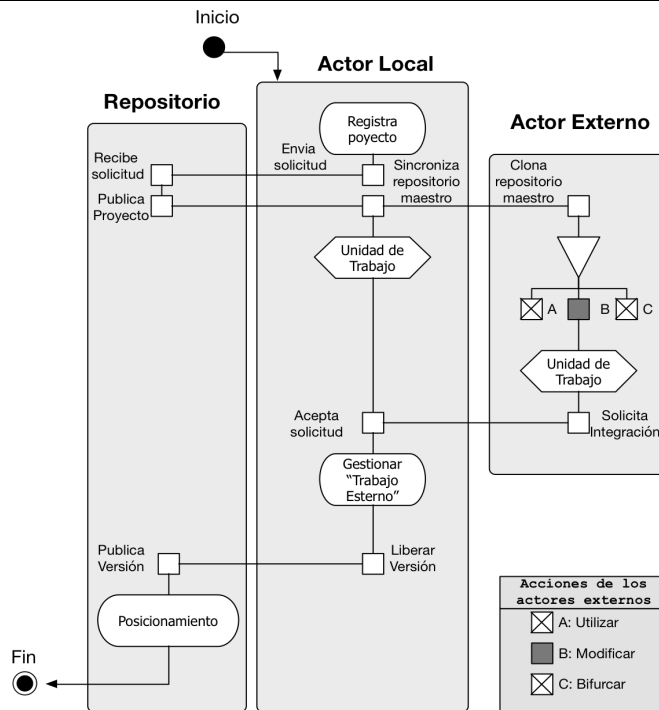
#### **4.3.4. Procedimiento de inmersión al mercado**

De acuerdo con la guía para *Github* de (Pipinellis, 2015), en el procedimiento de inmersión al mercado se identificó que a las interacciones entre los actores locales y los externos siguen un procedimiento automatizado en la plataforma de colaboración. La figura 4.9 ilustra las interacciones entre los roles involucrados; mientras que el procedimiento de inserción al mercado de la población estudiada se describe en la siguiente lista:

1. En primer lugar, el autor (actor local) registra el proyecto en la plataforma de colaboración;
2. Inmediatamente después, se solicita al actor local seleccionar el tipo de licencia;
3. La plataforma de colaboración registra el proyecto;
4. Los clientes tienen acceso al uso, modificación y re-distribución del producto;
5. Los actores locales y externos trabajan en paralelo hasta que se solicita la integración de recursos externos;
6. Al aceptar la integración de recursos externos, se procede con la integración de los recursos externos a la cadena local de valor;
7. Después de la integración, el actor local libera una nueva versión en la plataforma de colaboración.

Figura 4.9 El diagrama entidad relación (RAD) ilustra las interacciones entre los roles a lo largo del proceso de comercialización en la plataforma de la población (Github)

**Interacciones involucradas en el procedimiento de comercialización**



Fuente: elaboración propia

En esta etapa, se identificaron dos cuestiones que merecen atención; en primer lugar, la distribución del bien no es expuesta explícitamente por la plataforma de colaboración; y en segundo lugar, la opción de re-distribución (también conocida como “bifurcación”) del proyecto se utiliza como una de las principales funcionalidades en la plataforma.

Cabe mencionar, que la opción para re-distribuir el producto se encuentra deshabilitada en el entorno privado, ya que esta permite a los actores emplear las concesiones de la PI inclusiva para disponer de los recursos completos de un proyecto y comenzar uno nuevo, tal como indican (Nyman & Lindman, 2013).

### 4.3.5. Procedimiento de posicionamiento

De acuerdo con la documentación de la función de posicionamiento *PageRank* proporcionada por *Langville (2011)*, el procedimiento para ubicar a los productos abiertos se calcula a través de una fórmula que computa la popularidad de los productos para calcular su posición. La fórmula de posicionamiento es la siguiente:

*Fórmula 1 Posicionamiento de los proyectos abiertos (PageRank)*

|  |     |
|--|-----|
| $PR(A) = (1 - d) + d(PR(Ti)/C(Ti) + \dots + PR(Tn)/C(Tn))$ | (1) |
|--|-----|

*Fuente: Langville (2011)*

En la fórmula de posicionamiento, PR(A) es la posición en el ranking de un producto “A”, PR(Ti) es la posición de los clientes “Ti” que se enlazan al producto “A”, y C(Ti) es el número de enlaces salientes del cliente “Ti” hasta el “Tn” (la sumatoria), mientras que “d” es un factor de amortiguación que ajusta el resultado en un rango de 0 a 1.

**CAPÍTULO V**  
**CONCLUSIONES Y RECOMENDACIONES**

En este capítulo se toman los hallazgos más sobresalientes del análisis para explicar a la transferencia tecnológica del software en el entorno de la innovación abierta. Aunque los repositorios de innovación abierta no proporcionan información segmentada sobre los flujos de innovación, a través de un análisis de texto fue posible ordenar los registros de actividad de 100 proyectos de innovación abierta para analizarlos. A través de la metodología mixta fue posible describir a varios elementos de la innovación abierta por separado y articularlos en orden secuencial con base a los fundamentos teóricos encontrados en la revisión de literatura (capítulo II).

Asumiendo que la innovación se compone de cuatro elementos fundamentales: transferencia tecnológica, propiedad intelectual, gestión del conocimiento y comercialización, seguidamente se proporciona una explicación sobre las implicaciones del balance entre flujos locales y externos en la producción de bienes compartidos, las implicaciones del tipo de licencia, se describe la estructura de los flujos de innovación y la etapa de inmersión y posicionamiento de los productos. De esta manera se proporciona la explicación de la transferencia tecnológica en el ecosistema de innovación abierta.

Adicionalmente, se proveen algunas recomendaciones generales encaminadas a identificar aspectos clave y reducir la incertidumbre en la implementación de proyectos abiertos. Finalmente, se discute sobre las limitantes de la investigación y se puntualizan algunas líneas de investigación para futuras investigaciones.

## **5.1. La transferencia tecnológica del software en la innovación abierta**

En la industria del software, existen diversas interrogantes acerca de la transferencia de tecnología que se desarrolla en el entorno de la innovación abierta. Por una parte, la TTS se ha enfrentado, desde hace décadas, a un ritmo de actualización cada vez más apresurado, y también, a complicaciones para adaptar las estrategias de transferencia a los diferentes escenarios de aplicación. Por otra parte, la innovación abierta implica un cambio de paradigma en el desarrollo y comercialización de estos productos tecnológicos.

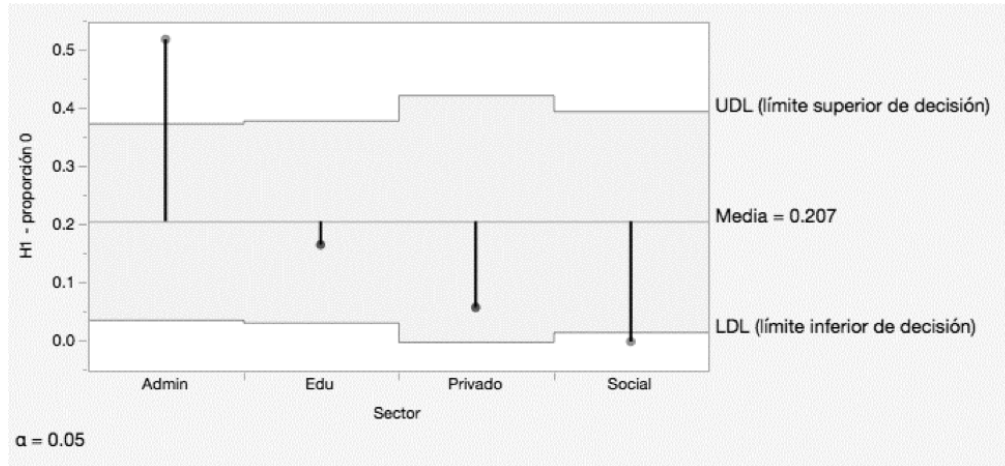
Actualmente, no se sabe hasta qué punto es recomendable para una organización renunciar a la exclusividad de un producto para acceder a los beneficios de los flujos externos de innovación. Investigaciones previas, señalan que cuando la recaudación de flujos externos es baja el proyecto no es redituable, no obstante, si un actor externo supera al autor del proyecto entonces corre el riesgo de perder la dirección del proyecto.

A partir de estos fundamentos, se planteó el criterio de “actor predominante” para estudiar el balance en la transferencia de flujos de innovación; específicamente, para identificar al actor que supera en rendimiento a todos los demás. Considerando que fueron seleccionados algunos de los proyectos abiertos más populares, se planteó como supuesto que el autor de un proyecto abierto tendría que ser el acreedor a ese rol.

En contraste con resultados obtenidos, se observó que los flujos de innovación locales solo superaron a los externos en 81 de los 100 proyectos seleccionados. En cuanto a los 19 casos de hipótesis alternativa, no se encontraron casos en el sector social, 1 en el sector privado, 5 en el de

investigación y 13 en el de administración. Como se puede ver en la figura 5.1, el rol del actor predominante varía ampliamente entre los sectores.

Figura 5.1 Umbral de decisión para el peso de la hipótesis alternativa frente a sector



Fuente: elaboración propia

El trabajo de gestión fue otro aspecto que se consideró en la transferencia tecnológica de los flujos de innovación, ya que estudios relacionados señalan que los flujos externos de innovación demandan esfuerzo para integrarlos a la cadena de valor. En términos generales se encontró que esta cuestión equivale a 2.57 registros de transacciones locales por cada externa, equivalentes a un 72% de costos de gestión para los actores locales. Al mismo tiempo, en tres de los actores se encontró una relación positiva para las transacciones locales del autor en función de los flujos externos de innovación.

## 5.2. Propiedad intelectual

Tomando en cuenta que el tipo de licencia define la configuración legal para el intercambio de recursos entre los colaboradores, se consideró pertinente determinar su relación con el balance de flujos de innovación, en particular, con el de actor predominante ( $H_0$ ) para describir como afectaba

el tipo de licencia al balance entre los flujos locales y externos. En 5 de ellas se encontró una relación significativa, la MIT es la que mejor explica la relación con una probabilidad de 0.0014, seguida de la Apache 2 con 0.00287, de ahí le siguen la BSD-3, ARR, y CC con 0.032, 0.047, y 0.084 respectivamente.

