

**TESIS DOCTORAL**

**2016**

**LA INCIDENCIA DE LA TOPOLOGÍA DE LA RED  
SOCIAL EN EL RESULTADO ECONÓMICO EN  
EMPRESAS PYME: EL CASO DEL AEROPUERTO  
DE MUENSTER Y OSNABRUECK**

**FRANCISCO RODRÍGUEZ AGUILERA  
INGENIERO DE SISTEMAS**

**PROGRAMA DE DOCTORADO EN ECONOMÍA Y  
EMPRESA**

**DIRECTORES: DRA. D<sup>a</sup> NIEVES ARRANZ PEÑA  
DR. D. JUAN CARLOS FERNÁNDEZ DE  
ARROYABE**

TESIS DOCTORAL

2016

LA INCIDENCIA DE LA TOPOLOGÍA DE LA RED  
SOCIAL EN EL RESULTADO ECONÓMICO EN  
EMPRESAS PYME: EL CASO DEL AEROPUERTO  
DE MUENSTER Y OSNABRUECK

FRANCISCO RODRÍGUEZ AGUILERA  
INGENIERO DE SISTEMAS

PROGRAMA DE DOCTORADO EN ECONOMÍA Y  
EMPRESA

DIRECTORES: DRA. D<sup>a</sup> NIEVES ARRANZ PEÑA  
DR. D. JUAN CARLOS FERNÁNDEZ DE  
ARROYABE

## **Agradecimientos**

Este trabajo que ha necesitado más de 6 años, no hubiera sido posible de realizar sin el permanente apoyo y motivación de mi esposa Annette. Con su paciencia, pero consecuente apoyo no hubiera sido posible lograrlo, ya que este fue realizado en forma paralela a mi trabajo diario como CIO del Aeropuerto de Muenster y Osnabrueck. El contenido, estructura y forma de esta tesis fueron resultado de largas y meticulosas charlas con los directores de la tesis, que llevando un horario fuerte de trabajo no perdieron nunca la confianza en que el autor a pesar de la distancia lograra un trabajo como el que estoy presentando - Nieves y Juan Carlos muchas gracias.

Asimismo, quisiera agradecer también la comprensión y la confianza de todas aquellas personas, que sin su participación este trabajo no hubiera podido concluirse. En particular, quiero destacar el apoyo de los empleados y compañeros del Aeropuerto de Muenster y Osnabrueck.

---

**ÍNDICE**

LISTA DE SÍMBOLOS, ABREVIATURAS Y SIGLAS	9
LISTADO DE GRÁFICAS	10
LISTADO DE TABLAS	15
CAPÍTULO I. INTRODUCCIÓN	17
1.1. Antecedentes	17
1.1.1. Delimitación en el uso de "redes sociales"	25
1.2. Planteamiento de la investigación	27
1.3. Planteamiento del problema	32
1.4. Objetivo específico	34
1.5. Limitaciones	41
1.6. La empresa ejemplo	42
1.7. Guía de la Tesis	44
CAPÍTULO II. MARCO TEÓRICO – REDES SOCIALES	48
2.1. Introducción - Red Social	48
2.2. Análisis de Redes Sociales	52
2.2.1. Conceptos básicos	52
2.2.2. En la historia	54
2.2.3. Análisis de redes como análisis estructural	56

## CAPÍTULO I. INTRODUCCIÓN

2.2.4.	Capital social y análisis de redes	58
2.2.5.	Indicadores en el análisis de redes sociales	60
2.2.5.1.	Díada	64
2.2.5.2.	Tríada	65
2.2.5.3.	Grado	69
2.2.5.4.	Grupos y subgrupos /Cliques	72
2.2.5.5.	Densidad	73
2.2.6.	Mediciones de la evolución de red social	74
2.2.6.1.	Tipos de redes / modo	75
2.2.6.2.	Tipos de redes / contexto global	78
2.2.6.3.	Ejemplos de redes sociales	81
2.3.	El sistema de correo electrónico: Una comunidad virtual	83

## CAPÍTULO III. MARCO TEÓRICO – DESARROLLO ECONÓMICO DE EMPRESAS PYME

88

3.1.	Introducción	88
3.1.1.	Definición de empresas PYME	89
3.1.2.	Desarrollo económico en la CEE	91
3.1.3.	Demografía empresarial - Tendencias en Europa	96
3.1.4.	Análisis en función del tiempo	99

---

## CAPÍTULO I. INTRODUCCIÓN

3.2.	Conclusiones	101
CAPÍTULO IV. METODOLOGÍA		102
4.1.	Introducción	102
4.2.	Herramientas de software	103
4.2.1.	El lenguaje R	103
4.2.2.	Microsoft Excel	104
4.2.3.	UCINET	106
4.2.4.	MemBrain	107
4.3.	Construcción de la red social	109
4.3.1.	Proceso de adquisición de datos del sistema de correo electrónico	109
4.4.	Propuesta de un método automático	117
4.4.1.	SPAM y otro correo no deseado	118
4.4.2.	Desarrollo a través de tiempo	120
CAPÍTULO V. ANÁLISIS EMPÍRICO		123
5.1.	Análisis de los datos con ayuda de la teoría de grafos	123
5.2.	La red Social	129
5.3.	Análisis de los Datos de la red social con UCINET y R	135
5.3.1.	Análisis con UCINET	141

## CAPÍTULO I. INTRODUCCIÓN

5.4.	Análisis de los datos económicos de la empresa	157
5.4.1.	Indicadores económicos / desarrollo en función del tiempo	158
CAPITULO VI. PRIMERA APROXIMACIÓN A UNA CORRELACIÓN		163
6.1.	Introducción	163
6.2.	Propuesta para un modelo	165
6.3.	Descripción del modelo	169
6.4.	Factores necesarios para una normalización	171
6.5.	Influencias no posibles de cuantificar	171
CAPITULO VII. VERIFICACIÓN CON AYUDA DE REDES NEURONALES		174
7.1.	Redes Neuronales	174
7.2.	Elección de una red adecuada	177
7.3.	La red neuronal	182
7.3.1.	Neuronas ocultas	183
7.3.2.	Neuronas de entrada	185
7.3.3.	Neurona de salida	185
7.3.4.	Resultado de la red	186
CAPITULO VIII. CONCLUSIONES Y RECOMENDACIONES		187
8.1.	Conclusiones generales	187

## CAPÍTULO I. INTRODUCCIÓN

8.1.1.	Dinámica de la red social y su evolución	188
8.1.2.	Déficit interpretativo	188
8.1.3.	Foco de las redes como variables aclaratorias	189
8.2.	Conclusiones – red social	190
8.3.	Conclusiones – efectos al desarrollo económico	192
8.4.	Puntos abiertos	194
	Bibliografía	197
A.	Apéndice – Programas en R	232
B.	Apéndice – Programa en Visual Basic	235
C.	Apéndice – Programa en VBA (Excel)	249
D.	Apéndice – Extracto de la Base de Datos (normalizados) 2002-01 a 2005-09	251
E.	Apéndice - Ponderación de las conexiones de la red neuronal	252

**LISTA DE SÍMBOLOS, ABREVIATURAS Y****SIGLAS**

ANOVA	Análisis de la varianza (ANalysis Of VAriance)
AWU	Cantidad de trabajadores promedio anual
C, C++,C#, VB	Lenguajes de programación
DFS	Compañía controladora de tráfico aéreo en Alemania
EBIDTA	Earnings Before Interest, Taxes, Depreciation, and Amortization (beneficio antes de intereses, impuestos, depreciaciones y amortizaciones)
ERP	sistemas de planificación de recursos empresariales
ICAO, ACI, IATA, ADV	ICAO: International Civil Aviation Organization ACI: Airports Council International IATA: International Air Transport Association ADV: Organización de Aeropuertos de habla alemana
PYME	Pequeña y mediana empresa
R	Paquete de Software con funciones estadísticas
SAP R/3	ERP Software de la compañía alemana SAP AG
SPAM	Correo basura o mensaje basura
UCINET	Paquete de software para el análisis de datos de redes sociales
USENET	Users Network (Red de usuarios)
OECD	Organización para la Cooperación y el Desarrollo Económicos

**LISTADO DE GRÁFICAS**

Gráfica 1 Modelo teórico .....	34
Gráfica 2 Visualización de una comunicación por email .....	38
Gráfica 3 Distribución de propietarios del Aeropuerto de Muenster y Osnabrueck .....	44
Gráfica 4 Estructura de la tesis .....	47
Gráfica 5 Ejemplo red con conexiones no dirigidas .....	62
Gráfica 6 Ejemplo de red con conexiones dirigidas .....	63
Gráfica 7 Las 16 diferentes clases de triadas en M-A-N notación según P. W. Holland y S. Leinhard.....	66
Gráfica 8 Tríada transitiva .....	69
Gráfica 9 Visualización de un actor importante .....	70
Gráfica 10 Ejemplo de una red social modo dos.....	77
Gráfica 11 Red social del gobierno iraní (2004) en drzaius.ics.uci.edu .....	81
Gráfica 12 Análisis de una red social en Arqueología en electricarchaeology.ca .....	81
Gráfica 13 Conexiones del Democracy Institute con los movimientos conservadores en powerbase.info .....	82
Gráfica 14 Análisis de una red social por Michael Wu en lithosphere.lithium.com .....	82

## CAPÍTULO I. INTRODUCCIÓN

---

Gráfica 15 El perfil de la PYME española frente a la de la UE-27 (Fuente: Comisión Europea, <a href="http://www.oecd.org">www.oecd.org</a> (2006)) .....	92
Gráfica 16 Número de PYMES, empleados y valor agregado de PYMES.....	93
Gráfica 17 Crecimiento anual de empleos, productividad y valor agregado por países del 2008 al 2011 (Fuente: (Wymenga, Spanikova, Barker, Konings, & Canton, 2012) .....	94
Gráfica 18 Número de empresas en España por año y clase .....	99
Gráfica 19 Descripción de los registros de la base de datos.....	114
Gráfica 20 Proceso de adquisición de datos.....	116
Gráfica 21 Diagrama de flujo de datos para borrar correos SPAM..	118
Gráfica 22 Pronóstico del uso de correo electrónico hasta el 2019.	120
Gráfica 23 Porcentaje de usuarios de correo electrónico en.....	121
Gráfica 24 Número estimado de mensajes electrónicos enviados y recibidos en el mundo a diario en miles de millones.....	122
Gráfica 25 Dígrafo ponderado o red a partir de dos correos electrónicos.....	124
Gráfica 26 Distribución por grupo de los integrantes de la red social.....	131
Gráfica 27 Distribución de mensajes enviados y recibidos .....	132

## CAPÍTULO I. INTRODUCCIÓN

---

Gráfica 28 Distribución de los integrantes de la red social según grupo (solo externos) .....	133
Gráfica 29 Cantidad de mensajes a través del tiempo por grupos y años .....	134
Gráfica 30 Visualización de la red de modo 2 con UCINET basada en intercambio de correo electrónico por el director de la empresa.....	135
Gráfica 31 Transitividad, Densidad, Grado y Diametro de la red normalizada.....	139
Gráfica 32 Componentes, Conexiones, Vértices y Aristas de la red normalizada.....	139
Gráfica 33 Correlación y Covarianza de la red normalizada .....	140
Gráfica 34 Cantidad de emails distribuidas por meses .....	141
Gráfica 35 Porcentaje de personas en el UE entre 16 y 74 años que usan regularmente email .....	142
Gráfica 36 Red social años 2002 y 2003 .....	143
Gráfica 37 Red social años 2004 y 2005 .....	144
Gráfica 38 Red social años 2006 y 2007 .....	144
Gráfica 39 Red social años 2008 y 2009 .....	145
Gráfica 40 Red social desde el año 2002 hasta el 2010.....	154
Gráfica 41 Red social sobre todos los tiempos visualizada con K-cores.....	156

## CAPÍTULO I. INTRODUCCIÓN

Gráfica 42 Curva del desarrollo de pasajeros y movimientos .....	157
Gráfica 43 Curva de ingresos normalizada .....	158
Gráfica 44 Curva de costos normalizada .....	158
Gráfica 45 Curva del balance general normalizada .....	159
Gráfica 46 Curva de cantidad de pasajeros y movimientos aeronáuticos.....	160
Gráfica 47 Curva de tendencia, temporada y restos del balance general .....	161
Gráfica 48 Balance general y el promedio en periodos de 24 .....	162
Gráfica 49 Desarrollo por mes del balance general.....	162
Gráfica 50 análisis lineal entre BALANCE y DENSIDAD.....	163
Gráfica 51 Análisis lineal entre Balance y Componentes.....	164
Gráfica 52 Correlación entre Costos y Correlación de la red social .	165
Gráfica 53 Comparación del comportamiento de la transitividad en referencia a los ingresos en transcurso del tiempo.....	167
Gráfica 54 Comparación del comportamiento del grado en referencia a los ingresos en transcurso del tiempo.....	168
Gráfica 55 Desarrollo del tráfico aéreo en Alemania desde el año 1993 .....	172

Gráfica 56 Neurona u: $w$ ponderación de las conexiones, $\varepsilon$ entrada efectiva, $a(t)$ función de activación y $o(t)$ señal de salida .....	176
Gráfica 57 Una red neuronal con 10 neuronas, 4 entradas y 4 salidas.....	176
Gráfica 56 Error de red.....	180
Gráfica 58 Red neuronal .....	183
Gráfica 59 Grafica de activación función logística .....	184
Gráfica 60 Pattern Error .....	186
Gráfica 61 GUI del Programa para extraer información de correos y guardarlos en una base de datos .....	235

---

**LISTADO DE TABLAS**

Tabla 1 Porcentaje de compañías controladas por familias .....	21
Tabla 2 Ejemplo análisis clásico de características.....	61
Tabla 3 Ejemplo red social: ¿quién llama a quién? .....	62
Tabla 4 Comparación de densidad y otros valores en una red.....	74
Tabla 5 Clasificación de PYMEs según la UE.....	89
Tabla 6 Formato de exportación desde Microsoft Outlook.....	110
Tabla 7 Características del contenido de la base de datos .....	117
Tabla 8 Correos significantes por año .....	121
Tabla 9 Ejemplo de miembros que no se tienen en cuenta en la evaluación .....	129
Tabla 10 Resumen de los datos a ser analizados.....	138
Tabla 11 Correo Electrónico 2002 .....	146
Tabla 12 Correo Electrónico 2003 .....	146
Tabla 13 Correo Electrónico 2004 .....	147
Tabla 14 Correo Electrónico 2005 .....	148
Tabla 15 Correo Electrónico 2006 .....	149
Tabla 16 Correo Electrónico 2007 .....	150
Tabla 17 Censo de tríadas año 2007 .....	151
Tabla 18 Censo de tríadas año 2008 .....	152
Tabla 19 Correo Electrónico 2008 .....	153

## CAPÍTULO I. INTRODUCCIÓN

---

Tabla 20 Correo Electrónico 2009 .....	153
Tabla 21 Correo Electrónico 2010 .....	155
Tabla 22 Mínimos y máximos de valores a ser analizados .....	170
Tabla 23 tipos de redes neuronales acuerdo a tipo de aprendizaje	178
Tabla 24 Características de la red neuronal.....	179
Tabla 25 método de activación de una neurona .....	181
Tabla 26 Neuronas ocultas.....	184
Tabla 27 Neuronas de entrada .....	185
Tabla 28 Neurona de salida.....	185

---

## CAPÍTULO I. INTRODUCCIÓN

### 1.1. Antecedentes

El impacto e influencias de las redes sociales al transcurso de la historia es posiblemente el hecho más contundente en el desarrollo de la especie humana (Burguière, Klapisch-Zuber, & Segalen, 1996), (Morris, 2011). El desarrollo del análisis de redes sociales inicialmente no fue de carácter técnico y su estructura inicial es basada en la teoría del antropólogo Radcliffe-Brown (Barnard, 1992). Desde los años 30 hasta mediados de los 70 una gran cantidad de antropólogos (Kuper, 1977) y sociólogos empezaron a construir alrededor del concepto de "estructura social" de Radcliffe-Brown, usando nuevas metáforas como "fábricas" y "web" que en su idioma original es entendido como "red". Partiendo de estos conceptos y entendiendo estos como una interactividad de relaciones, donde acciones sociales son organizadas, surge más y más el concepto, el término de "red".