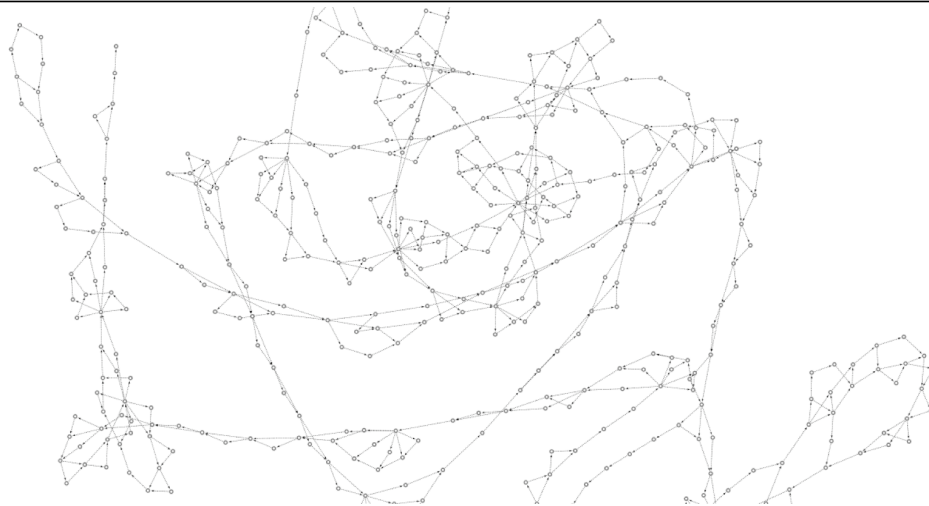
### 5.3. Gestión del conocimiento

En un producto de software, la gestión de conocimientos desempeña un rol fundamental ya que constituye la integración de valor agregado sobre el bien. En esta parte, se procedió a describir la estructura que respalda al proceso de transferencia de tecnológica; particularmente, los registros de trabajo y transacciones. Al modelar los registros de trabajo y vincularlos con las transacciones se verificó la estructura de los flujos de innovación. En la figura 5.2 se ilustra gráficamente la estructura de los flujos de innovación en un proyecto abierto.

*Figura 5.2 Red de colaboración.*

#### **Estructura de la integración entre flujos locales y externos**

---



*Fuente: elaboración propia. Desarrollado con Neo4j, y capturado de Cytoscape<sup>35</sup>.*

---

<sup>35</sup> <http://cytoscape.org>

A través del modelado de la estructura de registros se pudo corroborar lo siguiente:

- La innovación incremental se presenta como una cadena de registros integrados en orden secuencial;
- Incluso si los colaboradores pertenecen a diferentes sectores, sus actividades se desarrollan en paralelo para evitar conflictos de acceso concurrente;
- El paralelismo se compensa con la integración de recursos externos, representado por la intersección en los flujos de trabajo.

#### **5.4. Comercialización**

De acuerdo con la literatura consultada el procedimiento de inmersión y el de posicionamiento se encuentran automatizados, después de modelarlos se encontró:

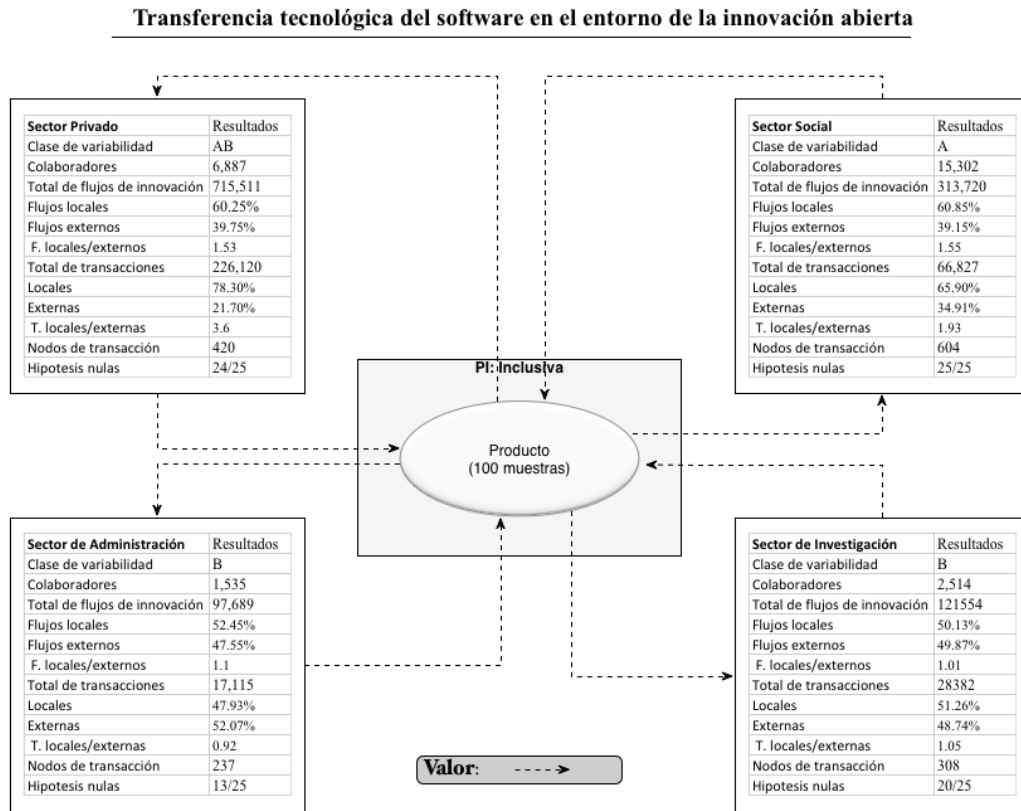
1. *Inmersión*. El procedimiento de inmersión se lleva a cabo a través de una serie de pasos específicos. A través de estos, el autor registra el producto, selecciona la licencia, y lo publica en la plataforma de colaboración para habilitar el acceso de los actores externos.
2. *Posicionamiento*. El posicionamiento de los productos se ejecuta con una fórmula integrada en (*PageRank*), esta calcula la popularidad basándose en el conteo de clientes enlazados al producto. Cabe mencionar que en la documentación de la fórmula no se especifica el tipo de usuario para los clientes, es decir, en lo que respecta al cálculo de la popularidad se trata por igual a los usuarios y a los colaboradores.

#### **5.5. La transferencia tecnológica en el entorno de innovación abierta**

Hasta este punto, se ha explicado que la innovación abierta se basa en la integración de recursos externos y a un procedimiento de transferencia tecnológica incremental, inclusivo, y de alcance global que se desenvuelve en ese entorno. Ahora, se procede a incorporar las características de los

actores involucrados. En la figura 5.3 se integran las métricas recolectadas sobre la transferencia tecnológica en cada sector.

Figura 5.3 Conceptualización del objeto de estudio en función de los resultados.



Nota. Cada elemento de la muestra es un producto independiente. Fuente: Elaboración propia.

La transferencia tecnológica a menudo es considerada un problema adaptativo, que depende de las condiciones del entorno de desarrollo y las características del producto. Asimismo, en la muestra seleccionada se identificaron algunas de esas particularidades. Con base a los resultados obtenidos en la siguiente lista describen algunas de las más distintivas:

- *Sector Social*
  - Posee más colaboradores que los otros tres sectores combinados (15,302 – 10,936).
  - En todos los casos resultó ser el predominante.

- Los autores contribuyen el 60.85% a la cadena de valor equivalente a 1.93 locales por cada externa.
- Los autores absorben el 65,90% de los costos de gestión.
- Es el sector que sufre una mayor variabilidad.
- *Sector Privado*
  - Es el sector con mejor rendimiento, tiene dos veces menos colaboradores que el sector social pero lo duplica en registros de trabajo.
  - Suele ser el actor predominante en la cadena de producción (24 de 25).
  - Sus proyectos exigen la mayor cantidad de costos de gestión (78.30%).
  - Independientemente de que es el que más registros de trabajo acumula también suele tener un comportamiento regular a nivel local, y comparte similitudes con el sector social en la variabilidad con respecto a la media de la muestra.
- *Sector de Administración*
  - El comportamiento más extraño ocurre en este sector, ya que el 52.07% de transacciones externas indica que la mayoría de los costos de gestión los absorben los actores externos.
  - Los autores contribuyen el 52.45% a la cadena de valor equivalente a 1.1 locales por cada externa.
  - No suele ser el actor predominante en la cadena de valor (12/25).
  - Es el sector con la menor actividad en los que respecta a la distribución de valores.
- *El sector de Investigación*
  - Sus proyectos recaudan la mayor proporción de flujos externos (49.87%).
  - Los autores absorben el 51.26% de los costos de gestión.
  - Fue el autor predominante en 20 de los 25 proyectos.
  - En comparación con los demás sectores su comportamiento es el más estable, tanto en variabilidad como en el cálculo del trabajo de integración en función de los flujos externos.

## 5.6. Hallazgos

A lo largo de este trabajo, se encontró que algunos de los rasgos característicos en cada actor se reflejan en el comportamiento de sus proyectos de innovación abierta. Esto indica que la

transferencia tecnológica está condicionada por el entorno en el que se desarrolla cada producto. Por ese lado, prevalecen los problemas adaptativos en el entorno de la innovación abierta. De acuerdo con la literatura consultada, la industria del software se ha caracterizado por dificultades de adaptabilidad en las estrategias de actualización tecnológica, y estos resultados indican que en el entorno de la innovación abierta ocurre lo mismo.

Por otro lado, se localizaron algunos hallazgos generales. En la siguiente lista se puntualizan algunos de los más relevantes:

- De los cien proyectos que se analizaron en este trabajo, a nivel general, los flujos externos superan el límite del 10% especificado por Mortara et al. (2009) con un 29% global.
- Una mayor cantidad de colaboradores no necesariamente significa más flujos de trabajo externos. Se encontró que el sector social tiene más colaboradores que el privado, pero el privado supera considerablemente al social en registros de trabajo.
- Es posible que el tipo de licencia afecte el balance de flujos.
- Es posible que el autor de un proyecto abierto pierda el rol de actor predominante y mantenga la dirección del producto, especialmente en el sector de administración.
- Como cliente, la ausencia del costo de uso y la posibilidad de modificarlo un incentivo fuerte para adoptar este modelo; pero para las organizaciones representa un desafío, ya que compensar los privilegios de la exclusividad con los flujos externos de innovación puede ser un incentivo complicado de asegurar.
- Los proyectos de innovación abierta suelen manejarse principalmente por popularidad. En este apartado, el cálculo de las posiciones se basa en la sumatoria de los clientes enlazados al producto, independientemente, de su participación en la construcción de valor.

## **5.7. Recomendaciones**

Con base a los puntos anteriores, se recomienda evaluar que el tipo de licencia se adapte al modelo de negocio antes se seleccionarla, y desarrollar estrategias para gestionar el trabajo externo, particularmente en el sector privado que es el más trabajo de gestión exige.

Cabe señalar, que en el planteamiento de esta investigación no se contemplaron los costos de gestión realizados por la parte externa. Sin embargo, en dos de los sectores se encontró un alto grado de transacciones externas, en el sector de administración incluso superando a las locales. Esto indica, que los actores externos pueden absorber los costos de gestión. Por ese lado, es recomendable concebir la actividad externa como algo más que flujos de trabajo que recaudar.