El antecedente del concepto de red social es la familia. Nosotros todos vivimos con el principio fundamental de la familia como parte fundamental de nosotros mismos (Huinink & Konietzka, 2007). De por sí podemos decir que el éxito o fracaso de una sociedad radica

---

en si los miembros de ella y por consiguiente su red es lo suficientemente fuerte (Leinhardt, 1977).

Analizando el éxito y caída de la familia Medici (Hibbert, 1999), se ve la complejidad, necesidad y efectividad de una red social. Pero también en el tiempo moderno hay ejemplos clásicos de cómo en forma efectiva redes sociales afectan la situación económica o desarrollo de una red. Burkart (Burkart, Panunzi, & Shleifer, 2002) analiza la influencia de las familias en las grandes firmas del mundo. No solo en Estados Unidos sino también en Japón o en general en Europa ofrecen los autores una teoría clara de cómo los clanes familiares han mantenido su influencia a través de los años. Este análisis de carácter matemático hace proposiciones y trata de demostrarlas con ayuda de corolarios, llegando a una conclusión esperada: *"La familia fundadora debe quedarse y dirigir la empresa, y sólo puede darse el lujo de ceder el control a un gestor profesional si lo convierten en un miembro de la familia. La separación entre la propiedad y la gestión es por lo tanto una indicación de un ambiente superior jerárquico corporativo. La falta de tal separación, y la prevalencia de las empresas familiares, es evidencia de subdesarrollo financiero"*.

Camiller y Monsutti (2005) analizan un aspecto totalmente diferente del desarrollo de las redes sociales. Ellos examinan las redes

en Afganistán. Lo interesante en este estudio, es el hecho que no se trata de un mundo occidental bien conocido y mucho menos de grandes empresas multinacionales, sino que estudia los Hazara que son un grupo étnico de lengua persa que reside en la región central de Afganistán y por lo tanto aislado y específico en medio de un conflicto bélico. Y lo sorprendente es el hecho que la red social de este grupo es parte fundamental de su supervivencia y por lo tanto de su desarrollo económico.

No obstante, nuestra cultura occidental y en general el desarrollo moderno de la economía que conocemos, tiene raíces muy antiguas. Constantakopoulou, Malkin y Panagopoulou (2012), en su estudio sobre las redes griegas y romanas en el mediterráneo en la edad antigua, analizan aspectos de la historia económica, social, religiosa y política de la época. En su conclusión dejan algunos aspectos abiertos: *“El concepto de red nos puede ayudar a escribir una nueva historia griega, tanto por debajo y más allá de la polis. Si vamos a ser capaces de beneficiarnos de esta problemática recíproca aún está por verse.”* No obstante, cabe anotar que la información disponible de aquella época es abundante pero no de contexto profundo, más sin embargo deja claro, que el poder desarrollado en aquella época, nunca hubiera sido posible sin el uso extenso y continuo de las redes sociales. El estudio muestra el factor definitivo del alcance del análisis de la red

## CAPÍTULO I. INTRODUCCIÓN

social, a pesar de tener solo datos fragmentados o incompletos. Cuanta más información se tiene, mejor pueden hacerse deducciones lógicas y comprobables.

Veamos otro estudio antes de profundizar el tema. Miller y Le Breton-Miller en su estudio sobre la búsqueda de los managers ganadores a largo plazo, encuentran que las compañías más exitosas en transcurso del tiempo están en su mayoría controladas por clanes familiares (Miller & Le Breton-Miller, 2005). Veamos esta tabla que hicieron en su estudio:

Prevalencia	Muestra	Definición
35%	S&P 500 (2003)	Familia fundadora es miembro del comité ejecutivo como ejecutivo o accionista mayoritario
32%	S&P 500 (1992-1999)	Miembros de la familia fundadora son miembros del comité ejecutivo y tienen acciones en ausencia de los bloques exteriores mayores de > 5%
37%	Fortune 500 (1992)	Ejecutivo más influyente es miembro de la familia o descendiente del fundador
35%	Fortune 500 (1987)	Familia fundadora en posición alta dirección o miembro del comité ejecutivo
47%	Fortune 500 (1965)	5% es propiedad de la familia, representación en el comité ejecutivo y las generaciones múltiples están envueltas
21%	BusinessWeek CEO 1000 (1993)	Ejecutivo más influyente es miembro de la familia o descendiente del fundador
60%	54.000 compañías públicas de Estados Unidos (1996)	Bajo control cerrado de la familia

## CAPÍTULO I. INTRODUCCIÓN

Prevalencia		Muestra	Definición
20%	Estados Unidos	20 compañías más grandes de sector público (1990)	Familias controlan > 20% de las acciones y son los mayores accionistas con un factor mayor de 2
25%	Canadá		
55%	Italia		
70%	Hong Kong		
44%	Europa Occidental	5.232 compañías públicas (1996)	Familias controlan > 20% de los votos
65%	Alemania		
65%	Francia		
56%	España		
57%		250 compañías del Paris Stock Exchange (1993-1998)	Familias controlan > 10% del patrimonio

**Tabla 1 Porcentaje de compañías controladas por familias**

Antes de sacar una conclusión de estos datos, quiero hacer un enlace al análisis de redes como tal. Uno de los precursores en el pensamiento sobre análisis de redes fue Georg Simmel. Simmel (1858-1918) vio en la llamada interacción el objeto principal de la sociología y la base de su autonomía (Simmel, 1890). Sociología debía ocuparse más de las formas sociales y menos en su contenido. Con este simple paradigma postuló Simmel el análisis de características relacionales de las relaciones entre individuos como centro de la sociología.

El concepto de análisis de redes sociales es formulado por primera vez a principios de los 70 por Lorrain, Breiger, Boormann y Burt (Scott, 1991). Esto sucede después de que el desarrollo en el tiempo haya tenido diferentes vías como lo fue en los años 50/60 por un lado Cartwright y Harary (1956) y por otro lado Harrison White en 1963 (Stegbauer, 2005).

Teniendo en cuenta estos estudios, concluye Miller y Le Breton-Miller (2005) que la red social es fundamental a largo plazo para el desarrollo exitoso de empresas que actúan a nivel global.

Ahora teniendo en cuenta esta información y la experiencia en sacar conclusiones a causas de formar parte de una red, de los efectos en grandes sociedades y empresas, se plantea el pensamiento siguiente: ¿Es posible hacer análisis parecidos en un contexto más definido y pequeño o puntual como lo es una empresa PYME? Y si el resultado de un análisis en el campo macro toma demasiado tiempo, y por ello solo sirve para responder preguntas del pasado, ¿No sería más útil hacer un análisis en forma más cercana al momento actual y no solo reflexionar sobre el pasado lejano sino sacar conclusiones inmediatas para apoyar la gestión de la empresa?

El proyecto más cercano a este trabajo fue realizado por van Alstyne y Zhang en el año de 2003 (Van Alstyne & Zhang, 2003), en el cual proponen un sistema para analizar la red social haciendo un grafo sobre el uso del correo electrónico, sin embargo, su enfoque supone el acceso a todos los mensajes de correo electrónico de una comunidad y no sólo de un usuario en particular. Desde el enfoque técnico esto sería posible, pero las barreras burocráticas y legislativas hacen que esto sea prácticamente imposible para una empresa en general.

También ha habido trabajos que consideran en el análisis de gráficos de coautoría (Newman, 2001) que son similares al trabajo de Culotta, Bekkermann y McCallum (2004) que se basan en la extracción de nombres de coautor de las páginas del publicaciones científicas.

Recientemente Angela Bohn, Kurt Hornik y Patrick Mair del Institute for Statistics and Mathematics del Department of Finance, Accounting and Statistics de la Universidad de Viena e Ingo Feinerer del Database and Artificial Intelligence Group del Institute of Information Systems de la Universidad Tecnológica de Viena (Bohn, Feinerer, Hornik, & Mair, 2011) hicieron un estudio con ayuda del programa de Software R para estudiar las relaciones entre personas en una red social, analizando el contenido del correo electrónico. Lo interesante del estudio es el hecho de que se usan herramientas de extracción de información de sistemas de correos electrónicos y estas son catalogadas en forma parecida a lo propuesto en este trabajo, cuyo foco es tenido en cuenta en el desarrollo del mismo.

En todos estos aspectos analíticos falta un aspecto fundamental para establecer un beneficio práctico de los resultados. Se trata de buscar una conexión a los resultados económicos de la red el cual es asunto de esta investigación. ¿Qué significa el hecho de identificar cliques en una red? ¿O cuál es el efecto inmediato de

---

determinar la actividad de los actores principales de la red en el desarrollo de la misma?

El aspecto social es bien descrito y es posible determinar las causas de, por ejemplo, la pobreza o riqueza de un grupo determinado, como se ha mostrado anteriormente (Miller & Le Breton-Miller, 2005). Estos y muchos otros aspectos sociales no son objeto de esta investigación. En especial por el hecho de tener una red social ya determinada desde el principio: La red de la empresa a ser estudiada.

Este trabajo tiene como objetivo determinar qué efectos en función del tiempo tiene el desarrollo de la red social de los miembros determinantes de una empresa PYME, tales como gerencia, dirección operativa y dirección de sistemas en el resultado económico de la misma. Estudios sobre los clanes de familias famosas e influyentes como por ejemplo en la antigua roma (Audioguiaroma, 2011) o en Estados Unidos (Kennedy, 2011) muestran la influencia de la red social en el desarrollo de las diferentes economías. En (Beaudreau, 2004) y (Malkin, Constantakopoulou, & Panagopoulou, 2012) se describe la influencia de las redes sociales en diferentes contextos económicos a nivel geográfico. Los autores llegan a la conclusión, que el análisis (clásico) de redes sociales, la inclusión de la geografía de la red, la densidad social y la calidad de las relaciones depende de la demarcación territorial. Sin embargo, en relación con el grupo de

investigación dirigido por Wellman (2002) se utiliza el argumento de que las redes son cada vez más de espacio independiente, especialmente con el apoyo de forma de comunicación entre sus integrantes, en la que Internet juega un papel importante.

### **1.1.1. Delimitación en el uso de “redes sociales”**

Dentro del hábito popular general y científico, el concepto de “redes sociales” ha sido usado en diferentes ámbitos y nociones:

- Redes sociales (Internet): se refiere en general a grupos de Internet de fácil acceso (comunidades en línea) (Welfens, Zoche, Jungmittag, Beckert, & Joisten, 2005). Las características típicas de ese tipo agrupaciones son, por ejemplo:
  - perfil personal
  - fácil accesibilidad
  - Envío y recepción de mensajes
  - Actualizaciones de estado
  - Juegos sociales
- Redes sociales (Administración de Empresas): es una forma específica, de forma deliberada y suelta de organizaciones con relaciones comunes con objetivos de establecer asociaciones

informales a través de grupos o personas con el propósito de obtener ventajas (Mackay, 1997).

- Redes sociales (Teoría de Sistemas): es a menudo entendida como un sistema (Weyer J. , 2000). Weyer dice que redes sociales (sistemáticas) se pueden dividir en cuatro categorías según su función en áreas de acción:
  - Redes estratégicas
  - Redes regionales
  - Redes políticas
  - Redes de innovación
- Redes sociales (Sociología): es una descripción de las interacciones sociales de cualquier tipo. El concepto fue usado por primera vez en la antropología social inglesa por J. Clyde Mitchell (1969), A. L. Epstein y Bruce Kapferer para describir organizaciones de inmigrantes sueltas e individuales en ciudades industriales coloniales como por ejemplo en Zambia. El término fue transferido a Europa entre otros, para estudiar factores de influencia de la vida laboral entre cónyuges (Bott E. , 1964). En los EE.UU., fue entablada la investigación empírica de redes sociales bajo el concepto sociológico inicialmente por Harrison C. White (White & Jorion, 1992) y popularizado sobre todo por Granovetter (1973).

La plataforma de este trabajo, a pesar de tener una familiaridad con Internet al usar correo electrónico, con el concepto en Administración de Empresas, por el hecho de que los actores son en su mayoría miembros de empresas en relación y con la teoría de sistemas al usar valores desarrollados por Weyer, es conceptualmente y básico en su desarrollo y objetivo entendido en el ámbito sociológico.

## **1.2. Planteamiento de la investigación**

Gran parte del análisis se basa en la recolección de datos. Los primeros trabajos en este campo fueron realizados con el sistema Ahoy (Shakes, Langheinrich, & Etzioni, 1997) que utilizaba técnicas heurísticas. Los siguientes estudios han sido en su mayoría enfocados en la extracción de información de las páginas de internet. Es así como Xi et al. (2002) hacen un enfoque a una máquina de aprendizaje utilizando árboles de decisión y de regresión logística, con resultados moderadamente bajos y Upstill, Craswell y Hawking (2003) describen una técnica de recuperación de información que aumenta la recuperación de contenido basado en el análisis de enlace (el PageRank de la búsqueda de Google) y las características de las URL.