Es posible que al delegar los costos de gestión la sustentabilidad de un producto libre pueda mejorar. Considerando lo anterior, se recomienda contemplar este aspecto en las estrategias de innovación abierta.

No hay mucho que agregar en lo que se refiere a los aspectos de gestión de conocimiento y comercialización, ya que se aplican las mismas reglas para cualquiera de los proyectos estudiados. Por ese lado, es recomendable concentrar las estrategias de transferencia tecnológica sobre la etapa de implementación.

Con base a las ideas anteriores, se puede concluir que una organización no puede simplemente lanzar un proyecto abierto y esperar que los recursos externos comiencen a llegar y se integren a la cadena de producción automáticamente; tal vez en el sector de administración, pero para los demás sectores implica costos de gestión y prestar atención a los riesgos asociados con el balance de flujos.

## **5.8. Limitantes**

En la etapa de análisis, se encontró un alto grado de variabilidad entre varios de los actores, principalmente en el privado y el social. Ya que los cálculos se realizaron sobre el comportamiento grupal, los resultados se limitan a estadísticos generalizados para cada actor.

Aunque la parte inferencial de este trabajo no resultó particularmente reveladora la situación se puede mejorar extendiendo el alcance de la muestra.

Otro de los inconvenientes que se encontraron en el desarrollo de esta investigación fue en la etapa de construcción de la base de datos, principalmente en el análisis de texto. Ya que cuando se rastrean datos en los que están involucrados diferentes idiomas pueden aparecer problemas en la codificación en los caracteres, en esta parte, fue necesario normalizar los registros a el estándar ISO/IEC 10646:2003. Otro contratiempo que hay que considerar son los caracteres especiales, por ejemplo, si las palabras clave tienen guiones “-” al analizarlas se pueden detectar como una resta por la plataforma de análisis.

## **5.9. Futuras líneas de investigación**

Considerando el alto nivel de variabilidad que se encontró en entre los actores, y que los resultados de este trabajo se limitan a la descripción de comportamientos generalizados en cada sector, es posible que se obtengan mejores resultados analizando la transferencia tecnológica en proyectos específicos. Similarmente, los resultados en las regresiones indican que puede haber variables endógenas que no fueron consideradas en esta investigación.

Por otro lado, aunque la mayoría de las investigaciones se concentran principalmente en atender las necesidades del sector privado, el paradigma de innovación abierta se ha comenzado a utilizar por diferentes sectores. Este escenario brinda la posibilidad de estudiar la participación de diferentes actores en un mismo proyecto. Sería interesante desarrollar estudios enfocados en la vinculación entre los sectores, ya que en el muestreo se identificaron algunos casos en los que participan centros de investigación y empresas, empresas con otros competidores, y centros de investigación con la comunidad social.

Esta última parte es importante, considerando que el sector social tiene la mayor cantidad de colaboradores, el privado tiene el mejor rendimiento, el de investigación es el que recauda el mayor porcentaje de flujos externos, y el de administración delega la mayoría de los costos de

gestión a los actores externos; es posible que cada actor pueda contribuir a la construcción de valor de un mismo proyecto desde diferentes perspectivas.

## REFERENCIAS

- Ahumada-Tello, E., Castañón-Puga, M., Castro, J. R., Suarez, E. D., Márquez, B. Y., Gaxiola-Pacheco, C., & Flores, D. L. (2011). On the Multi-Agent Modelling of Complex Knowledge Society for Business and Management System Using Distributed Agencies. In *Communications in Computer and Information Science* (Vol. 188, pp. 560–569). Berlin, Heidelberg: Springer Berlin Heidelberg. [http://doi.org/10.1007/978-3-642-22389-1\\_49](http://doi.org/10.1007/978-3-642-22389-1_49)
- Abulrub, A.-H. G., & Lee, J. (2012). Open innovation management: challenges and prospects. *Procedia - Social and Behavioral Sciences*, 41, 130–138. <http://doi.org/10.1016/j.sbspro.2012.04.017>
- Aguilar, L. J. (2016). Big Data, Análisis de grandes volúmenes de datos en organizaciones. Alfaomega Grupo Editor.
- Ameka, I., & Dhewanto, W. (2013). Technology Push vs. Market Pull in Technology University Innovation Commercialization Case Study: ITB. *Information Management and ...*
- Ali, M. S. M., Wahid, H., Subha, N. A. M., Sahlan, S., Yunus, M. A. M., & Wahap, A. R. (2017). *Modeling, Design and Simulation of Systems*. Springer.
- Andersson, J., Bengtsson, F., Ekman, J., Lindberg, E., Waldehorn, C., & Nilsson, F. (2011). Perception of innovation in companies - measuring the mindset of tangible and intangible innovation in companies (pp. 532–542). IEEE. <http://doi.org/10.1109/ITMC.2011.5996023>
- Arnkil, R., Jarvensivu, A., Koski, P., & Piirainen, T. (2011). Exploring Quadruple Helix. *Institute for Social Research*, 1–131.
- Audretsch, D. B., Lehmann, E. E., Link, A. N., & Starnecker, A. (2012). Technology Transfer in a Global Economy. (D. B. Audretsch, E. E. Lehmann, A. N. Link, & A. Starnecker, Eds.). Boston, MA: Springer Science & Business Media. <http://doi.org/10.1007/978-1-4614-6102-9>

- Bhardwaj, A., Scaria, V., Raghava, G. P. S., Lynn, A. M., Chandra, N., Banerjee, S., et al. (2011). Open source drug discovery--a new paradigm of collaborative research in tuberculosis drug development. *Tuberculosis (Edinburgh, Scotland)*, *91*(5), 479–486. <http://doi.org/10.1016/j.tube.2011.06.004>
- Biazzini, M., Monperrus, M., & Baudry, B. (2014). On Analyzing the Topology of Commit Histories in Decentralized Version Control Systems, 261–270.
- Boehm, B. W. (1984). Software Engineering Economics., *10*(1), 4–21. <http://doi.org/10.1109/TSE.1984.5010193>
- Carayannis, E. G., & Campbell, D. F. (2014). Developed democracies versus emerging autocracies: arts, democracy, and innovation in Quadruple Helix innovation systems. *Journal of Innovation and Entrepreneurship*, *3*(1), 1105. <http://doi.org/10.1186/s13731-014-0012-2>
- Carayannis, E. G., & Campbell, D. F. J. (2011). Mode 3 Knowledge Production in Quadruple Helix Innovation Systems. Springer Science & Business Media.
- Cardinal, R. N., & Aitken, M. R. F. (2013). ANOVA for the Behavioral Sciences Researcher. Psychology Press.
- Chambers, J. (2008). Software for Data Analysis. Springer Science & Business Media.
- Chang, W. (2012). R graphics cookbook. Sebastopol CA: O'Reilly Media. <http://doi.org/10.2139/ssrn.1738486>
- Chesbrough, H. W., & Appleyard, M. M. (2007). Open Innovation and Strategy. *California Management Review*, *50*(1), 57–76. <http://doi.org/10.2307/41166416>
- Dabbish, L. (2011). *Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository* (pp. 1–10).
- de Laat, P. B. (2007). Governance of open source software: state of the art. *Journal of Management*

& *Governance*, 11(2), 165–177. <http://doi.org/10.1007/s10997-007-9022-9>

Douglas, L., & Piotr, S. (2009). Innovation in the software sector. <http://doi.org/10.1787/9789264076761-en>

Easley, D., & Kleinberg, J. (2010). *Networks, Crowds, and Markets*. Cambridge University Press.

Fricke, M. (2008). The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of Information Science*, 35(2), 131–142. <http://doi.org/10.1177/0165551508094050>

Friedman, T. L. (2007). *The World Is Flat 3.0*. Picador.

Github. (2016) *GitHub State of the Octoverse: 2016*, 16 de Octubre de 2016, [en línea]. [Consulta: 2 de Noviembre de 2016]. Disponible en el World Wide Web: <https://octoverse.github.com/>.

Gomulkiewicz, R. W. (2013). *Software Law and its Application*. Wolters Kluwer Law & Business.

Hall, B. H. (2014). Open innovation and intellectual property rights–The two-edged sword.(2009).

Hu H. (2004). Software evolution based on software architecture (pp. 1092–1097). Presented at the The Fourth International Conference on Computer and Information Technology, 2004. CIT '04., IEEE. <http://doi.org/10.1109/CIT.2004.1357341>

Kalliamvakou, E., Damian, D. E., Blincoe, K., Singer, L., & German, D. M. (2015). Open Source-Style Collaborative Development Practices in Commercial Projects Using GitHub. *Icse*, 574–585. <http://doi.org/10.1109/ICSE.2015.74>

Kalliamvakou, E., Gousios, G., Blincoe, K., Singer, L., German, D. M., & Damian, D. (2016). An in-depth study of the promises and perils of mining GitHub. *Empirical Software Engineering*, 21(5), 1–38. <http://doi.org/10.1007/s10664-015-9393-5>

Katz, A. (2012). Towards a Functional Licence for Open Hardware. *International Free and Open Source Software Law Review*, 41–62. <http://doi.org/10.5033/ifosslr.v4i1.69>

Kock, N. F. (2007). *Encyclopedia of E-collaboration*, 1–751.

- Langville, A. N., & Meyer, C. D. (2011). *Google's PageRank and Beyond*. Princeton: Princeton University Press.
- Leydesdorff, L. (2012). *The Triple Helix of University-Industry-Government Relations* (February 2012). <http://doi.org/10.2139/ssrn.1996760>
- Longo, J., & Kelley, T. M. (2015). Use of GitHub as a platform for open collaboration on text documents (pp. 1–2). Presented at the the 11th International Symposium, New York, New York, USA: ACM Press. <http://doi.org/10.1145/2788993.2789838>
- Lowry, D. D., & Lowry, M. R. (1995). Legal issues in knowledge-based software engineering. 10th Knowledge-Based Software Engineering Conference, 61–69. <http://doi.org/10.1109/KBSE.1995.490120>
- Manteli, C., van Vliet, H., & Hooff, B. V. D. (2012). Adopting a Social Network Perspective in Global Software Development (pp. 124–133). Presented at the 2012 IEEE Seventh International Conference on Global Software Engineering, IEEE. <http://doi.org/10.1109/icgse.2012.10>
- Martínez, S. L. (n.d.). Bienes Comunes y Acceso Abierto: nuevas formas de acceder, generar contenidos y compartir obras en la arquitectura de Internet. *Acta científica. Servicioit. Cl.*
- Mendez J.M., Monroy F. A., & Zorrilla S. (1993). *Dinámica social de las organizaciones*. MC GRAW HILL.
- Moonan, S. (2013). *Technology Transfer*. Routledge.
- Metz, S. (2012). *Practical Object-Oriented Design in Ruby*. Addison-Wesley.
- Mortara, L., Minshall, T., Napp, J. J. (2010). *How to Implement Open Innovation*.
- Mortara, L., Napp, J. J., & Slacik, I. (2009). *How to implement open innovation: Lessons from studying large multinational companies. ... Of Cambridge*.