En un trabajo de Borkar, Deshmukh, y Sarawagi (2000) relacionado con la extracción de registros de contactos obtienen una alta precisión de extracción de datos de las páginas de internet,

buscando y usando un número limitado de información (número de casa, calle o carretera, ciudad, estado y código postal) que por lo general aparecen en forma muy constante y formalizada.

Una investigación similar de redes sociales fue llevado a cabo con datos de Usenet (Smith V. , 2009), cuyo objetivo fue el de caracterizar una comunidad en forma dinámica, así como de determinar la "autoridad" de un individuo en particular basándose en sus aportes.

La idea de esta investigación es tratar de encontrar un método claro y definido para recolectar información, analizar y clasificar esta misma con ayuda de herramientas de software que se encuentren en el mercado o que sean de fácil adaptación a la empresa objeto de investigación.

Para realizar un análisis de redes sociales se buscan estrategias para la adquisición de datos. Mi objetivo es demostrar que la ganancia de datos en forma electrónica, a pesar de ser puntal y poco descriptiva en su grado de intensidad, si puede ser usada como fuente de datos para realizar un análisis contundente y solo dependiente de la cantidad de información disponible. Es importante resaltar que el análisis de la red social comprende primero una depuración de los datos ganados. Es decir, la identificación de ellos y el determinar si tienen

---

influencia o no, es un objetivo principal de este estudio. De esta forma se establecen métodos, ayudas y sugerencias para ello.

Preguntas que se presentan en el análisis, como son el valor de una conexión basada en si el originario de un correo electrónico manda éste en forma directa o en copia o copia ciega y para ello si se trata de un determinado grupo, son tratadas específicamente.

El análisis de la red toma como referencia la teoría de grafos. Se analizan teorías que pueden ser aplicadas como son la posible influencia de cascadas y el modelo de pequeño mundo. La determinación de los factores medibles lineales del resultado del análisis de la red son objetivos principales para poder hacer una comparación matemática entre dos campos que en su forma aislada son normalmente analizados en forma diferente. Es así como el análisis de redes – sin importar su significado – tiende a ser puntual o multidimensional, por el contrario, el análisis económico por su calidad tiende a ser lineal o bidimensional. Cuando se habla de redes, se habla de expansión, densidad, de determinación cliques o grupos, de la determinación de nodos con mayor influencia. Estos factores vistos en su forma singular no los podemos usar para analizar la influencia en factor del tiempo en el desarrollo económico, que por su definición intrínseca se desarrolla en forma lineal continua. La solución o la recomendación de una solución para resolver este conflicto, forma

---

parte fundamental de este trabajo. Para ello se muestran diferentes estrategias, dando a conocer las ventajas y desventajas de cada una de ellas.

Atendiendo las leyes de protección de datos, desarrollaré una metodología científica, que, por su característica, podrá ser utilizado sin comprometer los datos personales de los integrantes. En especial el hecho de elaborar un algoritmo que elimine la información no relevante o poco relevante del sistema a ser investigado, ayudará a la comunidad científica en el futuro a desarrollar investigaciones posteriores sin invertir tiempo valioso en determinar el marco de ley en la que se pueda adquirir información.

El análisis económico de la empresa en cuestión toma como información datos acumulados mensuales de los últimos 10 años. Estos son evaluados y clasificados en su significado y las capacidades de ser influenciados por los actos de política de gestión a corto, mediano y largo plazo. A partir de esos determinantes se encuentran qué factores de costo o beneficio pueden ser influenciados mediante la intensidad de la red social y se establecen los parámetros precisos, que son actos para ser utilizados como variables en el contexto comparativo del análisis final. En el análisis se tendrán en cuenta varios aspectos que por separado deberán ser analizados o tomados en cuenta. Como primer factor se define qué tipo de resultado económico es influenciado

directamente, es decir, qué indicadores económicos son posibles de ser usados en el análisis científico y cuáles no tienen sentido en esta investigación. Naturalmente para algunas empresas son de crucial importancia algunos indicadores económicos, tales como la cuota de capital ajeno o de deuda sobre capital, pero estos factores no son en general deducibles a un límite menor del año fiscal y son influenciados por las políticas de balance general anual. Como segundo factor se calcula un EBIDTA (Weber & Schäffer, 2008) mensual y anual diferenciado. Si bien es cierto que este indicador financiero es criticado y muchas veces no es aceptado como base de un resultado económico de un empresa (Malik, 2003) desde mi punto de vista es un indicador que muestra en una forma directa si las actividades de la red social son influyentes o no. Todos los indicadores económicos (Fridson & Alvarez, 2002) que se utilicen, serán formalizados de tal forma que su valor sea mensual. Esto último tiene por efecto mostrar la consecuencia a mediano plazo, si lo hay.

Finalmente, con el objeto de verificar los resultados obtenidos, acudo a las redes neuronales. Especialmente con el propósito de comprobar los resultados del análisis estadístico clásico y hacer un recorrido de posibles alternativas para posteriores enfoques.

### **1.3. Planteamiento del problema**

Todos los trabajos anteriores en este campo, coinciden con un aspecto supremamente importante para poder hacer el análisis: La adquisición de los datos y su calidad son fundamento principal para lograr un resultado positivo de la investigación. ¿Pero a qué datos se tiene acceso?

El desarrollo de la informática en los últimos años, nos abre un campo absolutamente nuevo para adquirir información de posible uso. Si anteriormente la adquisición de datos se basaba en formularios y cuestionarios de preguntas, que con mucha intensidad humana se llenaban, teniendo como problema la subjetividad y la capacidad de responder las preguntas hechas por el personal entrevistado, hoy en día es posible acceder a información objetiva no manipulada como son los protocolos de la comunicación electrónica.

Resumiendo, se obtienen dos fases en el desarrollo de la investigación que la hacen que tenga un sentido interdisciplinario.

Por una parte, se adquieren datos de la situación económica de la empresa y se analiza su desarrollo a través del tiempo. Se deducen premisas, se destacan momentos de crisis o evolución en la misma. Se depuran los resultados en valores influenciados y no influenciados dando una visibilidad de la gestión.

Por otro lado, se toman datos del manejo de correo electrónico de los encargados de la gestión (tres de ellos) y de ellos se deduce y construye o visualiza una red social. Esta es medida y analizada. Se buscan factores lineales medibles, que describan la red social en su totalidad, buscando factores descriptivos de su evolución en transcurso del tiempo. Estos datos al igual que los datos de contabilidad, tienen que ser depurados en significativos y no significativos, basándonos en un proceso descriptivo formal aplicable a la mayoría de casos posibles de encontrar, dando indicaciones de cómo afrontar el proceso de adquisición y depuración.

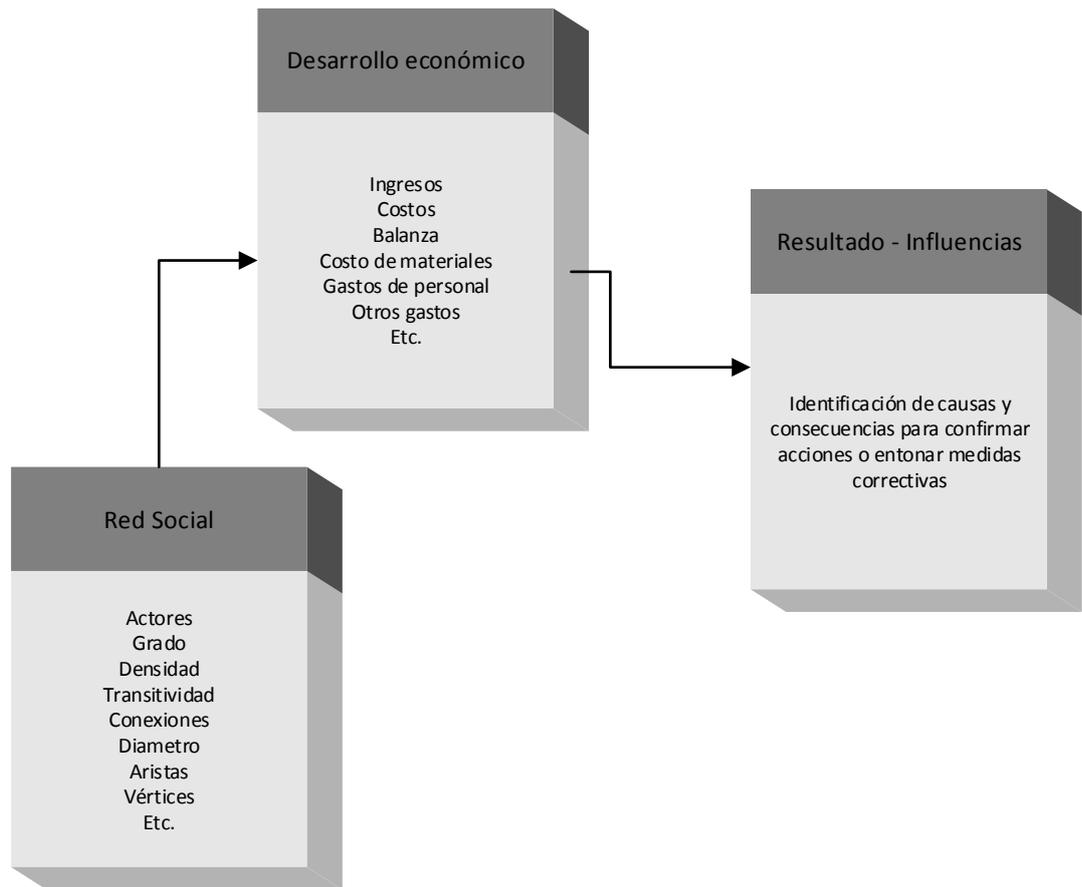
Posteriormente se busca establecer una correlación entre los valores ganados de la situación económica y los valores lineales que describen el desarrollo de la red social (Gráfica 1 Modelo teórico).

En el modelo se toman como:

**Variables dependientes:** El desarrollo económico de la empresa. En el modelo se buscan variables que describen el desarrollo económico de la empresa a través del tiempo.

**Variables independientes:** El desarrollo de la red social. La observación empírica de la red construida a base de la comunicación de correos electrónicos, nos proporciona un conjunto de parámetros. Estos constituyen el conjunto de variables independientes.

La determinación de qué valores son adecuados para ser tomados como variables independientes y dependientes, es en igual forma parte de la investigación.



Gráfica 1 Modelo teórico

#### 1.4. Objetivo específico

Tomando como base el desarrollo tecnológico de los últimos 10 años, nos damos cuenta que la forma y método de adquisición de datos se ha transformado radicalmente. Cuando en el pasado las investigaciones tenían como punto central la búsqueda de métodos

## CAPÍTULO I. INTRODUCCIÓN

para la consecución de información, desarrollando estrategias en el planteamiento y concepción de formularios de preguntas, hoy en día podemos adquirir datos para ser analizados de una forma mucho más rápida y segura - en su valor - a través de medios electrónicos.

La informática tiene por objeto proporcionar una base científica para el estudio del procesamiento de la información, la solución de problemas mediante algoritmos, y el diseño y la programación de ordenadores (Arbib, Kfoury, & Moll, 1984). En esta definición no se hace una diferencia entre hardware o software.

Después de desarrollar un concepto, se construye o programa un prototipo y se trata de mejorar el mismo en un tiempo determinado, y través del conocimiento ganado en esta etapa, se busca perfeccionar o ratificar el concepto original. Para hacer todo este proceso, se conciben desde un principio métodos para analizar su desarrollo. Este concepto llamado depuración se basa en la grabación de registros de los pasos seguidos en el transcurso o evolución del sistema en desarrollo. Estos registros son una fuente de información grande, pero como su objetivo no tiene un fin estadístico o analítico, es necesario analizar su construcción y establecer algoritmos y métodos de extracción y filtración.

En especial en este estudio se utilizarán los registros de email de la empresa a analizar, en particular los de tres miembros de la gestión de administración: del director de la empresa, del director de operaciones de la empresa y del jefe de sistemas, que en este caso es el mismo autor del trabajo.

En el análisis económico de la empresa se desarrolla una metodología basada en métodos empíricos tales como el análisis de la regresión, análisis de la varianza o ANOVA y análisis factorial con el propósito de establecer un pronóstico adecuado del desarrollo de la empresa (Hatzinger, Hornik, & Nagel, 2011). Con la ayuda de redes neuronales aplicadas año por año y emitiendo una corrección de predicción se comprobarán los resultados del análisis empírico. Para poder correlacionar el desarrollo de la red social y su efecto en el desarrollo económico es necesario encontrar un determinante o variable en función del tiempo con la cual la diferencia entre acción y resultado sea medida.

Por un lado, se usan métodos empíricos en el análisis económico en la evaluación de la empresa y por otro lado se usan métodos de análisis de redes sociales. Al determinar los nodos de la red y su característica hay que ver su transformación a través del tiempo. Para ello es necesario hacer un estudio de cada nodo en particular. El hecho de que un nodo tenga un valor específico en la red

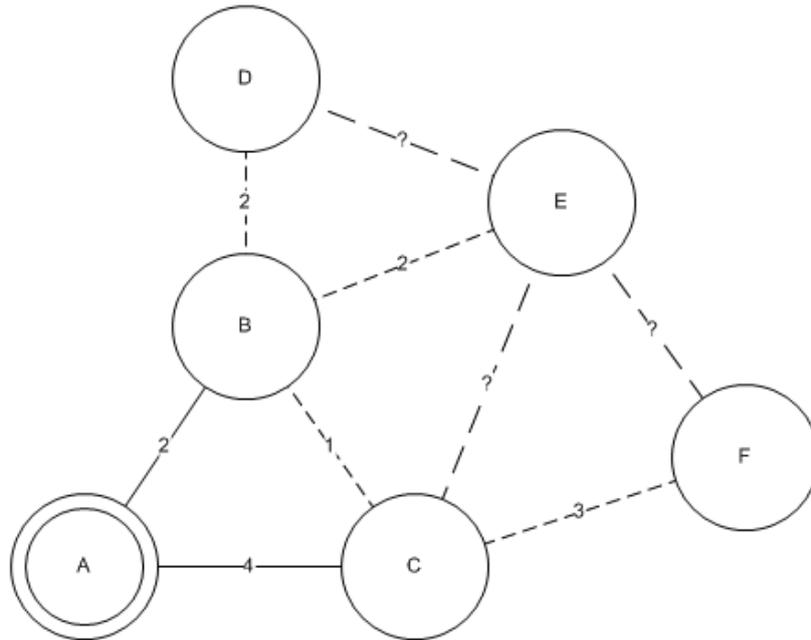
y su influencia sea clara, no significa al mismo tiempo que su influencia sea grande en el desarrollo de la empresa. Es por eso que la clasificación de los miembros de la red tiene que ser considerada en forma puntual, dándole un valor específico si su condición es de cliente, dueño, medio informativo o particular. Esta clasificación solo se puede hacer si existe información adecuada de carácter individual.

La influencia en particular de la red social en el desarrollo económico debemos postularla en diferentes ámbitos. Primero que todo se establece, qué poder particular tiene un nodo no fundamental en relación a su función dentro de la empresa. Al contrario de empresas de sociedad anónima con acciones dispersas y con una poca mayoritaria identificada, las empresas PMYE disponen de una influencia conocida particular y clara que puede ser ejercida por agentes conocidos y determinantes.

Buscando un método para medir la red y función del tiempo se llega a la teoría de grafos (Arbib, Kfoury, & Moll, 1984). Un problema en esa teoría clásica es el de determinar el camino más corto entre dos nodos (Korte & Vygen, 2008). Una idea para medir la red, puede ser determinar la media de todos los caminos más cortos. Este enfoque es, desafortunadamente, para el caso de este estudio inutilizable, ya que la fuente de todos los datos se basa en la comunicación de unos pocos (2-3) elementos o nodos centrales. Esa desventaja se encuentra en

forma parecida al tomar como referencia la centralidad de la red (Borgatti S. P., 2005).

Veamos el problema en forma gráfica:



**Gráfica 2 Visualización de una comunicación por email**

En este grafo se visualiza la red basada en la comunicación por email de un nodo (A). Por cada email a un miembro de la red en forma directa (To:) recibe un valor de 2 la conexión (valor de la arista) y si es conectado indirectamente (Cc: o Bcc:) se le da un valor de 1. Del protocolo de la comunicación por correo electrónico de A, se establece la intensidad (valor) de la comunicación y por tanto su grado nodal. Igualmente se puede determinar la cercanía a los nodos B y C en forma inequívoca. Mas, sin embargo, no es posible determinar si el nodo A, cuestión del análisis es el nodo de mayor influencia de la red,

ya que no existe información sobre la comunicación entre los otros nodos, sino de forma discriminante en cuanto que se puede protocolizar su intercomunicación solamente cuando el nodo A aparece en copia o cuando la comunicación es directa si los nodos adyacentes son tomados en copia vista (CC:). Podemos deducir, que el nodo A tendrá siempre en el análisis el mayor dominio y por lo tanto el resultado de una medición en función de centralidad no puede ser tomado en cuenta.

Existen características de una red (social) que en el trabajo se analizarán independientemente. El objeto es determinar que característica se puede tomar como referencia como medida central de la red. Candidatos son aquellos valores que con ayudada de la herramienta de software UCINET (Borgatti, Everett, & Freeman, 2012) se pueden obtener fácilmente como son el coeficiente de agrupación, reciprocidad, transitividad en sus diferentes expresiones, cantidad de puentes de tipo t-local, o los porcentajes generales de centralidad. En el capítulo II se explican estos coeficientes detalladamente. Otros aspectos que pueden tenerse en cuenta son valores tales como el valor medio de distancia entre dos nodos o la cantidad de triadas, sin embargo, estos últimos por la característica de la red se espera que sean muy similares y no se alteren en factor del tiempo, que es lo

fundamental para lograr una comparación con el desarrollo económico de la empresa a analizar.

Para el manejo de los datos una vez estos hayan sido verificados puntualmente con UCINET, se utiliza la herramienta de software R, que ofrece facultades de programación y de esta forma obtener los valores de referencia en cuestión, sobre el tiempo. Como se trata de un proceso de red en evolución, se utilizan dos mediciones del campo de teoría de grafos (Butts & Carley, 2001) y (Backhaus, Erichson, Plinke, & Weiber, 2011), las cuales son la covariancia y la correlación, definidas por (Butts & Carley, 2001) como

$$Cov(H_i, H_j) = \frac{1}{|Vu|^2} \sum_{x=1}^{|Vu|} \sum_{y=1}^{|Vu|} \left( (\delta_i(x, y) - \overline{\delta_{Hi}}) (\delta_{ji}(x, y) - \overline{\delta_{Hj}}) \right) \quad (1)$$

y

$$Cor(H_i, H_j) = \frac{Cov(H_i, H_j)}{\sqrt{Var(H_i)Var(H_j)}} \quad (2)$$

donde  $H_i$  y  $H_j$  son grafos,  $Vu$  el conjunto de vértices,

$\delta_i(x, y) = \begin{cases} 1 & \text{if } (V_x, V_y) \in E(H) \text{ (conjunto de esquinas)} \\ 0 & \end{cases}$  que permite contar la

cantidad de aristas del grafo y  $Var(H) = Cov(H, H)$ .

En el campo académico se abre una nueva forma de ver el desarrollo económico de las empresas PMYE. El catálogo de posibles adaptaciones en este campo en el futuro, está en función de la

posibilidad y capacidad de obtener datos, lo cual supone un reto para un futuro contingente de científicos. Es por ello que la adaptabilidad de la metodología, por su apertura en el planteamiento del algoritmo seguido para la depuración de la información ganada, es parte fundamental y meta del trabajo a ser desarrollado.

### **1.5. Limitaciones**

El problema de establecer un concepto generalizado, es, por aquello de la diversidad tanto en tamaño como en calidad y características de las empresas PYME, un aspecto que se mostrara en el transcurso descriptivo de la solución.

La adquisición de los datos y por consecuencia el posible conflicto con leyes de protección de datos, impide una evaluación puntual de determinados eventos. Es así, que la identificación de clics en la red social, solo puede ser detallada si se conocen la función de los nodos y por hecho el nombre del usuario. Esta limitación es de gran importancia en el desarrollo de las conclusiones, ya que de por si es imposible y por asunto de ley no querido, establecer "culpas" de mala gestión o poca reacción de por ejemplo los dueños de la empresa o del consejo administrativo, si esto condujera a un resultado parecido. Es decir que los resultados obtenidos tienen un sentido macro, general.

Por otro lado, en el proceso de identificación de grupos o cliques, podemos hacer una distinción entre interno y externo. Es entonces posible, solo por este hecho sacar conclusiones que pueden conducir a un resultado más adecuado.

El abarque de los efectos de correo electrónico en masa, como son los boletines informativos y correo clasificado SPAM es objeto principal en el avance de la investigación.

### **1.6. La empresa ejemplo**

Esta investigación tiene como foco a las empresas PMYE, ya que estas constituyen una gran parte de la base de la economía europea (Vetter & Köhler, 2014). Por su diversidad tanto en organización, tamaño y radio de acción, ha sido difícil en el pasado realizar trabajos que den un resultado de amplia generalidad y que pueda ser usado para todo tipo de empresas sin importar su especialidad. El proyecto tiene como plan de trabajo hacer el análisis en cuestión en una empresa, generalizando su resultado. Este será ampliado y evaluado teniendo en cuenta su carácter específico por la característica de la empresa en cuestión, dando resultados de metodología y desarrollo de aplicabilidad a empresas de otra rama o especialización.

## CAPÍTULO I. INTRODUCCIÓN

---

El aporte a la comunidad científica interesada en investigar los efectos de gestión administrativa de empresas PMYE en su desarrollo futuro, es finalmente la especificación de herramientas tales como metodología de la investigación, campo de aplicación y algoritmos para la adquisición de datos en marcos legales.

El Aeropuerto de Muenster y Osnabrueck es una empresa, que por su definición tamaño y balance general se puede clasificar como empresa mediana, sin embargo, por la calidad de los dueños de la empresa se puede considerar como empresa gubernamental, ya que todos ellos son igualmente empresas gubernamentales. Por su forma legal y fuero podemos considerarla ideal para el estudio.

El Aeropuerto de Muenster y Osnabrueck es una "sociedad con responsabilidad limitada"<sup>1</sup> con una estructura central con aproximadamente 120 empleados y una serie de empresas afiliadas especializadas en diferentes áreas tales como el servicio de rampa (100%), el servicio de carga (33%), el servicio de Check-In, tiquetes y abordaje (33%) y el servicio de control de seguridad (100%).

---

<sup>1</sup> Una GmbH es una sociedad con responsabilidad limitada definida en el derecho mercantil alemán desde 1892. (<http://www.comercio.mityc.es/tmpDocsCanalPais/C392056B518401B47300F0DC5C57CCAB.pdf>).



---

económico para la comunidad económica europea. En especial se hace un análisis de la economía PMYE en función de tiempo de acuerdo a diferentes fuentes estadísticas.

En el capítulo cuatro se describen las herramientas de software que fueron usadas durante la investigación, dando una evaluación a cada una de ellas. Se analiza los problemas en la adquisición de los datos necesarios para el análisis y la forma como fueron preparados para poder ser usados en las herramientas previstas. Especialmente las limitaciones de cada una ellas específicamente y el significado de este, es mostrado detalladamente.

En el capítulo quinto se describe el procedimiento del análisis, haciendo una descripción de los datos recolectados. En este capítulo se propone un método para normalizar los datos que componen la red social, deducida de los datos adquiridos por medio del análisis de correo electrónico, para que con ellos se construya la red social. Igualmente se propone una forma de mostrar el desarrollo de los datos de la red social y del desarrollo económico y de productividad de la empresa objeto de estudio.

El capítulo sexto muestra una primera aproximación a un análisis con ayuda del análisis de multivariante. En esta parte de la investigación se propone un modelo y se describe el mismo.

Igualmente se describen los factores influyentes no cuantificables y las posibles ventajas en el uso de este método como herramienta de predicción.

El capítulo séptimo toma los datos ganados en forma tradicional en el capítulo anterior y con ayuda de redes neuronales se busca un método alternativo de comprobar lo hallado. Para ello se busca una red neuronal adecuada y sencilla. La red encontrada comprueba un resultado, pero por la característica de la misma, el camino a ella es de método arbitrario y por consiguiente solo recomendable para efectos de comprobación.

Por último, en el capítulo octavo se relatan las conclusiones y enseñanzas adquiridas durante la investigación. Una parte fundamental del estudio es el posible desarrollo del mismo en un futuro. Es por ello que algunos puntos abiertos detectados durante la investigación mas no tomados en consideración en forma profunda, son mostrados detalladamente, acompañados de posibles recomendaciones para resolver los mismos. Un ejemplo de un punto abierto es la consideración de tomar otros datos electrónicos como son los adquiridos de redes sociales de la web 2.0 como complemento o suplemento de los datos adquiridos de la red de correos electrónicos. Las recomendaciones a seguir son mostradas detalladamente.

Capítulo I	• INTRODUCCIÓN
Capítulo II	• MARCO TEÓRICO – REDES SOCIALES
Capítulo III	• MARCO TEÓRICO – DESARROLLO ECONÓMICO DE EMPRESAS PYME
Capítulo IV	• METODOLOGÍA
Capítulo V	• ANÁLISIS EMPÍRICO
Capítulo VI	• PRIMERA APROXIMACIÓN A UNA CORRELACIÓN
Capítulo VII	• VERIFICACIÓN CON AYUDA DE REDES NEURONALES
Capítulo VIII	• CONCLUSIONES Y RECOMENDACIONES

Gráfica 4 Estructura de la tesis

## **CAPÍTULO II. MARCO TEÓRICO – REDES SOCIALES**

### **2.1. Introducción - Red Social**

La red social es un concepto bien definido en la literatura. Una red es un conjunto de puntos denominados nodos, con conexiones entre ellos denominados vínculos (Borgatti & Lopez-Kidwell, 2011) (Newman M. E., 2010). En el contexto de este estudio, el emisor y el receptor del correo electrónico son los nodos de la red, y el correo electrónico es el vínculo. Como resultado de la multiplicidad de agentes (edad, sexo, etc.) que pueden intervenir en una red social, se espera que las conexiones entre ellos sean heterogéneas (Newman M. E., 2010). Dicha heterogeneidad en la distribución de las conexiones es consecuencia de la afinidad y de las relaciones privilegiadas entre los nodos de la red, que resultan de los diferentes roles que éstos adoptan (Wassermann & Faust, 1999), (Borgatti S. P., 2009), (Borgatti & Foster, 2003). Cuando los nodos comparten similares características interactúan más entre ellos que con aquellos que son diferentes (Coleman, 1986), (Haythornthwaite & Wellman, 1988) y por tanto existe un patrón definido que los vincula preferentemente. Dicha afinidad, siguiendo estos autores, tiene como consecuencia la

existencia de zonas más densas en la red, derivadas de un mayor nivel de interconexión entre los nodos.

La heterogeneidad de las redes sociales justifica su estudio como un sistema complejo (Knoke & Yang, 2008). El estudio de la red social como un sistema complejo permite analizar las reglas estocásticas que influyen en la formación de los vínculos, teniendo en cuenta además de la existencia de algún tipo de aleatoriedad; y los rasgos característicos de los nodos (por ejemplo, su grado de conectividad -número de vínculos- o su centralidad) medida en función de la importancia de un nodo que, a su vez, puede verse afectada por sus vínculos con otros nodos (Coleman, 1986).

Según Jansen (2006) el análisis de redes es un instrumento que puede capturar recursos sociales o capital social. El factor del capital social es descrito por primera vez por Bourdieu en 1983 y coloca este al mismo nivel que el capital económico y cultural. Sin embargo, Bourdieu enfoca sus estudios hacia el capital cultural. El solo hecho de describir el capital social, hace que Bourdieu sea uno de los pioneros en este campo. Weyer (2000) establece que, en las sociedades modernas, las redes sociales tienen una gran influencia en la generación de soluciones y que la sociedad cree que estas son más eficaces que por ejemplo la burocracia común. Weyer aclara en su trabajo sobre conceptos y métodos en investigación de redes en

ciencias sociales, que las redes sociales enlazan la flexibilidad y la eficiencia en una forma única. Weyer en sus trabajos iniciales en 1993 afirma que la red social es un fenómeno emergente que sigue sus propias reglas y que ninguno de los actores de esa red controla en forma exclusiva y dicha red puede desarrollar una dinámica propia.

En el año 2011 en el R Journal (Bohn, Feinerer, Hornik, & Mair, 2011) un grupo de investigadores publicaron una aproximación en un análisis de redes sociales usando el intercambio de correo electrónico en una red específica. Si bien el objetivo de esa investigación no es una comparación con elementos de tipo económico, el hecho de obtener información mediante extracción de metadatos del tráfico de correo electrónico, muestra una forma para su uso. En ese estudio se evalúa con ayuda de herramientas del paquete de software "R" la posible relación entre personas que tienen interés parecido y su intercomunicación. Sin embargo, en su conclusión llegan a un punto que es de suma importancia para el desarrollo de este trabajo: La actividad de los actores de la red específica solo puede tenerse en cuenta si la actividad de los mismos es de forma parecida. De tal manera, que, si un autor no es muy comunicativo en la red a ser analizada, aun teniendo en cuenta que puede ser básica su contribución a un tema específico, su importancia no podrá ser definida. Igualmente, en ese estudio se tiene en cuenta el hecho que

el contenido de los correos analizados, no se puede usar, ya que este contiene mucha información inútil.