- NeoChange (2012). Recuperado en Septiembre 13, 2016, de <http://sandhill.com/article/2012-it-adoption-insight-survey-reveals-disturbing-level-of-enterprise-productivity-loss/>
- Nyman, L., & Lindman, J. (2013). Code forking, governance, and sustainability in open source software. *Technology Innovation Management* ....  
<http://doi.org/10.1080/13549830601183289>
- O'Connell, A. A. (2006). *Logistic Regression Models for Ordinal Response Variables*. SAGE.
- OECD. (2008). *Open Innovation in Global Networks*. OECD Publishing.  
<http://doi.org/10.1787/9789264047693-en>
- OECD, & Communities, S. O. O. T. E. (2014). *Oslo Manual Guidelines for Collecting and Interpreting Innovation Data, 3rd Edition (Chinese version)*. OECD Publishing.  
<http://doi.org/10.1787/9789264213081-zh>
- Olsen, R. (2010). *OmniGraffle 5 Diagramming Essentials*. Birmingham: Packt Publ.
- Perkel, J. (2016). Democratic databases: science on GitHub. *Nature*, 538(7623), 127–128.  
<http://doi.org/10.1038/538127a>
- Pfleeger, S. L. (1998). The nature of system change [software]. *IEEE Software*, 15(3), 87–90.  
<http://doi.org/10.1109/52.676964>
- Pipinellis, A. (2015). *GitHub Essentials*. Packt Publishing Ltd.
- Punter, T., Krikhaar, R. L., & Bril, R. J. (2006). *Sustainable Technology Transfer* (pp. 15–18). New York, New York, USA: ACM. <http://doi.org/10.1145/1138046.1138052>
- Redwine, S. T., Jr, & Riddle, W. E. (1985). Software Technology Maturation. *Icse*, 189–200.
- Rifkin, J. (2014). *The Zero Marginal Cost Society*. St. Martin's Press.
- Robinson, I., Webber, J., & Eifrem, E. (2015). *Graph Databases*. “O'Reilly Media, Inc..”
- Roebuck, K. (2011). *MapReduce: High-impact Strategies - What You Need to Know*.

- Rowley, J. (2007). The wisdom hierarchy - representations of the DIKW hierarchy. *J. Information Science*, 33(2), 163–180.
- Sassen, S. (2000). Digital Networks and the State Some Governance Questions. *Theory, Culture & Society*, 17(4), 19–33. <http://doi.org/10.1177/02632760022051293>
- Schuh, G., & Aghassi, S. (2013). Technology transfer portals: A design model for supporting technology transfer via social software solutions, 43–47. <http://doi.org/10.1109/IEEM.2013.6962371>
- Schumpeter, J. (1947/2012). *Capitalism, Socialism, and Democracy*. Simon and Schuster. <http://doi.org/10.4324/9780203857090>
- Shmueli, E., Altshuler, Y., & Pentland, A. (2014). Temporal Dynamics of Scale-Free Networks. *Sbp*, 8393(Chapter 44), 359–366. [http://doi.org/10.1007/978-3-319-05579-4\\_44](http://doi.org/10.1007/978-3-319-05579-4_44)
- Solinger, M., Engberts, A., & Ning, J. Q. (1994). Transferring re-engineering technology to a software development and maintenance organization: an experience report. *Proceedings International Conference on Software Maintenance* (pp. 100–108). IEEE. <http://doi.org/10.1109/ICSM.1994.336785>
- Solleiro, J. L., & Teran-Bustamante, A. (2012). Buenas prácticas de gestión de la innovación en centros de investigación tecnológica (pp. 1–75). Instituto de Investigaciones Eléctricas y Universidad Nacional Autónoma de México.
- Stallman, R. M. (2007). *Software libre para una sociedad libre*.
- Thung, F., Bissyande, T. F., Lo, D., & Lingxiao Jiang. (2016a). Network Structure of Social Coding in GitHub (pp. 323–326). Presented at the 2013 17th European Conference on Software Maintenance and Reengineering (CSMR 2013), IEEE. <http://doi.org/10.1109/CSMR.2013.41>

- Thung, F., Bissyandé, T. F., Lo, D., & Jiang, L. (2016b). Network Structure of Social Coding in GitHub. *2013 17th European Conference on Software Maintenance and Reengineering (CSMR 2013)*, 323–326. <http://doi.org/10.1109/CSMR.2013.41>
- Vanhaverbeke, W. (2009). Business Models in Open Innovation and Commercialization – a dynamic approach. In *Innovationsführerschaft durch Open Innovation* (pp. 147–154). Berlin, Heidelberg: Springer, Berlin, Heidelberg. [http://doi.org/10.1007/978-3-540-87755-4\\_11](http://doi.org/10.1007/978-3-540-87755-4_11)
- Weisberg, S. (2013). *Applied Linear Regression*. John Wiley & Sons.
- West, J., & Gallagher, S. (2006). Patterns of open innovation in open source software. *Open Innovation: Researching a New Paradigm*.
- Wilms, T. (2012, Enero 10). SAPVoice: [Infographic]: The Evolution of Technology & Its Impact on the Development of Social Businesses. Recuperado en Agosto 11, 2015, de <https://www.forbes.com/sites/sap/2012/01/10/infographic-the-evolution-of-technology-its-impact-on-the-development-of-social-businesses/#6bf5be9466f9>
- WIPO. (2015). *The Global Innovation Index 2015* (pp. 1–453). Presented at the Effective Innovation Policies for Development.
- Yu, Y., Yin, G., Wang, H., & Wang, T. (2014). Exploring the patterns of social behavior in GitHub. *CrowdSoft@SIGSOFT FSE*, 31–36. <http://doi.org/10.1145/2666539.2666571>
- Zafarani, R., Abbasi, M. A., & Liu, H. (2009). *Social Media Mining*. Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9781139088510>
- Zelkowitz, M. V., Wallace, D. R., & Binkley, D. (1998). Culture conflicts in software engineering technology transfer. *NASA Goddard Software ....*

**ANEXOS**

## ANEXO A

### FUNCIÓN DE MAPEO

La función que se utilizó para extraer los flujos de innovación fue registrada con el nombre “MailMapper” en el repositorio: <https://rubygems.org/profiles/os Caryunue>; la siguiente figura proporciona información sobre su versión, dependencias, y el tipo de licencia.

The screenshot shows the RubyGems page for the `mailMapper` gem, version `0.1.2`. The page includes a description of the gem's functionality, version information, dependencies, and installation instructions.

**mailMapper 0.1.2**

Search and count emails from a text file, Ex. `MailMapper.find("source", option=int);` option 0 «default value» returns a hash with domain-key counter-value, option 1 returns a hash for `user@mail -> counter-value`, option 2 returns an array with all mails

**VERSIONES:**

- 0.1.2 - April 07, 2016 (20 KB)
- 0.1.0 - April 07, 2016 (9 KB)

**DEPENDENCIAS DE DEVELOPMENT (3):**

- `bundler` -> 1.11
- `rake` -> 10.0
- `rspec` -> 3.0

**AUTORES:**

- os Caryunue

**ADMINISTRADORES:**

- [Profile Icon]

**TOTAL DOWNLOADS**  
**1.010**

**PARA ESTA VERSIÓN**  
**626**

**GEMFILE:**

```
gem 'mailMapper',
```

**INSTALAR:**

```
gem install mailMa
```

**LICENCIA:**  
**MIT**

**REQUIRED RUBY VERSION:**  
**>= 0**

Esta función posee tres funcionalidades:

- La opción “0”. Recolecta a todos los colaboradores y los categoriza por tipo de dominio.
- La opción “1”. Extrae a todos los colaboradores.
- La opción “2”. Proporciona un contador de registros de trabajo por colaborador.

En este trabajo se utilizó la opción “0”; las otras dos opciones se indican en el parámetro “*option*” al ejecutar la función. Por ejemplo, “*MailMapper.find("EI\_ARCHIVO", option=ElNúmeroDeOpción)*”.

## ANEXO B

## FUNCIÓN DE REDUCCIÓN

La siguiente captura de pantalla ilustra un fragmento del código fuente utilizado en la función de reducción de datos.

```

1 # Inputs: sector data
2 reduceFunction <- function(commits =n, merges = c) {
3   length <- length(commits)
4   t_df <- data.frame(1:length)
5   for (i in 1:length) {
6     lActors <- n[intersect(names(n), names(m))]
7     eActors <- m[intersect(names(n), names(m))]
8     owner <- max(unlist(n))
9     mowner <- max(unlist(m))
10    cttotal <- sum(unlist(n))
11    mttotal <- sum(unlist(m))
12    complement <- cttotal - owner
13    mcomplement <- mttotal - mowner
14    t_df[i, 1] <- owner; t_df[i, 2] <- cttotal; t_df[i, 3] <- sum(lactors); t_df[i, 4] <- complement
15    t_df[i, 5] <- sum(lactors + complement); t_df[i, 6] <- eActors
16    t_df[i, 7] <- count(has.transactions(lActors) + count(has.transactions(eActors)))
17    t_df[i, 8] <- sum(lActors.transactions); t_df[i, 9] <- sum(eActors.transactions)
18    t_df[i, 10] <- if((unlist(lActors)) > each(eActors)) { 1 } else {0} # external proportion
19  }
20  colnames(t_df) <- c("ID", "totalColaborators",
21                    "totalWork", "totalTransactions", "localWork", "externalWork", "transactionNodes", "localTRansactions",
22                    "externalTRansactions", "H0")
23  return(t_df)
24 }

```

*Fuente propia. Capturado de RStudio.*

De igual manera, en el siguiente enlace se suministra el código fuente y al espacio de trabajo utilizado en esta investigación:

- Código fuente: <https://github.com/os Caryunue/redF>
- Licencia: MIT

## ANEXO C

### MUESTRAS SELECCIONADAS

#### Sector privado

| Autor          | Repositorio   | Descripción  | Licencia |
|----------------|---------------|--|----------|
| RH, Inc.       | ansible       | <i>Ansible</i> es una plataforma de automatización de IT | GLP3     |
| Amazon Inc.    | aws-cli       | Interfaz de línea de comandos universal para AWS         | Apache2  |
| Twitter Inc.   | bootstrap     | Complementos JavaScript para móviles en la web           | MIT      |
| Adobe Sys.     | brackets      | Es un editor de código fuente abierto para la web        | MIT      |
| Harvest        | chosen        | Librería para facilitar consultas en servicios web       | MIT      |
| Uber T. Inc.   | deck.gl       | Capas de visualización basadas en WebGL                  | MIT      |
| Yahoo!         | elide         | Biblioteca de Java para servicios web JSON API           | Apache2  |
| GitHub Inc.    | gitignore     | Colección de plantillas para grupos de trabajo           | CC       |
| LinkedInCorp.  | gobblin       | Plataforma de integración de datos                       | Apache2  |
| Facebook, Inc. | hhvm          | Ejecuta programas escritos en Hack y PHP                 | BSD      |
| Netflix        | Hystrix       | Biblioteca diseñada para el acceso a sistemas remotos    | Apache2  |
| Spotify        | luigi         | Módulo de Python para la gestión de macro datos          | Apache2  |
| Meteor DG      | meteor        | Plataforma de aplicaciones JavaScript.                   | MIT      |
| Nylas          | nylas-mail    | Es una aplicación de correo de escritorio extensible.    | GPL3     |
| Thoughtbot     | paperclip.git | Código fuente para la actualización de BD dinámicas      | Apache2  |
| Android        | frameworks    | Código fuente del sistema operativo Android              | MIT      |
| Apple          | react.git     | Es una biblioteca para crear interfaces de usuario       | BSD3     |
| Yelp.com       | elastalert    | Modulo para la integración de servicios web              | Apache2  |
| The Guardian   | frontend      | Es una fuente del portal theguardian.com                 | Apache2  |
| Intel SDI      | snap          | Es un marco abierto de telemetría                        | Apache2  |
| Samsung        | GearVRf       | Código fuente para aplicaciones de realidad virtual      | Apache2  |
| Valve Soft.    | steamos_mesa  | Código fuente del sistema operativo <i>Steam</i>         | BSD3     |
| Microsoft      | vscode        | Código fuente de <i>Visual Studio</i>                    | MIT      |
| Heroku Soft.   | legacy-cli    | Código fuente para la integración de servicios web       | MIT      |

#### Sector social

| Autor            | Repositorio       | Descripción   | Licencia   |
|------------------|-------------------|---|------------|
| alamofire.org    | AFNetworking      | Es una estructura de red para iOS, OS X, watchOS y tvOS.      | MIT        |
| atom.io          | atom              | Es un editor de textos  | MIT        |
| vinta.chen       | awesome-python    | Es una lista recursos para el lenguaje Python                 | CC         |
| bitcoin.org      | bitcoin           | Árbol de integración / estatificación de <i>Bitcoin</i> Core. | MIT        |
| d3js.org         | d3                | Trae datos a la vida con SVG, lienzo y HTML.                  | BSD3       |
| electron.atom.io | electron          | Crea aplicaciones multiplataforma de escritorio               | MIT        |
| expressjs.com    | express           | Marco web minimalista para el nodo.                           | MIT        |
| Mozilla F.       | firefox-ios       | Firefox para iOS  | MozillaPL2 |
| fontawesome.io   | Font-Awesome      | La fuente icónica y el juego de herramientas CSS.             | SIL        |
| Free Ebook       | free-prog-books   | Libros de programación disponibles libremente.                | CC         |
| freeCodeCamp     | freeCodeCamp      | Es un código fuente abierto y currículo.                      | BSD 3      |
| golang.org       | go                | El lenguaje de programación Go                                | BSD 3      |
| Apache SF        | hadoop            | Espejo de Apache Hadoop p/ gestión de macro datos             | Apache2    |
| H5BP             | html5-boilerplate | Es una plantilla de aplicaciones o sitios web adaptables.     | MIT        |
| Airbnb           | javascript        | Guía de estilo JavaScript                                     | MIT        |
| Jekyll           | jekyll            | Es un generador de sitio estático y blog-aware en Ruby        | MIT        |
| jQuery           | jquery            | jQuery JavaScript Library                                     | MIT        |
| Moby O.F.        | moby              | Es un proyecto colaborativo para el ecosistema                | Apache2    |
| Node.js Found.   | node              | Node.js Tiempo de ejecución de JavaScript                     | ICU        |
| rubyonrails.org  | rails             | Ruby on Rails   | MIT        |
| hakim.elhatab    | reveal.js         | El marco de presentación HTML.                                | MIT        |
| Semantic Org.    | Semantic-UI       | Es un marco de componentes de la interfaz de usuario.         | MIT        |
| Socket.IO        | socket.io         | Marco de aplicaciones en tiempo real                          | MIT        |
| Symfony          | symfony           | El framework PHP de Symfony.                                  | MIT        |
| fastlane.tools   | fastlane          | Forma fácil publicación de app para iOS y Android             | MIT        |

**Sector de administración**

| Autor                | Repositorio       | Descripción  | Licencia |
|----------------------|-------------------|--|----------|
| Almofire Org.        | Alamofire         | Las políticas oficiales de Fundación Alamofire.            | MIT      |
| digipalvelutehdas.fi | APIKA             | Es una arquitectura nacional de servicios en Finlandia     | MIT      |
| Francia Gov.         | beta.gouv.fr      | Sistema del Ministerio de Educación y Cultura              | CC       |
| CKAN Org             | ckan              | Fuente abierta bancos de datos                             | AGLP3    |
| NYC.gov              | NYCOpenRecords    | Es una aplicación web para presentar solicitudes           | Apache2  |
| Ayu. Barcelona       | decidim.barcelona | Es una plataforma web de democracia directa                | AGLP3    |
| Belgium              | eid-mw            | Integración de sistemas federales en Bélgica               | GLP3     |
| FR                   | mdph              | Sistema de indemnizaciones en Francia                      | AGLP3    |
| FR                   | mes-aides-ui      | Simulador general beneficios sociales                      | AGLP3    |
| openaddresses.io     | openaddresses     | Es un repositorio de datos globales                        | BSD3     |
| OpenLayers           | openlayers        | OpenLayers es una librería cartográfica                    | BSD      |
| Belgium              | openpolice        | Sistema web de código abierto para las fuerzas policiales. | AGLP3    |
| Australian Tax. Off. | RAM               | Sistema de Autorización de Impuestos en Australia          | MIT      |
| Interlegis           | sapl              | Sistema de regulación en Brasil                            | ARR      |
| SNIESE(Ecuador)      | sniese-ui         | Cliente web de la Secretaria de Educación Superior         | ARR      |
| Swagger              | swagger-ui        | Generador dinámico de documentación                        | Apache2  |
| Taiga.io             | taiga-front       | Contruye aplicaciones de gestión                           | AGLP3    |
| GOV.UK               | whitehall         | Publica contenido gubernamental en GOV.UK                  | MIT      |
| OpenGov              | madison           | Plataforma para la legislación ciudadana                   | AGLP3    |
| NY Senate OS         | OpenLegislation   | Servicio web de información legislativa                    | GLP3     |
| C for America        | adopt-a-hydrant   | Sistema de Participación ciudadana                         | BSD 3.0  |
| HospitalRun          | frontend          | Aplicación cliente para <i>HospitalRun</i>                 | GLP3     |
| dados.gov.br         | kit               | Plan de datos abiertos                                     | CC       |
| Refuge Restrooms     | refugerestrooms   | Alberga índices y mapas de ubicaciones de baños            | AGLP3    |
| UK Government DS     | e-petitions       | Servicio de peticiones electrónicas en Inglaterra          | MIT      |

**Sector de investigación**

| Autor  | Repositorio      | Descripción  | Licencia |
|--|------------------|--|----------|
| COMBINE-Australia                                  | Brisbane         | Herramientas de investigación para bioinformática y campos relacionados. | CC       |
| National Institute of Standards and Tech.          | 800-63-3         | Directrices para implementar servicios de identidad digital              | MIT      |
| Art Institute of Chicago                           | lakeshore        | Sistema de gestión para colecciones de arte                              | CC       |
| Harvard University                                 | economic-complex | Herramienta para visualizar transacciones económicas a escala global     | Apache2  |
| Berkeley Vision and Learning Center                | caffe            | Herramienta de análisis para aprendizaje profundo                        | ARR      |
| British Columbia IOT                               | CodeIgniter      | Repositorio de herramientas de análisis                                  | BSD3     |
| edX  | edx-platform     | Plataforma de aprendizaje en línea fundada por Harvard/MIT               | AGLP3    |
| National Renewable Energy Laboratory               | EnergyPlus       | Es un programa de simulación de energía para modelar el consumo de agua  | BSD3     |
| Masaryk University                                 | munimap          | Plataforma para mostrar mapas interactivos de la Universidad Masaryk     | Apache2  |
| Institute of Medical Genetics and Applied Genomics | ngs-bits         | Herramientas para analizar transmisión de datos                          | GLP2     |
| NASA Tensegrity Robotics Toolkit                   | NTRTsim          | Simulador basado en la física para el diseño y control de los robots     | Apache2  |
| National Climate Prediction Portal                 | ocgis            | Herramientas de geo-procesamiento y cálculo de datos climáticos          | ARR      |
| North Rhine-Westphalian                            | oerworldmap      | Repositorio para recursos globales de educación abierta                  | Apache2  |
| National Library of Sweden                         | openapc-se       | Sistema para publicaciones de acceso abierto                             | Apache2  |
| ARM Climate Research Facility                      | pyart            | Colección de algoritmos y utilidades de radar meteorológico              | ARR      |
| Materials Project                                  | pymatgen         | Sistema de análisis para representar estructuras y moléculas             | MIT      |
| Institute of Neuroinformatics Zurich               | pyNCS            | Herramientas para modelar Sistemas Neuromórficos                         | GLP2     |
| Maryland Institute                                 | sga              | Iniciativa para reconstruir obras de literatura antiguas                 | ARR      |
| Lawrence Livermore National Laboratory             | spack            | Gestor de herramientas para sistemas de investigación                    | GLP2     |
| NCBI - National Center for Biotech.                | sra-tools        | Herramientas de análisis del Centro Nacional de Biotecnología            | ARR      |
| LIBRIS   | swepubanalys     | Interfaz web para sistemas de bibliometría                               | MIT      |
| tensorflow.org                                     | tensorflow       | Herramientas de aprendizaje automático                                   | Apache2  |
| University of Lübeck                               | testbed-runtime  | Sistema de análisis para sensores inalámbricos                           | BSD      |
| chainer.org  | chainer          | Herramientas para diseñar redes neuronales                               | MIT      |
| Zenodo Org.  | zenodo           | Plataforma de servicio para alojar bancos de datos científicos           | GLP2     |