Los muchos estudios el foco principal ha sido primordialmente la extracción de información de las páginas de internet. Es así como Xi et al. (2002) hace un enfoque a una máquina de aprendizaje utilizando árboles de decisión y de regresión logística, con resultados moderadamente bajos y Upstill, Craswell y Hawking (2003) describen una técnica de recuperación de información que aumenta la recuperación de contenido basado en el análisis de enlace (el PageRank de la búsqueda de Google) y las características de las URL.

En un trabajo de Borkar, Deshmukh, y Sarawagi (2000) relacionado con la extracción de registros de contactos obtienen una alta precisión de extracción de datos de las páginas de internet, buscando y usando un número limitado de información (número de casa, calle o carretera, ciudad, estado y código postal) que por lo general aparecen en forma muy constante y formalizada.

Una investigación similar de redes sociales fue llevada a cabo con datos de Usenet (Smith M. , 1999), cuyo objetivo fue el de caracterizar una comunidad en forma dinámica, así como de determinar la "autoridad" de un individuo en particular basándose en sus aportes.

También ha habido trabajos que consideran en el análisis de gráficos de coautoría (Newman, 2001), que son similares al trabajo de Culotta, Bekkermann y McCallum (2004) que se basan en la extracción de nombres de coautor de las páginas del publicaciones científicas.

## **2.2. Análisis de Redes Sociales**

### **2.2.1. Conceptos básicos**

En el campo del análisis de redes sociales, hay algunos conceptos básicos que son importantes de conocer. Estos son actores, relaciones, díada, tríada, subgrupo, grupo, relación y la red (Wassermann & Faust, 1999).

□ **Actor:** los nodos para ser examinados de la red se conocen como actores, sus relaciones y sus consecuencias son examinados en el contexto del análisis de redes sociales.

□ **Relaciones:** Los actores están en una relación social vinculada, en lo que respecta a su contenido, su intensidad o forma diferentes. Los ejemplos de las relaciones son relaciones personales (amistad, respeto, etc.), de afiliación (grupos, departamentos, etc.), las relaciones formales de poder (autoridad, la relación de supervisor, etc.) o las conexiones físicas (ríos, carreteras, etc.).

□ **Díada**: una díada es una conexión o relación entre dos actores y, por lo tanto, la unidad más pequeña posible en el análisis de redes sociales (2.2.5.1).

□ **Tríada**: la tríada muestra la conexión de tres actores entre sí, con todas las conexiones posibles y las relaciones entre ellos. Al igual que la díada, es considerado como parte fundamental de la investigación de una red social (2.2.5.2).

□ **Grupo**: En un grupo se concentran todos los actores, cuya relación común se quiere investigar (2.2.5.4).

□ **Subgrupo**: Es un subconjunto de los actores de una red cuyas relaciones se pueden agrupar.

□ **Relación**: A la colección de relaciones de un tipo determinado, entre un grupo de actores se le llama relación. Un ejemplo son las amistades de los trabajadores individuales dentro de una empresa.

□ **Red**: El concepto de red incluye a todos los actores, los grupos y sus relaciones que son objeto de estudio.

Algunos de estos conceptos básicos así como conceptos métricos y de análisis de redes sociales, son descritos detalladamente más adelante (2.2.5).

### **2.2.2. En la historia**

A parte de los descrito en el capítulo anterior, David Knoke y Song Yang (Knoke & Yang, 2008) definen una red social como una estructura compuesta de un conjunto de actores, donde algunos de sus miembros están conectados por un conjunto de una o más relaciones. El trabajo principal del análisis de redes sociales es identificar la forma de estas relaciones y encontrar si las relaciones son dirigidas o no dirigidas, o la intensidad que pueden tener en su interacción.

Cabe mencionar que el desarrollo del análisis de redes sociales ha ido evolucionando a través del tiempo. Inicialmente como parte de estudios de cognición y psicología social de Wolfgang Kohler (Kohler, 1924 (Org.)), que en forma interesante realizó un estudio con chimpancés, postuló los posibles problemas que se encuentran al pasar de experimentos en circunstancias de laboratorio y cuando éstos son realizados sin ese perfecto entorno. La teoría postulada por Kohler mostró la importancia de las relaciones informales e interpersonales en un sistema social (Scott, 2000).

El desafío en el análisis de redes sociales, es encontrar una cantidad de datos confiables y accesibles para analizar y poder llevar a cabo la investigación. Para obtener información general, científicos utilizan cuestionarios, pidiendo a los actores a ser incluidos en el análisis responder preguntas formalizadas. Los cuestionarios más

comunes para este efecto fueron discutidos en los años 80 por Burt y Marsden (Burt, 1992) y se han utilizado en muchas encuestas. Otra forma de obtener los datos a analizar es el uso de documentos de archivo como en las páginas web de Internet, las citas de patentes, artículos y referencias, etc. Internet es el objeto de importante número de investigaciones (Adamic & Adar, 2003), buscando por ejemplo contenido específico de determinadas páginas web.

Actualmente se busca poner el foco a la información extraída del tráfico de correo electrónico. Una gran cantidad de investigadores utilizan cartas o material escrito para investigar las estructuras sociales en el pasado. Por ejemplo leer el contenido de las cartas a y de Cicerón en la antigua Roma en los años 44 A.C. es parte del análisis que hace (Standage, 2013) de la red social y la influencia de aquel personaje en Roma. Tratar de hacer lo mismo con los mensajes de correo electrónico en la actualidad no es fácil. En primer lugar, porque la cantidad de los mismos es muy alta y en segundo lugar por la privacidad de información o las leyes de privacidad de datos que hacen que sea muy difícil o prácticamente imposible.

Existen bases de datos formalizadas que pueden ser usadas a efectos de la investigación y disponibles para todos. La más conocida es la base de datos de correos electrónicos de Enron, publicada por William Cohen de la Universidad de Carnegie-Mellon y que se distribuyó

inicialmente en 2004. Esta fuente ha sido ya corregida en algunos aspectos. Por ejemplo, algunas direcciones se han corregido en `user@enron.com` en vez de la menos adecuada `no_address@enron.com`. La base de datos actualmente se puede obtener en Internet <http://www-2.cs.cmu.edu/~Enron/>.

En la literatura se diferencian varios tipos de análisis de la información (Scott, 2000). En términos generales podemos diferenciar el análisis de las redes en tres niveles:

- Análisis de los actores solitarios en la red
- Análisis de grupos en la red
- Análisis general de la red o sociedad

Este estudio se basa por sus características y disposición de datos en el análisis general de la red social.

### **2.2.3. Análisis de redes como análisis estructural**

Una frase muy común en el ámbito de la perspectiva en el análisis de redes en estructuras sociales es “El todo es más que la suma de sus partes”. O sea que tenemos que analizar la red como todo para explicar el comportamiento de sus partes, tales como sus actores, individuos o elementos. Existe la plena convicción que la red tiene

características que los miembros de la misma no tienen. Estas características son llamadas emergentes y surgen en nivel de agregación superior. Esto no significa que los actores de niveles inferiores de agregación sean independientes. Esa relación entre los individuos y su proceder y de las estructuras, instituciones o sistemas a un nivel macro es el problema fundamental de la teoría sociológica (Parsons, 1976).

Por otro lado, existe la perspectiva micro, donde el proceder del individuo como tal juega un punto central (Homans, 1972); (Hummell & Opp, 1972). El análisis de la red es en este contexto, en cuanto a la agregación, menos importante que el planteamiento del individuo como producto o no de la fuerza de la red. Estas dos perspectivas Micro-Macro y una posible integración de las mismas han sido criticadas por Coleman (1986), Wilson (1982) y Esser (1987).

En la investigación empírica se centra el comportamiento individual como variable dependiente. Los resultados de esta para la teoría macro-sociológica son considerados muy limitados, ya que sustituye la acción por el comportamiento y poco se interesa en las consecuencias de la acción individual a los fenómenos macro.

El análisis de red es por naturaleza afín al concepto macro, debido a sus características y dinámicas a niveles de agregación. Una diferencia básica entre el análisis de redes y el método clásico de

estudio científico es, que no está en primera línea una ordenación de variables y su relación como es buscado en diagramas causales, sino ante todo una ordenación relacional de individuos que es mostrado generalmente en grafos. En el centro del análisis no están las características individuales de los miembros de la red, sino las características relacionales. El individuo como tal no interesa sino su relación con otros y su incrustación en una estructura son motivo de la investigación (Jansen D. , 2006)

#### **2.2.4. Capital social y análisis de redes**

Bajo el término de capital social se entiende un aspecto de la estructura social, que abre a actores individuales o corporativos una posibilidad de acción, que les permite ganancias o les facilita la coordinación de planes de acción colectivos (Coleman, 1998), (Bourdieu, 1983). El análisis de redes es un instrumento que puede registrar recursos sociales o capital social. Por ejemplo, se abren perspectivas para mejor manejo del beneficio o para mejor coordinación de acciones de intención (Hurlbert, Beggs, & Haines, 2008).

El carácter estructural del capital social, requiere que el proceso de su producción a menudo no sea consciente. A veces se produce casualmente junto con otras acciones (Jansen D. , 2007). Si bien esto tiene la ventaja de que el capital social se produce sin ningún

coste adicional, tiene también la desventaja de que sólo puede ser producido específicamente.

Esta autonomía estructural que Burt (1992) postuló consecuentemente, da un poder a la red basada en su capital social. Burt en su estudio también hace una diferenciación entre el poder en el contexto de influencia social y poder en el sentido de autonomía estructural. Esta diferencia ya había sido descrita por Weber cuando describió la diferencia entre dominación (caudillaje) y mero poder per se. Con la disposición del capital social, es decir a través de la incorporación en redes con características específicas, el poder y la influencia de los individuos y los agentes empresariales pueden explicarse. Las redes son entonces un factor de competitividad tanto a nivel de las organizaciones como en agrupaciones organizativas o sociedades enteras.

Qué tipo de estructuras de redes suministran beneficios es bastante controvertido (Jansen D. , 2007). ¿Es la posición de corredor (Burt, 1992) con sus muchas relaciones débiles e indirectas en el borde de clúster densos que son los que dan ventajas a las organizaciones o? ¿O son las redes densas las que permiten una cooperación confiada y reducen los costos de transacción? Así que ¿Cuál es el papel, por ejemplo, de la red individual de un empleado o gestor en la red social

de la empresa y que influencia tienen al éxito y la influencia en la organización?

### **2.2.5. Indicadores en el análisis de redes sociales**

Existe un gran número de conceptos dentro del Análisis de Redes Sociales que es necesario conocer para poder investigarlas. Algunos son básicos, como actor y conexiones, y otros resultan de la combinación de estos conceptos básicos como son díadas, tríadas, grupos y subgrupos. Existen otros conceptos que, incluyendo de la misma forma actor y conexiones, se relacionan con el nivel de análisis.

Los elementos básicos ya los he nombrado varias veces: actores, redes y sus relaciones (Rürup, Röbbken, Emmerich, & Dunkake, 2015). Estos son los elementos indispensables para formar una red. La combinación entre los actores y sus relaciones conforman la red social. **Actores** pueden ser personas, colectivos, organizaciones, animales, naciones, palabras etc. **Relaciones** también llamadas “conexiones”, pueden ser también de diferente naturaleza de acuerdo al caso a analizar: A nivel individual, por ejemplo, mostrando una relación familiar, a nivel de cognición (saber, alguna vez ha oído hablar de), de relación afectiva (respeto, confianza en), interactividades comunes (ir a navegar, miembros del mismo grupo que participan en el mismo

evento), o roles sociales tales como la amistad o relación de trabajo. Entre organizaciones pueden existir intercambios entre sus miembros. A través de sus empleados pueden estar organizaciones en relación, ya sea entre sus miembros (aeropuerto A y aeropuerto B intercambian información en grupos de interés colectivo) o a través de una membresía paralela (empleados trabajan en más de una empresa). Relaciones entre empresas pueden ser también a través de subcontrataciones, relaciones de ventas, las relaciones de propiedad o alianzas. En las relaciones entre organizaciones juegan a menudo un papel importante el flujo de recursos y de información, así como jerarquía y relaciones de cooperación. También pueden jugar un rol significativo la amistad, el parentesco o las relaciones interactivas.

El último elemento fundamental, la **red**, es la forma de mostrar los actores y sus relaciones, de tal manera que para el objeto de estudio pueda ser visible.

Los datos de una red son generalmente puestos en una matriz rectangular en donde las intersecciones entre las columnas y las filas muestran la interactividad entre los actores.

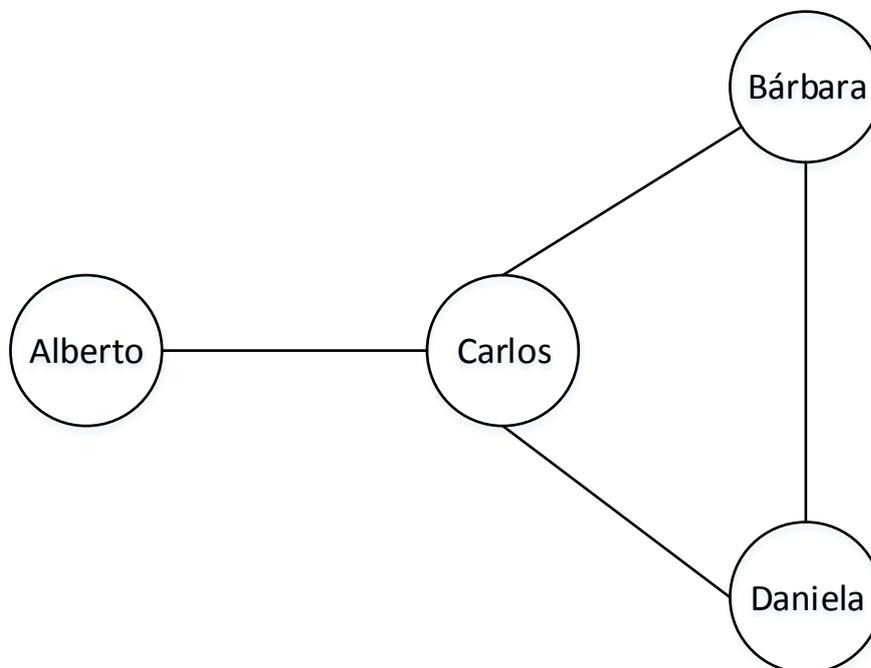
Nombre	Edad	Sexo	Cantidad de conocidos
Alberto	32	M	25
Bárbara	28	F	44
Carlos	41	M	16
Daniela	33	F	27

**Tabla 2 Ejemplo análisis clásico de características**

	Alberto	Bárbara	Carlos	Daniela
Alberto	X	0	1	0
Bárbara	0	X	0	1
Carlos	1	1	X	1
Daniela	0	0	1	X

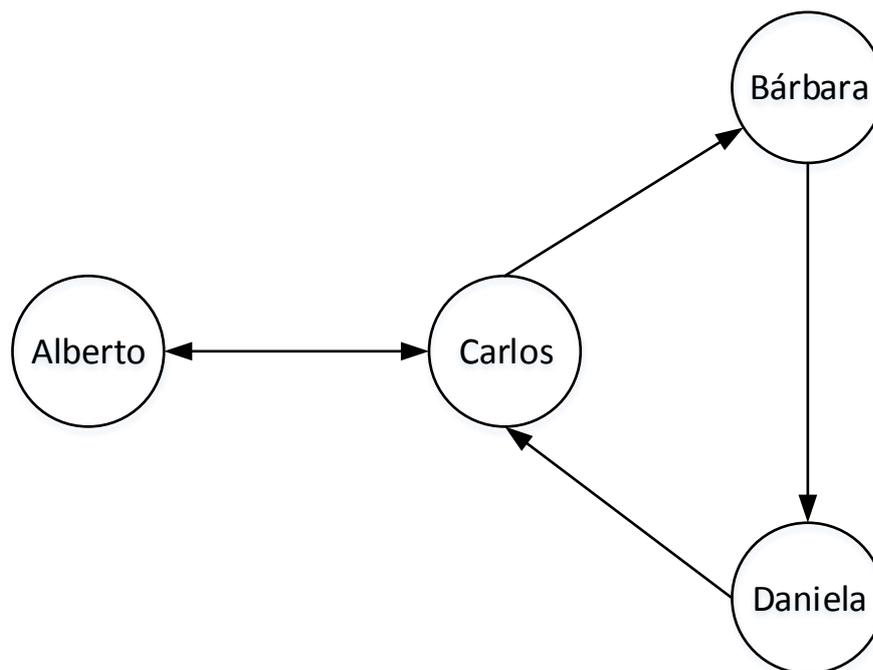
**Tabla 3 Ejemplo red social: ¿quién llama a quién?**

Con la ayuda de grafos se pueden visualizar las redes sociales. Estos son llamados también sociogramas. Es así que podemos visualizar la Tabla 3 con conexiones no dirigidas:



**Gráfica 5 Ejemplo red con conexiones no dirigidas**

o con conexiones dirigidas:



**Gráfica 6 Ejemplo de red con conexiones dirigidas**

El objeto de análisis no es siempre la red sino también la relación entre sus actores, la cual se pueden diferenciar entre contenido, intensidad y forma (Knoke & Yang, 2008), sin embargo, el foco lo colocaremos en la red y sus características.

En el campo de análisis de redes se diferencian cinco niveles que van más allá del nivel del actor: díada, tríada, grado de centralidad, grupos/subgrupos y densidad (Jansen D. , 2006). La centralidad, si bien es cierto que es uno de los componentes más utilizados en el Análisis de Redes Sociales, debido a que el objetivo principal de este concepto, es hacer referencia a la posición del actor en la red (Borgatti, Everett, & Freeman, 2002), en este estudio no se tendrá en cuenta.

### 2.2.5.1. Díada

La díada viene definida por un par de actores y la posible conexión entre ellos (Wassermann & Faust, 2007). Atendiendo a esta definición existen cuatro formas o posibilidades de conexión tal y como se observa en la Gráfica 6: Alberto y Daniela no conectan, es decir existe una carencia de díada o díada nula, Carlos conecta con Bárbara y Bárbara conecta con Daniela estas son díadas asimétricas y en el caso de Carlos y Alberto que se conectan simultáneamente se denomina una díada mutua o recíproca.

En una red con N actores existen

$$\binom{N}{2} = \frac{(N^2 - N)}{2} \quad (3)$$

diferente de díadas.

En el contexto de la red en general se mide la cantidad de díadas de cada tipo. En nuestro ejemplo la Gráfica 6 vemos el siguiente resultado:

Díadas nulas ( $D_{ij}=(0,0)$ )	2
Díadas asimétricas ( $D_{ij}=(1,0)$ y ( $D_{ij}=(0,1)$ )	3
Díadas simétricas ( $D_{ij}=(1,1)$ )	1

### 2.2.5.2. Tríada

El segundo nivel de análisis de redes es la tríada que es definida como un subconjunto de la red social compuesto por tres actores y los posibles lazos entre ellos (Knoke & Yang, 2008). Las tríadas dan lugar a más tipologías que las díadas. Hanneman (2005) contabiliza 64 ( $2^6$ ) diferentes cuando las conexiones son dirigidas y cada actor es tomado en el concepto como diferente. En el caso contrario son solo 16 formas diferentes de conexiones (Gráfica 7). La calculación de estructuras de diferentes clases de tríadas en una red se denomina **censo de tríadas**.

Una red con N elementos tiene

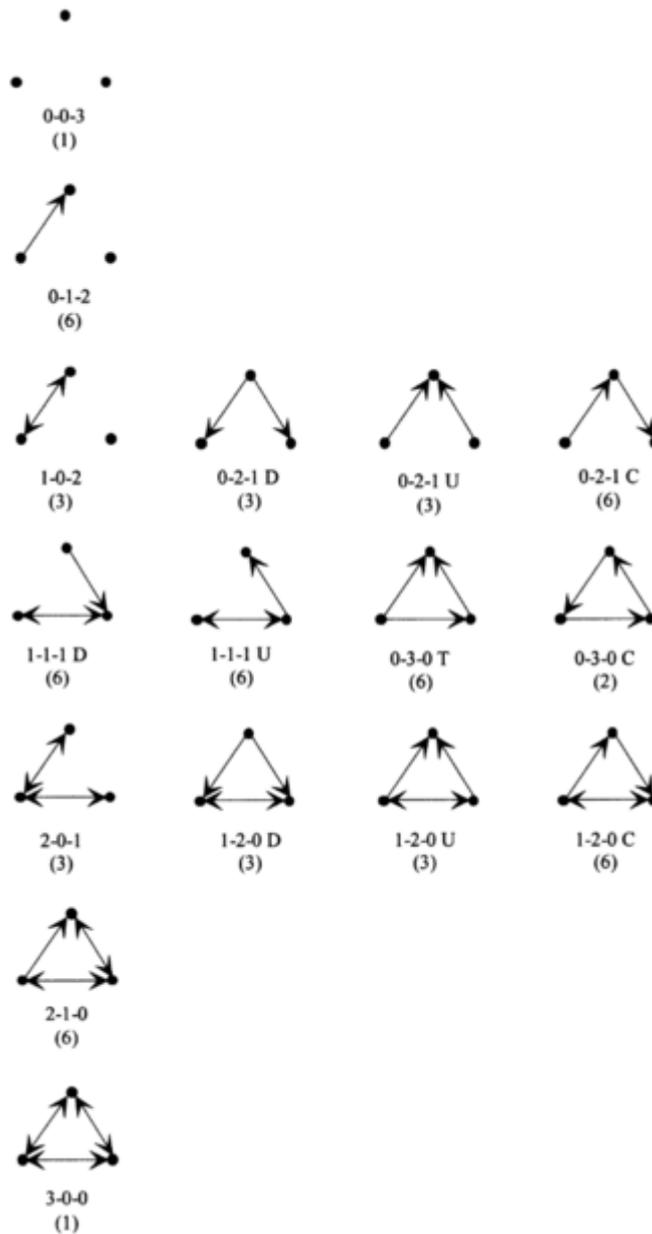
$$\binom{N}{3} = \frac{1}{6} * N * (N - 1) * (N - 2) \quad (4)$$

tríadas.

Tanto para deducir la cantidad de díadas como de triadas en una red, se utiliza el término  $\binom{n}{k}$  que se conoce como un factor en la ecuación binomial y se calcula así:

$$\frac{n!}{(n-k)!*k!} \quad (5)$$

El término  $\binom{n}{k}$  indica cuántas posibilidades hay de elegir  $n$  elementos de un número total  $k$  de elementos diferentes, siendo siempre  $n > k$ .



Gráfica 7 Las 16 diferentes clases de triadas en M-A-N notación según P. W.

Holland y S. Leinhard

Volviendo a nuestro ejemplo en la Gráfica 6, obtenemos el siguiente censo de triadas:

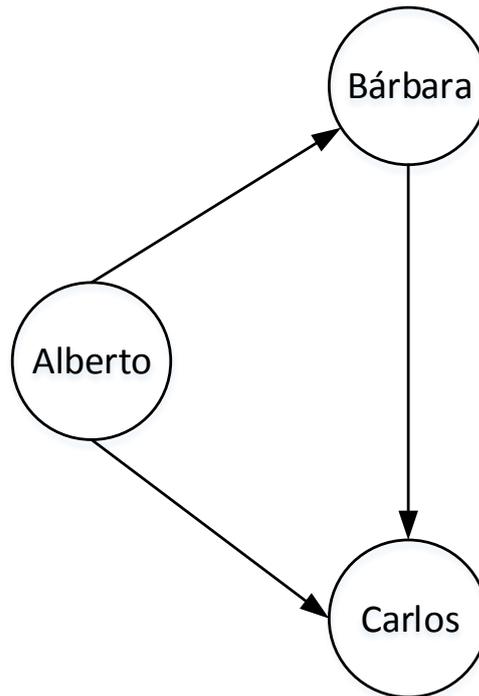
Tríadas 003	0
Tríadas 012	1
Tríadas 102	0
Tríadas 021D	0
Tríadas 021U	0
Tríadas 021C	0
Tríadas 111D	1
Tríadas 111U	1
Tríadas 030T	0
Tríadas 030C	0
Tríadas 201	0
Tríadas 120D	0
Tríadas 120U	0
Tríadas 120C	1
Tríadas 210	0
Tríadas 300	0
Suma	4

Con respecto a las triadas se pueden observar un nuevo parámetro: **Transitividad** (Jansen D. , 2007). Una triada transitiva es aquella que se forma entre tres nodos implicando que si por ejemplo Alberto conecta con Bárbara y Bárbara conecta con Carlos entonces Alberto conectaría con Carlos.

Tríadas son según Jansen en diferente grado transitivas. Esto debido a que una triada tiene 6 diferentes tripletes. Estos son los diferentes métodos en que se pueden agrupar: ABC, ACB, BAC, BCA,

CAB y CBA. Si solo uno de esa triplete es intransitiva, se dice que la tríada en si es intransitiva. Existen casos en los cuales la transitividad de la tríada no es explicita pero que el contexto no la contradice. Igualmente, según la forma en que se hace la observación puede o no una tríada ser catalogada como transitiva o no. Por ejemplo si tomamos la estructura 2-1-0 (Gráfica 7) y enumeramos los actores de la siguiente forma:  $a \leftrightarrow b, a \rightarrow c, b \leftrightarrow c$  entonces tenemos que

1.  $a$  conecta a  $b$ ,  $b$  conecta a  $c$  y  $a$  conecta a  $c$ . Esto significa que la triplete es transitiva
2.  $a$  conecta a  $c$ ,  $c$  conecta a  $b$  y  $a$  conecta a  $b$ . Esto significa que la triplete también es transitiva
3.  $b$  conecta a  $c$ ,  $c$  no conecta a  $a$ . Esto significa que no se sabe si existe transitividad o no
4.  $b$  conecta a  $a$ ,  $a$  conecta a  $c$  y  $c$  conecta a  $b$ . Esto significa que la triplete también es transitiva
5.  $c$  conecta a  $b$ ,  $b$  conecta a  $a$  y  $c$  no conecta a  $a$ . Esto significa, que no se sabe si existe transitividad o no
6.  $c$  no conecta a  $a$ . Esto significa que la triplete no es transitiva

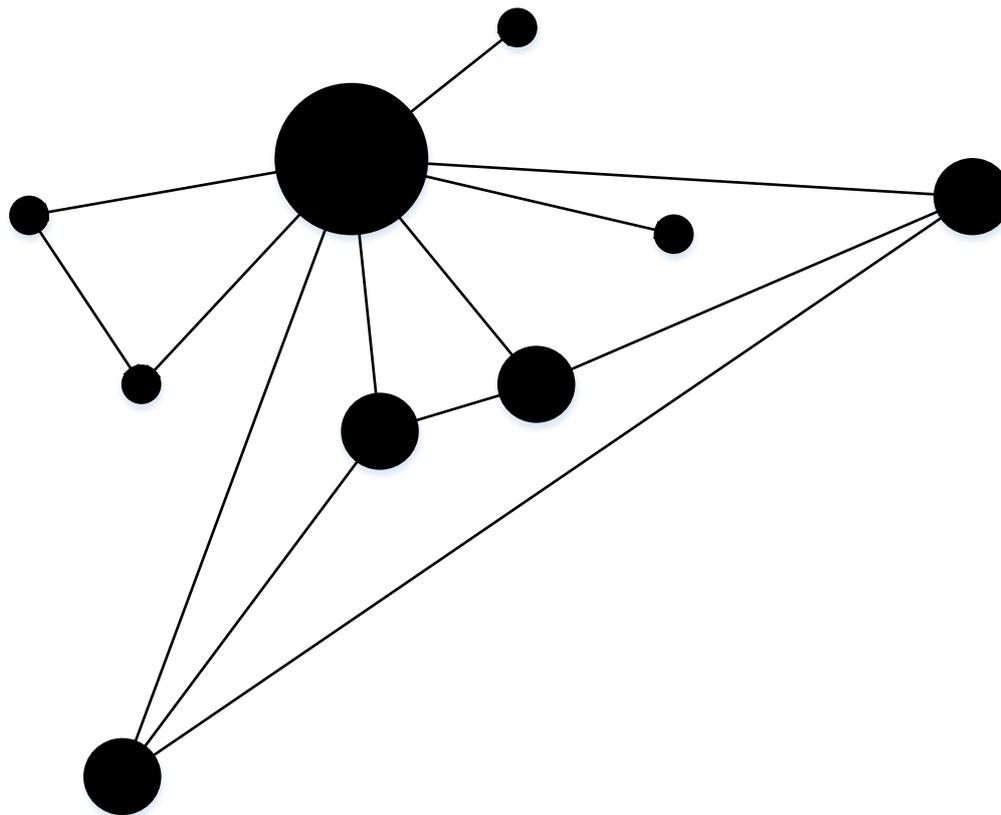


**Gráfica 8 Tríada transitiva**

El concepto de transitividad se usa para establecer grupos o cliques y la fuerza y tamaño de los mismos.