**ANEXO D**

## BASE DE DATOS

### Sector privado

| Identificador     | Total de Colaboradores | Trabajo | Transacciones | Trabajo Local | Trabajo Externo | Nodos de transacción | Transacciones Locales | Transacciones Externas | H0 |
|-------------------|------------------------|---------|---------------|---------------|-----------------|----------------------|-----------------------|------------------------|----|
| s1_c_ansible      | 274                    | 29374   | 5427          | 11550         | 17824           | 38                   | 3033                  | 2394                   | 1  |
| s1_c_aws_cli      | 350                    | 4755    | 1531          | 2107          | 2648            | 8                    | 627                   | 904                    | 1  |
| s1_c_bootstrap    | 241                    | 15398   | 4000          | 6493          | 8905            | 28                   | 2065                  | 1935                   | 0  |
| s1_c_brackets     | 446                    | 16974   | 5729          | 8737          | 8237            | 31                   | 3425                  | 2304                   | 1  |
| s1_c_chosen       | 119                    | 1177    | 311           | 553           | 624             | 9                    | 193                   | 118                    | 1  |
| s1_c_deck.gl      | 605                    | 781     | 33            | 439           | 342             | 3                    | 15                    | 18                     | 1  |
| s1_c_elide        | 209                    | 1107    | 272           | 233           | 874             | 3                    | 175                   | 97                     | 1  |
| s1_c_gitignore    | 89                     | 2369    | 1059          | 639           | 1730            | 26                   | 438                   | 621                    | 1  |
| s1_c_goblin       | 82                     | 3221    | 1210          | 1276          | 1945            | 11                   | 412                   | 798                    | 1  |
| s1_c_hhvm         | 65                     | 1541    | 27            | 893           | 648             | 5                    | 19                    | 8                      | 1  |
| s1_c_hystrix      | 603                    | 1997    | 621           | 1003          | 994             | 6                    | 597                   | 24                     | 1  |
| s1_c_luigi        | 818                    | 3290    | 1140          | 1407          | 1883            | 11                   | 401                   | 739                    | 1  |
| s1_c_meteor       | 99                     | 18626   | 1602          | 9896          | 8730            | 29                   | 733                   | 869                    | 1  |
| s1_c_nylas_mail   | 214                    | 5818    | 156           | 3584          | 2234            | 6                    | 107                   | 49                     | 1  |
| s1_c_paperclip    | 339                    | 1688    | 186           | 693           | 995             | 14                   | 104                   | 82                     | 1  |
| s1_c_framework    | 709                    | 215512  | 155866        | 191111        | 24401           | 30                   | 140731                | 15135                  | 1  |
| s1_c_react        | 262                    | 7992    | 2403          | 2243          | 5749            | 29                   | 962                   | 1441                   | 1  |
| s1_c_elastalert   | 113                    | 1251    | 415           | 899           | 352             | 12                   | 386                   | 29                     | 1  |
| s1_c_frontend     | 151                    | 71323   | 21460         | 33082         | 38241           | 34                   | 7462                  | 13998                  | 1  |
| s1_c_snap         | 66                     | 7992    | 878           | 1970          | 6022            | 7                    | 707                   | 171                    | 1  |
| s1_c_GearCRF      | 208                    | 2329    | 656           | 2062          | 267             | 4                    | 644                   | 12                     | 1  |
| s1_c_steamos_mesa | 109                    | 78938   | 736           | 18865         | 60073           | 31                   | 283                   | 453                    | 1  |
| s1_c_vscode       | 334                    | 19575   | 1436          | 14882         | 4693            | 16                   | 805                   | 631                    | 1  |
| s1_c_legacy_cli   | 142                    | 4283    | 664           | 2204          | 2079            | 15                   | 334                   | 330                    | 1  |

### Sector social

| Identificador      | Total de Colaboradores | Trabajo | Transacciones | Trabajo Local | Trabajo Externo | Nodos de transacción | Transacciones Locales | Transacciones externas | H0 |
|--------------------|------------------------|---------|---------------|---------------|-----------------|----------------------|-----------------------|------------------------|----|
| s2_c_AFNetworking  | 240                    | 198200  | 18302         | 114301        | 83899           | 24                   | 12407                 | 5895                   | 1  |
| s2_c_atom          | 274                    | 2783    | 701           | 1555          | 1228            | 10                   | 445                   | 256                    | 1  |
| s2_c_awesomepython | 350                    | 29374   | 5427          | 11550         | 17824           | 38                   | 3033                  | 2394                   | 1  |
| s2_c_bitcoin       | 241                    | 1065    | 380           | 855           | 210             | 10                   | 370                   | 10                     | 1  |
| s2_c_d3            | 446                    | 11918   | 4842          | 8411          | 3507            | 14                   | 4365                  | 477                    | 1  |
| s2_c_electron      | 119                    | 4084    | 684           | 3147          | 937             | 6                    | 659                   | 25                     | 1  |
| s2_c_express       | 605                    | 14303   | 3148          | 11393         | 2910            | 25                   | 2166                  | 982                    | 1  |
| s2_c_firefox_ios   | 209                    | 1466    | 239           | 887           | 579             | 47                   | 129                   | 110                    | 1  |
| s2_c_font_Awesome  | 89                     | 5462    | 1498          | 2443          | 3019            | 13                   | 644                   | 854                    | 1  |
| s2_c_free_books    | 82                     | 1142    | 151           | 754           | 388             | 5                    | 65                    | 86                     | 1  |
| s2_c_freeCodeCamp  | 883                    | 3691    | 1479          | 2725          | 966             | 27                   | 1424                  | 55                     | 1  |
| s2_c_go            | 603                    | 9432    | 2964          | 5659          | 3773            | 27                   | 1461                  | 1503                   | 1  |
| s2_c_hadoop        | 818                    | 32872   | 79            | 25541         | 7331            | 2                    | 63                    | 16                     | 1  |
| s2_c_boilerplate   | 99                     | 12425   | 396           | 11776         | 649             | 4                    | 380                   | 16                     | 1  |
| s2_c_javascript    | 214                    | 1514    | 197           | 1162          | 352             | 10                   | 162                   | 35                     | 1  |
| s2_c_jekyll        | 339                    | 1356    | 295           | 721           | 635             | 5                    | 115                   | 180                    | 1  |
| s2_c_jquery        | 709                    | 8676    | 1957          | 4658          | 4018            | 27                   | 1002                  | 955                    | 1  |
| s2_c_moby          | 262                    | 6148    | 251           | 5345          | 803             | 7                    | 204                   | 47                     | 1  |
| s2_c_node          | 1688                   | 32740   | 13112         | 8174          | 24566           | 88                   | 3770                  | 9342                   | 1  |
| s2_c_rails         | 1389                   | 17830   | 349           | 8216          | 9614            | 8                    | 136                   | 213                    | 1  |
| s2_c_reveal        | 3307                   | 63039   | 12773         | 35793         | 27246           | 128                  | 9102                  | 3671                   | 1  |
| s2_c_semantic_UI   | 208                    | 1952    | 374           | 1672          | 280             | 7                    | 351                   | 23                     | 1  |
| s2_c_socket_io     | 109                    | 6189    | 544           | 5478          | 711             | 2                    | 512                   | 32                     | 1  |
| s2_c_symfony       | 140                    | 1446    | 282           | 1321          | 125             | 5                    | 268                   | 14                     | 1  |
| s2_c_fastlane      | 1454                   | 29951   | 12241         | 23732         | 6219            | 51                   | 11971                 | 270                    | 1  |

**Sector de administración**

| Identificador            | Total de Colaboradores | Trabajo | Transacciones | Trabajo Local | Trabajo Externo | Nodos de transacción | Transacciones Locales | Transacciones externas | H0 |
|--------------------------|------------------------|---------|---------------|---------------|-----------------|----------------------|-----------------------|------------------------|----|
| s3_c_alamofire           | 665                    | 12862   | 2464          | 7923          | 4939            | 38                   | 1246                  | 1218                   | 1  |
| s3_c_APIKA               | 137                    | 1166    | 168           | 761           | 405             | 6                    | 104                   | 64                     | 0  |
| s3_c_beta.gouv.fr        | 21                     | 6383    | 938           | 2320          | 4063            | 12                   | 299                   | 639                    | 1  |
| s3_c_ckan                | 73                     | 2219    | 638           | 498           | 1721            | 11                   | 194                   | 444                    | 0  |
| s3_c_DACOpenRec          | 160                    | 15828   | 3322          | 6461          | 9367            | 27                   | 1584                  | 1738                   | 0  |
| s3_c_debarcelon          | 7                      | 1605    | 378           | 843           | 762             | 5                    | 176                   | 202                    | 1  |
| s3_c_eid_mw              | 33                     | 3913    | 944           | 3412          | 501             | 7                    | 871                   | 73                     | 0  |
| s3_c_mdph                | 10                     | 2656    | 132           | 2577          | 79              | 3                    | 126                   | 6                      | 1  |
| s3_c_mes_aides_ui        | 10                     | 2620    | 256           | 1801          | 819             | 5                    | 168                   | 88                     | 1  |
| s3_c_openaddresses       | 14                     | 2745    | 256           | 1498          | 1247            | 5                    | 113                   | 143                    | 0  |
| s3_c_openlayers          | 113                    | 5980    | 1646          | 3353          | 2627            | 16                   | 702                   | 944                    | 0  |
| s3_c_openpolice_platfórm | 108                    | 6456    | 508           | 2351          | 4105            | 11                   | 239                   | 269                    | 0  |
| s3_c_RAM                 | 9                      | 3307    | 729           | 3296          | 11              | 2                    | 728                   | 1                      | 1  |
| s3_c_sapl                | 12                     | 3237    | 849           | 1721          | 1516            | 6                    | 279                   | 570                    | 1  |
| s3_c_snieste_ui          | 13                     | 2872    | 261           | 1747          | 1125            | 5                    | 158                   | 103                    | 1  |
| s3_c_swagger_ui          | 221                    | 1857    | 52            | 987           | 870             | 4                    | 32                    | 20                     | 0  |
| s3_c_taiga_front         | 205                    | 2136    | 661           | 1249          | 887             | 17                   | 247                   | 414                    | 0  |
| s3_c_whitehall           | 32                     | 4918    | 576           | 3445          | 1473            | 5                    | 303                   | 273                    | 1  |
| s3_c_madison             | 135                    | 16563   | 3264          | 4779          | 11784           | 49                   | 816                   | 2448                   | 1  |
| s3_c_OpenLegislation     | 21                     | 2168    | 314           | 1932          | 236             | 7                    | 280                   | 34                     | 0  |
| s3_c_adopt_a_hydrant     | 16                     | 2025    | 278           | 1661          | 364             | 5                    | 263                   | 15                     | 0  |
| s3_hospitalrun_frontend  | 16                     | 1709    | 12            | 1691          | 18              | 2                    | 12                    | 0                      | 0  |
| s3_c_kit                 | 86                     | 2398    | 221           | 1749          | 649             | 5                    | 162                   | 59                     | 1  |
| s3_c_refugereestrooms    | 5                      | 335     | 30            | 186           | 149             | 2                    | 18                    | 12                     | 1  |
| s3_c_e_petitions         | 49                     | 863     | 215           | 565           | 298             | 11                   | 160                   | 55                     | 0  |

**Sector de investigación**

| Identificador                     | Total de Colaboradores | Trabajo | Transacciones | Trabajo Local | Trabajo Externo | Nodos de transacción | Transacciones Locales | Transacciones externas | H0 |
|-----------------------------------|------------------------|---------|---------------|---------------|-----------------|----------------------|-----------------------|------------------------|----|
| s4_c_2017_06_28_R_Brisbane        | 137                    | 1166    | 168           | 761           | 405             | 6                    | 104                   | 64                     | 0  |
| s4_c_800_63_3                     | 21                     | 6383    | 938           | 2320          | 4063            | 12                   | 299                   | 639                    | 1  |
| s4_c_aicdams_lakeshore            | 73                     | 2219    | 638           | 498           | 1721            | 11                   | 194                   | 444                    | 0  |
| s4_c_atlas_economic_complexity    | 160                    | 15828   | 3322          | 6461          | 9367            | 27                   | 1584                  | 1738                   | 0  |
| s4_c_caffe                        | 7                      | 1605    | 378           | 843           | 762             | 5                    | 176                   | 202                    | 1  |
| s4_c_ckan                         | 33                     | 3913    | 944           | 3412          | 501             | 7                    | 871                   | 73                     | 0  |
| s4_c_codeigniter                  | 10                     | 2656    | 132           | 2577          | 79              | 3                    | 126                   | 6                      | 1  |
| s4_c_commit_top_github_users_data | 10                     | 2620    | 256           | 1801          | 819             | 5                    | 168                   | 88                     | 1  |
| s4_c_edx_platform                 | 14                     | 2745    | 256           | 1498          | 1247            | 5                    | 113                   | 143                    | 0  |
| s4_c_energyPlus                   | 113                    | 5980    | 1646          | 3353          | 2627            | 16                   | 702                   | 944                    | 0  |
| s4_c_libmesh                      | 108                    | 6456    | 508           | 2351          | 4105            | 11                   | 239                   | 269                    | 0  |
| s4_c_munimap                      | 9                      | 3307    | 729           | 3296          | 11              | 2                    | 728                   | 1                      | 1  |
| s4_c_nginx_bits                   | 12                     | 3237    | 849           | 1721          | 1516            | 6                    | 279                   | 570                    | 1  |
| s4_c_NRTsim                       | 13                     | 2872    | 261           | 1747          | 1125            | 5                    | 158                   | 103                    | 1  |
| s4_c_ocgis                        | 221                    | 1857    | 52            | 987           | 870             | 4                    | 32                    | 20                     | 0  |
| s4_c_oerworldmap                  | 205                    | 2136    | 661           | 1249          | 887             | 17                   | 247                   | 414                    | 0  |
| s4_c_openapc_se                   | 32                     | 4918    | 576           | 3445          | 1473            | 5                    | 303                   | 273                    | 1  |
| s4_c_pyart                        | 135                    | 16563   | 3264          | 4779          | 11784           | 49                   | 816                   | 2448                   | 1  |
| s4_c_pymatgen                     | 21                     | 2168    | 314           | 1932          | 236             | 7                    | 280                   | 34                     | 0  |
| s4_c_pyNCS                        | 16                     | 2025    | 278           | 1661          | 364             | 5                    | 263                   | 15                     | 0  |
| s4_c_sga                          | 16                     | 1709    | 12            | 1691          | 18              | 2                    | 12                    | 0                      | 0  |
| s4_c_spack                        | 86                     | 2398    | 221           | 1749          | 649             | 5                    | 162                   | 59                     | 1  |
| s4_c_sra_tools                    | 5                      | 335     | 30            | 186           | 149             | 2                    | 18                    | 12                     | 1  |
| s4_c_swepubanalysis               | 49                     | 863     | 215           | 565           | 298             | 11                   | 160                   | 55                     | 0  |
| s4_c_tensorflow                   | 29                     | 1730    | 467           | 367           | 1363            | 9                    | 169                   | 298                    | 1  |

## ANEXO E

### RESULTADOS DEL AJUSTE LINEAL

#### Ajuste lineal para el sector Privado

Escala de rango[Transacciones locales] = 0.1198298 + 0.258767\*Escala de rango[Trabajo externo]

#### Resumen del ajuste

|                                 |          |
|---------------------------------|----------|
| R cuadrado                      | 0.043859 |
| R cuadrado ajustado             | 0.002288 |
| Raíz del error cuadrático medio | 0.245416 |
| Media de respuesta              | 0.132827 |
| Observaciones (o suma de pesos) | 25       |

#### Análisis de varianza

| Fuente   | Grados de libertad | Suma de cuadrados | Media de los cuadrados | Razón F            |
|----------|--------------------|-------------------|------------------------|--------------------|
| Modelo   | 1                  | 0.0635438         | 0.063544               | 1.0550             |
| Error    | 23                 | 1.3852637         | 0.060229               | <b>Prob &gt; F</b> |
| C. Total | 24                 | 1.4488075         |                        | 0.3150             |

#### Estimaciones de los parámetros

| Término                          | Estimación | Error estándar | Razón t | Prob >  t |
|----------------------------------|------------|----------------|---------|-----------|
| Constante del modelo             | 0.1198298  | 0.050688       | 2.36    | 0.0269*   |
| Escala de rango[Trabajo externo] | 0.258767   | 0.251927       | 1.03    | 0.3150    |

#### Ajuste lineal

Escala de rango[Transacciones locales] = 0.0853615 + 0.6392166\*Escala de rango[Trabajo externo]

#### Resumen del ajuste

|                                 |          |
|---------------------------------|----------|
| R cuadrado                      | 0.32428  |
| R cuadrado ajustado             | 0.294901 |
| Raíz del error cuadrático medio | 0.230968 |
| Media de respuesta              | 0.176548 |
| Observaciones (o suma de pesos) | 25       |

#### Análisis de varianza

| Fuente   | Grados de libertad | Suma de cuadrados | Media de los cuadrados | Razón F            |
|----------|--------------------|-------------------|------------------------|--------------------|
| Modelo   | 1                  | 0.5888208         | 0.588821               | 11.0378            |
| Error    | 23                 | 1.2269593         | 0.053346               | <b>Prob &gt; F</b> |
| C. Total | 24                 | 1.8157801         |                        | 0.0030*            |

#### Estimaciones de los parámetros

| Término                          | Estimación | Error estándar | Razón t | Prob >  t |
|----------------------------------|------------|----------------|---------|-----------|
| Constante del modelo             | 0.0853615  | 0.053732       | 1.59    | 0.1258    |
| Escala de rango[Trabajo externo] | 0.6392166  | 0.192401       | 3.32    | 0.0030*   |

#### Ajuste lineal para el sector de administración

Escala de rango[Transacciones locales] = 0.1813052 - 0.1070384\*Escala de rango[Trabajo externo]

#### Resumen del ajuste

|                                 |          |
|---------------------------------|----------|
| R cuadrado                      | 0.012429 |
| R cuadrado ajustado             | -0.03051 |
| Raíz del error cuadrático medio | 0.246817 |
| Media de respuesta              | 0.156847 |

|                                 |    |
|---------------------------------|----|
| Observaciones (o suma de pesos) | 25 |
|---------------------------------|----|

**Análisis de varianza**

| Fuente   | Grados de libertad | Suma de cuadrados | Media de los cuadrados | Razón F            |
|----------|--------------------|-------------------|------------------------|--------------------|
| Modelo   | 1                  | 0.0176341         | 0.017634               | 0.2895             |
| Error    | 23                 | 1.4011296         | 0.060919               | <b>Prob &gt; F</b> |
| C. Total | 24                 | 1.4187637         |                        | 0.5957             |

**Estimaciones de los parámetros**

| Término                          | Estimación | Error estándar | Razón t | Prob >  t |
|----------------------------------|------------|----------------|---------|-----------|
| Constante del modelo             | 0.1813052  | 0.067107       | 2.70    | 0.0127*   |
| Escala de rango[Trabajo externo] | -0.107038  | 0.198947       | -0.54   | 0.5957    |

**Ajuste lineal para el sector de investigación**

Escala de rango[Transacciones locales] = 0.0345182 + 1.0435273\*Escala de rango[Trabajo externo]

**Resumen del ajuste**

|                                 |          |
|---------------------------------|----------|
| R cuadrado                      | 0.819367 |
| R cuadrado ajustado             | 0.811513 |
| Raíz del error cuadrático medio | 0.099694 |
| Media de respuesta              | 0.12706  |
| Observaciones (o suma de pesos) | 25       |

**Análisis de varianza**

| Fuente   | Grados de libertad | Suma de cuadrados | Media de los cuadrados | Razón F            |
|----------|--------------------|-------------------|------------------------|--------------------|
| Modelo   | 1                  | 1.0369320         | 1.03693                | 104.3299           |
| Error    | 23                 | 0.2285963         | 0.00994                | <b>Prob &gt; F</b> |
| C. Total | 24                 | 1.2655283         |                        | <.0001*            |

**Estimaciones de los parámetros**

| Término                          | Estimación       | Error estándar  | Razón t      | Prob >  t         |
|----------------------------------|------------------|-----------------|--------------|-------------------|
| Constante del modelo             | <b>0.0345182</b> | <b>0.021901</b> | <b>1.58</b>  | <b>0.1287</b>     |
| Escala de rango[Trabajo externo] | <b>1.0435273</b> | <b>0.102164</b> | <b>10.21</b> | <b>&lt;.0001*</b> |