### **2.2.5.3. Grado**

Para entender ese coeficiente de una red social, tenemos que ver primero el concepto de centralidad que ya había nombrado anteriormente. Según (Knoke & Yang, 2008) al usar la teoría de grafos en el análisis de redes sociales se busca identificar los actores más importantes o prominentes. La importancia de un actor refleja su gran visibilidad sobre los otros actores de la red:



**Gráfica 9 Visualización de un actor importante**

Los casos más populares en la computación de centralidad son según Freeman (1977) (1979) son grado, cercanía e intermediación.

Grado o grado nodal o grado de vértice, como se denomina en diferentes literaturas, es definido por Freeman (1979), como el número de vértices adyacentes a un vértice dado en un gráfico simétrico. Para datos no simétricos, el grado de entrada de un vértice  $u$  es el número de contactos recibidos por  $u$  y el grado de salida es el número de lazos iniciados por  $u$ . Además, si los datos son valorados, los grados (de entrada y salida) consistirán en las sumas de los valores

de los contactos. La centralidad de grado normalizada, es el grado dividido por el máximo grado posible expresado como un porcentaje. Los valores normalizados se utilizan solamente con datos binarios (Borgatti, Everett, & Freeman, 2012).

Para calcular el grado de centralidad de la red global se calcula primero el grado de centralidad de cada vértice. En una red binaria dada con vértices  $v_1 \dots v_n$  y un grado máximo de centralidad  $c_{max}$ , el grado o grado de centralidad de la red se define como

$$c = \frac{\sum(c_{max} - c(v_i))}{\max \sum(c_{max} - c(v_i))} \quad (6)$$

Si aplicamos esta fórmula a nuestro ejemplo (Gráfica 6) con ayuda del programa UCINET 6 (4.2.3), se obtiene el siguiente resultado:

```

FREEMAN'S DEGREE CENTRALITY MEASURES
-----
Diagonal valid?          NO
Model:                   ASYMMETRIC
Input dataset:           Example-2016 (Y:\Doctorado\FINAL 2016\Example-2016)

      1          2          3          4
      OutDegree  InDegree  NrmOutDeg  NrmInDeg
-----
2  B          2.000      2.000      66.667    66.667
1  A          1.000      1.000      33.333    33.333
3  C          1.000      1.000      33.333    33.333
4  D          1.000      1.000      33.333    33.333
    
```

DESCRIPTIVE STATISTICS

		1	2	3	4
		OutDegree	InDegree	NrmOutDeg	NrmInDeg
		-----	-----	-----	-----
1	Mean	1.250	1.250	41.667	41.667
2	Std Dev	0.433	0.433	14.434	14.434
3	Sum	5.000	5.000	166.667	166.667
4	Variance	0.188	0.188	208.333	208.333
5	SSQ	7.000	7.000	7777.777	7777.777
6	MCSSQ	0.750	0.750	833.333	833.333
7	Euc Norm	2.646	2.646	88.192	88.192
8	Minimum	1.000	1.000	33.333	33.333
9	Maximum	2.000	2.000	66.667	66.667
10	N of Obs	4.000	4.000	4.000	4.000

Network Centralization (Outdegree) = 33.333%

Network Centralization (Indegree) = 33.333%

Actor-by-centrality matrix saved as dataset Y:\Doctorado\FINAL 2016\Example-2016-deg

-----  
 Running time: 00:00:01  
 Output generated: 13 Mrz 16 19:53:13  
 Copyright (c) 2002-12 Analytic Technologies

### 2.2.5.4. Grupos y subgrupos /Cliques

El concepto de grupo social puede ser utilizado para estudiar las propiedades de un conjunto de actores que están dentro de la red (Jansen D. , 2006). En general dice Jansen, que el análisis de subgrupos o cliques se basa en la identificación de cohesión<sup>2</sup> de subgrupos dentro de una red. Clique se entiende en la misma forma como se usa normalmente: un conjunto abarcable de actores, muchas veces con relaciones directas y estrechas con intereses comunes que puede ser identificado claramente. En sociología este tipo de cliques son también denominados como grupos.

---

<sup>2</sup> Por cohesión de subgrupos se entiende "un conjunto de actores entre quienes hay unos lazos relativamente fuertes, directos, intensos, frecuentes o positivos" (Knoke & Yang, 2008, pág. 72)

### 2.2.5.5. Densidad

Mitchell (1969) introdujo una serie derivados de la transición de la teoría de grafos al lenguaje sociológico, que puede ser usado en el contexto de redes sociales. Densidad es uno de ellos. Mitchell ve Densidad como la completad de la red. El término define el número total de conexiones existentes en relación al número máximo posible que podrían existir en la red. Anteriormente Barnes (1954) y Bott (1955) habían tratado de explicar esto con otras notaciones tales como mallas o conectividad.

La ecuación para medir la densidad es entonces

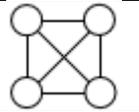
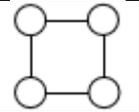
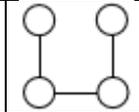
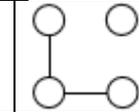
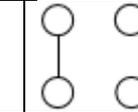
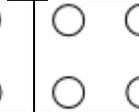
$$d = \frac{l}{n(n-1)/2} \quad (7)$$

Donde  $l$  es el número de líneas (conexiones, relaciones etc.) presentes y  $n$  es el número de puntos (actores).

En el caso de un grafo con conexiones dirigidas la densidad se calcula algo diferente. La matriz con datos dirigidos es asimétrica, ya que la línea de  $a$  a  $b$  no necesariamente es recíproca. Por esta razón el número máximo de conexiones es calculado como el número de pares que contiene  $n(n - 1)$ . Por lo tanto, la ecuación para una red con conexiones es

$$d = \frac{l}{n(n-1)} \quad (8)$$

En la siguiente gráfica podemos observar los diferentes valores de grado y densidad en una red con cuatro actores (puntos):

						
Nº de puntos conectados	4	4	4	3	2	0
Inclusividad <sup>3</sup>	1,0	1,0	1,0	0,75	0,5	0
Grado	12	8	6	4	2	0
Nº de líneas	6	4	3	2	1	0
Densidad	1,0	0,7	0,5	0,3	0,1	0

**Tabla 4 Comparación de densidad y otros valores en una red**

### 2.2.6. Mediciones de la evolución de red social

En un estudio empírico sobre la evolución de una red social (Kossinets & Watts, 2006) se muestran cómo es posible utilizar algunos valores de la red y no exclusivos de un actor de esa red para observar los cambios de la misma en razón del tiempo. En su estudio se analiza el desarrollo de la misma basada en el intercambio de correo electrónico entre más de 43.000 estudiantes y miembros de una universidad durante un año académico. Los miembros de la red fueron complementados con información adicional tal como afiliaciones y atributos escolares y de género. Según el estudio, se utilizó solamente la cabecera de los correos electrónicos. Dentro de las medidas tomadas

<sup>3</sup> Inclusividad es definida como el número total de actores menos el número de actores aislados en porcentaje (Scott, 1991, pág. 70)

para disminuir el monto de información, está el hecho de filtrar aquellos mensajes con más de 4 receptores. Kossinets y Watts sostienen que el resultado de ese análisis también tiene implicaciones en el análisis de la red transversalmente, basándose en la suposición de que las propiedades de la red están en equilibrio (Doreian, 1997). Entre los valores utilizados para medir los cambios de la red están el promedio del grado de vértice, el tamaño fraccional del mayor componente, la media de longitud del camino más corto y además el coeficiente de agrupación. Los investigadores tratan de evitar falsos desarrollos o valores que son específicos en determinados tiempos, como lo es en su caso el fin de semestre. Estas presunciones son de una gran importancia en todo tipo de observaciones.

### **2.2.6.1. Tipos de redes / modo**

Según el tipo de actores que conforman la red y las propiedades de las conexiones, una red se puede clasificar de forma que es denominada **modo**. El modo es el número de conjuntos de entidades en las cuales son medidas las variables estructurales en función de las conexiones establecidas (Borgatti, Everett, & Freeman, 2012), (Wassermann & Faust, 2007) . Los modos más usuales usado especialmente en los diferentes paquetes de software son el modo uno y el modo dos, pero existen más posibilidades según el conjunto de actores, y las clases de entidades que participen en el análisis.

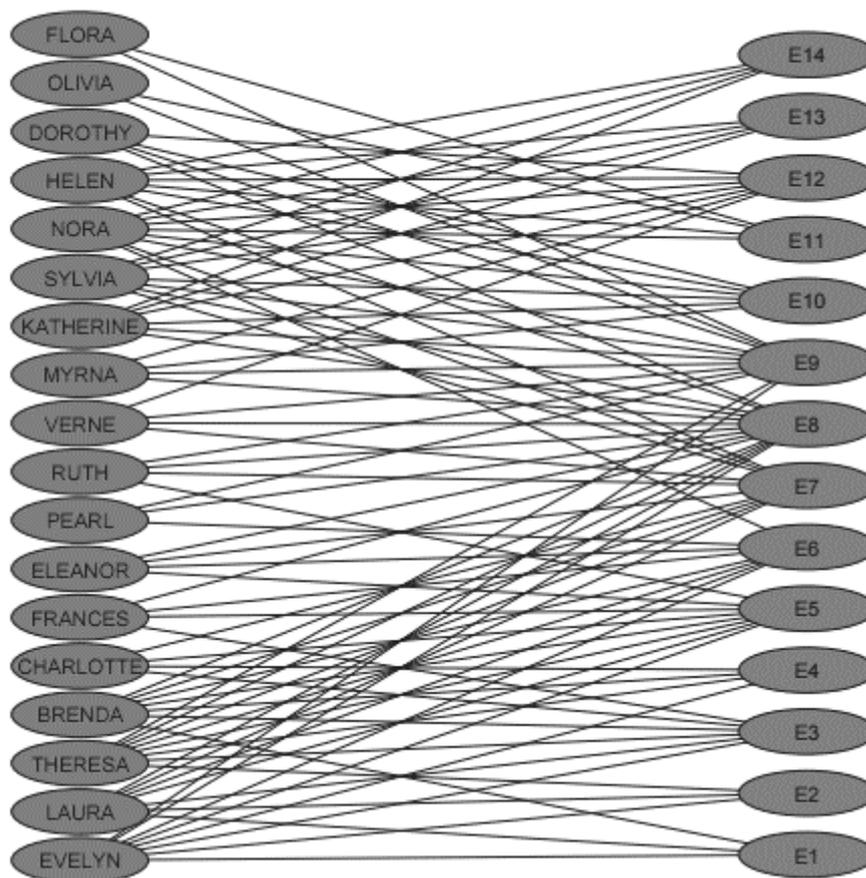
Las redes tipo modo uno, están compuestas por un único conjunto de entidades o actores.

Una red de dos modos contiene dos conjuntos de unidades (por ejemplo, gente y acontecimientos), la relación conecta a los dos conjuntos, por ejemplo, su participación en los eventos sociales:

Existen varias redes de dos modos:

- Membresía en instituciones  $\Rightarrow$  personas son miembros de diferentes instituciones
- Votación  $\Rightarrow$  personas, políticos, votos por sugerencias
- Compra de artículos  $\Rightarrow$  consumidores compran artículos
- Lectores  $\Rightarrow$  Lectores leen revistas
- Redes de coautoría  $\Rightarrow$  autor es un (co) autor de artículo

Un gráfico de una red de modo 2 se le denomina grafo bipartito. Las líneas conectan únicamente los vértices de uno a otro conjunto, mientras que dentro del grupo no hay conexiones.



**Gráfica 10 Ejemplo de una red social modo dos<sup>4</sup>**

En cuanto al peso o influencia en las relaciones establecidas entre los pares de actores, estas pueden ser medidas en términos positivos o negativos, o atendiendo al grado de conexión.

Existen estrategias para convertir redes de modo 2 a modo 1. Redes de modo dos, rara vez se analizan sin transformarlas. Esto se debe a que la mayoría de las medidas de red se definen únicamente para las redes de modo uno, y sólo unos pocos de ellos se han

---

<sup>4</sup> Fuente: A. Mrvar: Network Analysis using Pajek

redefinido para las redes de dos modos (Borgatti & Everett, Network analysis of 2-mode data, 1997). La transformación de una red de modo dos, en una red de modo uno, se hace a menudo con un método conocido como proyección. Este método funciona mediante la selección de uno de los dos conjuntos de nodos (a menudo el conjunto de nodos primario) y la vinculación de los nodos entre sí, sí al menos estos estuvieran conectados al menos aun nodo común en el otro conjunto. Aunque la estructura de dos modos se descarta en este proceso, es posible definir los pesos de unión basadas en ella. Específicamente, los pesos de las conexiones se definen a menudo como el número de nodos comunes. Este método fue extendido originalmente por Newman (2001), quien argumentó que los pesos de unión entre los autores en redes de colaboración científica deben ser descontados si los autores colaboraron en papeles con muchos otros.

#### **2.2.6.2. Tipos de redes / contexto global**

Dos importantes puntos de enfoque en el análisis de redes sociales fue desarrollado por Barnes en 1974 (Wilhelm, 2000):

- Red socio céntrica: Esta se centra en los patrones de relación en un conjunto para ser examinado. Describe una estructura social distinta para cada contexto institucional. Es decir, si existe un conjunto de actores a partir de los cuales se estudian las

relaciones entre ellos, cada relación origina una red diferente (Molina, 2004).

- En general estas redes pueden ser mostradas como un grafo  $G = (V, E)$  donde  $V$  es el conjunto de todos los Actores y  $E = \{(u, v): u, v \in V\}$  el conjunto de todas las relaciones o conexiones entre  $u$  y  $v$ . Esto es denominado en general un Grafo social (Krumke & Noltemeier, 2012).
- Red egocéntrica: En la cual se estudian las conexiones directas de un actor con otros actores, así como los vínculos entre esos actores entre sí, pero no las conexiones hacia el exterior.
  - La llamada red egocéntrica es en general una forma especial de red personal, que puede ser establecida por medio de investigaciones basadas en encuestas. El análisis de redes egocéntricas es adecuado cuando la investigación es enfocada en medidas, tipos y las consecuencias de la integración de los actores en su entorno social (Jansen D., 2007).
  - La red egocéntrica de un actor un Actor  $u$  lo conforma el conjunto de Actores  $V_u = \{u \cup \Gamma(u)\}$  y el conjunto de todas sus conexiones  $E_u = \{(u, v): u, v \in V_u\}$ . Esta red incluye además de las conexiones del usuario y su vecindad también las conexiones de Actores disjuntos de la vecindad (Dürr, 2013).