## ANEXO F

### CÓDIGO FUENTE

```

###   FUNCIÓN DE REDUCCIÓN
###   LENGUAGE DE PROGRAMACIÓN: R
###   LICENCIA: MIT

# Entrada: «Diccionario de datos»
construct <- function(commits, merges) { # La función requiere de dos argumentos
  length <- length(commits)
  t_df <- data.frame(1:length)
  for (i in 1:length) { # Inicio: Calcular desde el primer registro hasta el fin del archivo

    n <- get(commits[i]) # Recibe los flujos de trabajo
    m <- get(merges[i]) # Recibe las transacciones

    c_vals <- unlist(n[1:length(n)],use.names=F)
    m_vals <- unlist(m[1:length(m)],use.names=F)

    uc <- unlist(n[1:length(n)],use.names=F)
    nc <- unlist(n[1:length(n)],use.names=T)
    um <- unlist(m[1:length(m)],use.names=F)
    nm <- unlist(m[1:length(m)],use.names=T)

    cActors <- n[intersect(names(n), names(m))]
    mActors <- m[intersect(names(n), names(m))]

    owner <- max(unlist(n))
    mowner <- max(unlist(m))

    cttotal <- sum(unlist(n))
    mttotal <- sum(unlist(m))
    externaltotal <- mttotal
    complement <- cttotal - owner
    mcomplement <- mttotal - mowner

    L_CommitProportion <- max(uc)*100/cttotal
    L_MergeProportion <- max(um)*100/mttotal
    E_CommitProportion <- complement*100/cttotal
    E_MergeProportion <- mcomplement*100/mttotal
    L_CommitProportion
    t_df[i, 2] <- commits[i]
    t_df[i, 3] <- "owner"
    t_df[i, 4] <- sum(unlist(cActors))

    t_df[i, 5] <- cttotal
    t_df[i, 6] <- owner
    t_df[i, 7] <- complement
    t_df[i, 8] <- if(owner > each(externaltotal)){1}else{0} # Cálculo de la Hipótesis
    t_df[i, 9] <- L_CommitProportion
    t_df[i, 10] <- E_MergeProportion

    t_df[i, 11] <- mttotal # total of transactions
    t_df[i, 12] <- length(unlist(mActors)) # total of transactionNodes
    t_df[i, 13] <- mowner # transactions of the owner
    t_df[i, 14] <- mttotal - mowner # external transactions
    t_df[i, 15] <- L_MergeProportion # local proportion
    t_df[i, 16] <- if(is.numeric(E_MergeProportion)) { E_MergeProportion } else {0} # external proportions

    t_df[i, 17] <- max(unlist(cActors)) / max(unlist(mActors))
    # t_df[i, 18] <- sum(uc[uc != max(uc)]) / sum(um[um != max(um)])

    cA <- unlist(cActors[1:length(cActors)],use.names=F)
    cM <- unlist(mActors[1:length(mActors)],use.names=F)
    eWork <- sum(cM[cM != max(cM)]) * 100 / sum(cA[cA != max(cA)])
    t_df[i, 19] <- if(!is.nan(eWork)) { eWork } else {0}
  }
  t_df[1] <- NULL

```

```

# Define el nombre de las columnas
colnames(t_df) <- c("ID", "totalColab",
  "totalWork", "totalTransactions", "localWork", "externalWork", "transactionNodes", "localTransactions",
  "externalTransactions", "H0")
return(t_df) # Salida. Regresa la instancia con la base de datos
}

###   FUNCIÓN DE MAPEO
###   LENGUAGE DE PROGRAMACIÓN: RUBY
###   LICENCIA: MIT

module MailMapper
require 'json'

def self.find(path=" ", option = 0)
  text=File.open(path).read
  if option == 0
    mails = mapMails(text)
    return reduceMails(mails)
  elsif option == 1
    mails = mapMails(text)
    return mails.each_with_object(Hash.new(0)) { |mail,counts| counts[mail] += 1 }
  elsif option == 2
    mails = mapMails(text)
    return mapMails(text)
  end
end

def mapMerges(text)
  merges = []
  #Map text
  line_num=0
  text.gsub!(/\r\n?/, "\n")
  text.each_line do |line|
    if line[/((Merge)/)]
      merges << line[/((Merge)/)]
      nextLine = line + 1
      mails << nextLine[/[a-zA-Z_]+@[a-zA-Z_]+?[a-zA-Z]{2,3}\.[a-zA-Z]{2,3}|[a-zA-Z_]+@[a-zA-Z_]+?[a-zA-Z]{2,3}/]
    end
  end
  merge.compact!
  # BEGIN map-email function
  return merges
end

def reduceMerges(merges=@merges)
  mailExt = []
  mapHash = {}
  domain = ""
  mails.each {|x|
    # Set domain extension (after "@")
    domain = x[/@[a-zA-Z_]+?[a-zA-Z]{2,3}\.[a-zA-Z]{2,3}|@[a-zA-Z_]+?[a-zA-Z]{2,3}/]
    search = mapHash.select {|key, value| key[domain]}
    if search.empty?
      @r = mapHash.merge!({ domain => 1})
    else
      mapHash[domain] += 1
    end
  }
  return @r
end

def self.mapMails(text)
  mails = []
  #Map text
  line_num=0
  text.gsub!(/\r\n?/, "\n")
  text.each_line do |line|
    mails << line[/[a-zA-Z_]+@[a-zA-Z_]+?[a-zA-Z]{2,3}\.[a-zA-Z]{2,3}|[a-zA-Z_]+@[a-zA-Z_]+?[a-zA-Z]{2,3}/]end
  end
end

```

```

        mails.compact!
        # BEGIN map email function
        return mails
    end

def self.reduceMails(mails=@mails)
    mailExt = []
    mapHash = {}
    domain = ""
    mails.each {|x|
        # Set domain extension (after "@")
        domain = x[/@[a-zA-Z_]+?\.[a-zA-Z]{2,3}\.[a-zA-Z]{2,3}|@[a-zA-Z_]+?\.[a-zA-Z]{2,3}/]
        search = mapHash.select{|key, value| key[domain]}
        if search.empty?
            @r = mapHash.merge!({ domain => 1})
        else
            mapHash[domain] += 1
        end
    }
    return @r
end
end

3)

###      REPLICACIÓN DE LA ESTRUCTURA DE RED DE COLABORACION
###      LENGUAGE DE PROGRAMACIÓN: PYTHON / CYPHER
###      RE-DISTRIBUIDO DE: https://gist.github.com/knutwalker/d0ef536ec6ceafba570a#file-git2neo-py

#!/usr/bin/env python

from __future__ import print_function

from collections import namedtuple
from itertools import tee
from operator import itemgetter
from subprocess import Popen, PIPE
import re

Commit = namedtuple('Commit', ['sha1', 'hash', 'parents', 'author_email', 'author_name', 'refs', 'subject', 'timestamp', 'date_time'])
GIT_LOG_FORMAT = '%x1E'.join(['%H', '%h', '%P', '%ae', '%an', '%d', '%s', '%at', '%ai'])
# Define la variable que almacena los registros
seen_shas = set()

def as_list(elements):
    items = [e for e in elements if e]
    return items or None

def fix_parents(parents_string):
    return as_list(parents_string.split())

def fix_refs(refs_string):
    return as_list(refs_string.strip()[1:-1].split(', '))

def mk_commit(commit_line):
    commit_line = re.sub(r'^\x00-\x7F+', "", commit_line)
    commit = Commit(*commit_line.split("\x1E"))
    parents = fix_parents(commit.parents)
    refs = fix_refs(commit.refs)
    return commit._replace(parents=parents, refs=refs)

def create_node(neo4j_op, sha1_ident, property_pairs):
    if property_pairs:
        properties = ','.join('{0}: {1r}'.format(k, v) for k, v in property_pairs)
        return neo4j_op + '(:Commit {{{0}}});'.format(properties)
    else:
        return neo4j_op + '(:Commit);'
# Define la relación de los flujos de trabajo
def create_rel(neo4j_op, sha1, parent):

```

```
return 'MATCH (parent:Commit {sha1:"'+sha1+'"}), (child:Commit {sha1:"'+parent+'"}) '+neo4j_op+' (parent)<-[[:PARENT]]-(child);' #'La relación"
```

```
def mk_node_stmts(neo4j_op, sha1_ident, **properties):
    if sha1_ident not in seen_sha1s:
        seen_sha1s.add(sha1_ident)
        props = {k:v for k, v in properties.iteritems() if v is not None}
        yield create_node(neo4j_op, sha1_ident, sorted(props.items(), key=itemgetter(0)))
```

```
def node_stmts(neo4j_op, commit_or_sha1):
    if (hasattr(commit_or_sha1, 'sha1')):
        sha1, properties = (commit_or_sha1.sha1, commit_or_sha1._asdict())
    else:
        sha1, properties = (commit_or_sha1, {})
    return mk_node_stmts(neo4j_op, sha1, **properties)
```

```
def nodes(neo4j_op, cs):
    for c in cs:
        for s in node_stmts(neo4j_op, c):
            yield s
```

```
def missing_parents(neo4j_op, parents):
    for parent in parents or []:
        for n in node_stmts(neo4j_op, parent):
            yield n
```

```
def rel_stmts(neo4j_op, sha1, parents):
    for parent in parents or []:
        yield create_rel(neo4j_op, sha1, parent)
```

```
def join_rel_stmts(neo4j_op, cs):
    it = (s for c in cs for s in rel_stmts(neo4j_op, c.sha1, c.parents))
    a, b = tee(it)
    next(b, None)
    for _ in b:
        yield next(a)
    yield next(a)
```

```
def rels(neo4j_op, cs):
    for c in cs:
        for s in missing_parents(neo4j_op, c.parents):
            yield s
    for s in join_rel_stmts(neo4j_op, cs):
        yield ' ' + s
```

```
def mk_cmd(limit, branch):
    git_log = ['git', 'log', '--format=format: {0}'.format(GIT_LOG_FORMAT)]
    if limit is not None:
        git_log.append('-n')
        git_log.append('--all')
        git_log.append(str(limit))
    if branch is not None:
        git_log.append(str(branch))
    return git_log
```

```
def commits(limit, branch):
    proc = Popen(mk_cmd(limit, branch), stdout=PIPE)
    for line in proc.stdout:
        yield line.strip()
```

#### # Procedimiento para convertir los registros en nodos y enlaces

```
def main(neo4j_op, limit=None, branch=None):
    cs = map(mk_commit, commits(limit, branch))
    statements = []
    if cs:
        statements.append('BEGIN')
        statements.extend(nodes(neo4j_op, cs))
        statements.append('COMMIT')
        statements.append('BEGIN')
        statements.append('MATCH (n:Commit) SET n.date = substring(n.date_time,0,10);')
        statements.append('COMMIT')
```

```
statements.append('BEGIN')
statements.append('create index on :Commit(hash);')
statements.append('create index on :Commit(author_name);')
statements.append('create index on :Commit(author_email);')
statements.append('create index on :Commit(date);')
statements.append('create constraint on (c:Commit) assert c.sha1 is unique;')
statements.append('COMMIT')
statements.append('BEGIN')
if len(cs) > 1:
    statements.extend(rels(neo4j_op, cs))
statements.append('COMMIT')
return statements
# envía parámetros a Neo4J
if __name__ == "__main__":
    import argparse
    parser = argparse.ArgumentParser()
    parser.add_argument('-n', '--limit', type=int, metavar='N', help='only N latest commits', default=None)
    parser.add_argument('--merge', dest='neo4j_op', const='MERGE', default='CREATE', action='store_const', help='use MERGE instead of CREATE')
    parser.add_argument('--json', action='store_true', help='print JSON output for the Cypher REST API.')
    parser.add_argument('branch', help='which branch to examine - defaults to HEAD', nargs='?', default=None)
    arguments = parser.parse_args()
    res = main(arguments.neo4j_op, arguments.limit, arguments.branch)
    if arguments.json:
        import json, sys
        json.dump({"query": ''}.join(res), sys.stdout)
    else:
        print('\n'.join(res))
```