Desde el punto de vista del nivel de análisis (Wassermann & Faust, 1999) se categoriza estos en tres niveles: Micro, Meso y Macro.

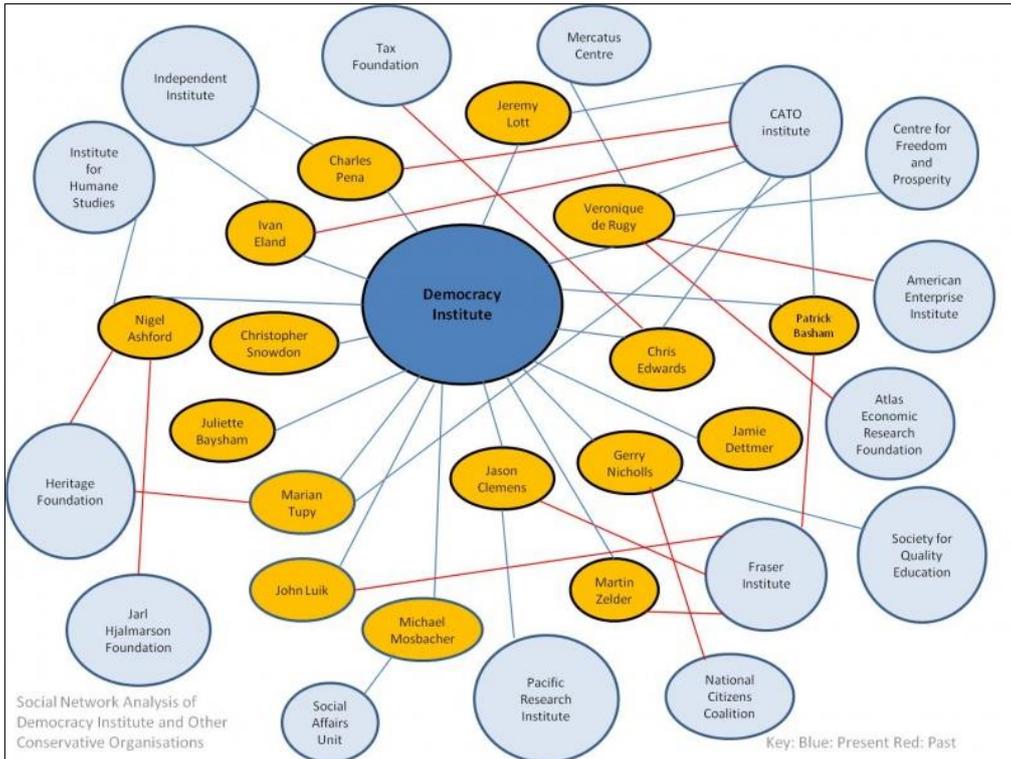
En el nivel micro, la investigación de redes sociales suele comenzar con un individuo, trazando sus relaciones sociales o se puede comenzar con un pequeño grupo de individuos en un contexto social particular.

El nivel macro comienza generalmente con un pequeño tamaño de población que cae entre los niveles micro y macro. Sin embargo, meso-nivel también puede hacer referencia a los análisis que están específicamente diseñados para revelar conexiones entre micro y macro niveles. Estas redes son de baja densidad y pueden presentar procesos causales distintos de las redes a nivel micro interpersonales.

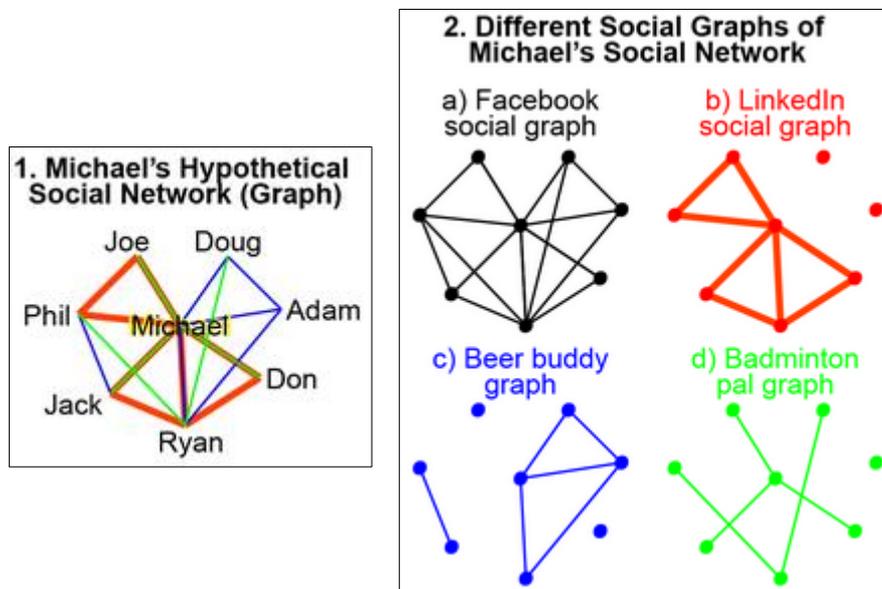
El nivel macro clasificado por Wassermann, en lugar de trazar las interacciones interpersonales, traza los resultados de las interacciones, tales como interacciones de transferencia de recursos económicos o de otro tipo más de una población grande.

En este estudio se mezclan diferentes enfoques: Se toman tres redes desde el punto de vista egocéntrico, conformando entre si una red sociocéntrica, para ser analizada a un nivel macro ya que lo que pasa en la red en si es de poco interés.





**Gráfica 13 Conexiones del Democracy Institute con los movimientos conservadores en powerbase.info**



**Gráfica 14 Análisis de una red social por Michael Wu en lithosphere.lithium.com**

## **2.3. El sistema de correo electrónico:**

### **Una comunidad virtual**

Los sistemas de correo electrónico, como vínculo entre personas, son un campo de investigación importante (Alstyne & Zhang, 2003), (Baym, Zhang, & Lin, 2004), (Gross & Acquisti, 2005), (Bird, Gourley, Devanbu, Swaminathan, & Gertz, 2006), (Kim, Kim, Park, & Rice, 2007), (Gruzd, 2012). Una primera línea de investigación, ha ido dirigida a la extracción y procesamiento de la información del sistema de correo electrónico. Así, los primeros trabajos en este campo fueron realizados con el sistema "Ahoy!" (Shakes, Langheinrich, & Etzioni, 1997) que desarrollaron una metodología de recuperación de la información (DRS<sup>5</sup>), con el objetivo de extraer las direcciones de correo electrónico de páginas de Web. Otros estudios han estado en su mayoría enfocados a la extracción de información de las páginas de internet (Xi, Fox, Shu, & Tan, 2002). Borkar et al. (2000) obtuvo una alta precisión en la extracción de datos buscando y usando un número limitado de información (número de casa, calle o carretera, ciudad, estado y código postal). Upstill et al. (2003), por su parte, describen una técnica de recuperación de información que aumenta la extracción de contenidos basándose en las características de las URL. McCallum

---

<sup>5</sup> Data Recovery System

et al. (2007) chequean los flujos de email con MD5 digest<sup>6</sup>. Shetty y Adibi (2004), y Klimt y Yang (2004), desarrollan una metodología con el objetivo de explorar la distribución de mensajes entre usuarios y tiempo. En general, están proliferando los estudios que tienen como objetivo captar los datos de correo electrónico, sin embargo, se observa una baja eficiencia en la recuperación de datos fundamentalmente, como consecuencia de carecer de algoritmos adecuados para ello (Bird, Gourley, Devanbu, Swaminathan, & Gertz, 2006), (Alstyne & Zhang, 2003) y de la necesidad de proteger la privacidad de los datos (Gross & Acquisti, 2005).

Una segunda línea de investigación se ha ido dirigiendo al estudio del sistema de correo electrónico desde el punto de vista estructural (Núñez & Cárdenas, 2013), (Castells, 2001). Desde esta perspectiva se considera que las interacciones creadas por el sistema de correo electrónico entre usuarios conforman una red social (Gruzd & Haythornthwaite, 2013). En este sentido, Contractor y Monge (2003), introducen el concepto de red social de comunicación, la cual refleja el flujo de mensajes a través del espacio y tiempo. La noción de red social ha implicado un primer aspecto que es el sentido de comunidad entre los miembros que la conforman. Freeman et al. (1989) consideran que la comunidad implica además de unas

---

<sup>6</sup> algoritmo de reducción criptográfico de 128 bits

relaciones específicas entre personas, un cierto grado de similaridad entre ellas. En este sentido, internet está creando unas comunidades virtuales, en las cuales la vía de relación entre otros son los emails, transformando la interacción física en interacción virtual. Hasta 1970, la comunidad se entendía como un grupo de personas que interaccionaban físicamente (Knoke & Yang, 2008), pero posteriormente se fue introduciendo la clave de distancia, por lo que en la comunidad no era necesaria la interacción física (Wellman & Leighton, 1979). Con la incorporación de las tecnologías de la información, se llega al concepto de comunidad virtual (Guinalú & Flavián, 2003). En este sentido, tal como señala Gruzd (2012), es importante en estas comunidades virtuales, la creación de un espacio común, en el cual la confianza, seguridad y la colaboración estén presentes, y donde las interacciones desarrollen grupos de identidad. Jones (1977) y McMillan y Chavis (1986), por su parte, argumentan que para la existencia de una comunidad virtual es necesario que se cumplan cuatro condiciones: interactividad; más de dos agentes; un espacio público de comunicación y, por último un sentimiento de pertenencia. Granovetter (1973) considera que la interactividad supone el intercambio, el diálogo y la dependencia entre los agentes implicados en esta comunicación. Cada agente presta atención al flujo de información, poniendo énfasis en el flujo bidireccional, así como en las conversaciones surgidas entre agentes, creando sensaciones de

comunidad (Burbules & Smith, 2005). Tales interacciones construyen una red de información compartida, de aprendizaje y debate, que refleja como las personas interactúan entre ellas y se implican en esa red social (Gruzd, 2012).

Por tanto, asumiendo que el sistema de correo electrónico, crea una comunidad online, la combinación del análisis de redes sociales con las técnicas de extracción de correo electrónico, proporciona una perspectiva analítica que provee información y visualiza como son las interacciones entre agentes. Desde esta perspectiva, se considera que el sistema de correo electrónico, conforma una estructura de interacción entre diversos agentes o actores, lo cual nos permite extraer información sobre aspectos organizativos (Klimt & Yang, 2004), (Bekkerman, McCallum, & Huang, 2004), (Berry & Browne, 2005), (Priebe, Conroy, & Marchette, 2005), (Keila & Skillicorn, 2005). McCallum et al (McCallum, Wang, & Corrada-Emmanuel, 2007), combinan la red social y email buscando similitudes entre personas en una organización. Smith (2009), analiza la red de correo electrónico de una comunidad, con el objetivo de obtener los niveles de autoridad en la misma, en función de la posición de los agentes en la red de comunicación. Frank (2005), analizan el caso del departamento de Computer Science, con el objetivo de estudiar las interacciones entre los miembros del departamento. Trier (2008), en un estudio longitudinal demuestra

como la combinación del análisis de redes y emails, permite no solo el estudio del proceso de comunicación en una organización, sino también como esta estructura reacciona ante eventos externos. Diesner et al. (2005) analizan el caso de Enron, estudiando los comportamientos del sistema de correo electrónico y estableciendo un paralelismo con la evolución de la situación económica. Estos Autores encuentran, que, en situaciones de crisis, la red de correos electrónico es más diversificada, siendo más homogénea cuando nos alejamos de la crisis.

## **CAPÍTULO III. MARCO TEÓRICO – DESARROLLO ECONÓMICO DE EMPRESAS PYME**

### **3.1. Introducción**

Una muy buena descripción del desarrollo de la economía y por consiguiente de las empresas PYME en la historia antigua, hace Mario Caius Amschlinger en su estudio empírico (Amschlinger, 2011) sobre el desarrollo de inversiones directas de estas empresas en Rumania. Amschlinger describe como a través de la historia, comenzando con la producción por necesidad propia, expandiendo esto a la función del trueque y posteriormente con la enseñanza de la misma postulada por Aristóteles, las empresas PYME han sido el fundamento para el desarrollo de cualquier civilización.

En el área geográfica europea, se han encontrado restos de lo que se puede deducir son pequeñas y medianas empresas, en especial las excavaciones en Pompeya muestran algunos indicios de ello (Mugler, 2008). Sin embargo existen teorías sobre el desarrollo de la economía a través de la historia (García Ruiz, 1995) que tratan la materia desde el punto de vista del gigantismo. Es decir, las PYME no juegan un papel importante en el desarrollo económico y se refieren más a trabajos de gestión como lo fueron la construcción de las pirámides en Egipto o de la Gran Muralla china. De por si García Ruiz

dedica una gran parte de su estudio en mostrar como una empresa como Siemens después de la II Guerra Mundial pudo llegar a ser tan grande y poderosa.

### 3.1.1. Definición de empresas PYME

La empresa tomada como ejemplo en esta investigación es una empresa típica PYME europea. Según la Unión Europea en la Recomendación del 6 de mayo de 2003, basándose en la Carta de la Pequeña Empresa emitida en el Consejo Europeo de Santa María da Feira en junio de 2000, con entrada en vigor el 1 de enero de 2005, una PYME: *“es la unidad económica con personalidad jurídica o física que reúna los siguientes requisitos”* (Comisión Europea de la UE, 2003):

Tipo de empresa	Empleados (AWU) <sup>7</sup>		Facturación (Millones de €)		Total de Balance general (Millones de €)
Microempresa	< 10	y	≤ 2	o	≤ 2
Pequeña empresa	< 50	y	≤ 10	o	≤ 10
Media empresa	< 250	y	≤ 50	o	≤ 43

**Tabla 5 Clasificación de PYMEs según la UE**

Además, para ser reconocida una empresa como pequeña y mediana empresa de la UE, también es necesario que hasta un 25% de la compañía esté en propiedad de las empresas que no cumplen con

---

<sup>7</sup> AWU: Unidad de trabajo promedio anual

esta definición. De esta forma, se toma como referencia en la definición de una empresa PYME los valores cuantitativos dados por la comunidad económica europea.

El REGLAMENTO (UE) No 651/2014 DE LA COMISIÓN de 17 de junio de 2014 por el que se declaran determinadas categorías de ayudas compatibles con el mercado interior en aplicación de los artículos 107 y 108 del Tratado, no altera básicamente lo asumido anteriormente. Sino que amplía las posibilidades de ayudas a las PYMES:

- de finalidad regional
- a la inversión, de funcionamiento y para el acceso a la financiación
- para la protección del medio ambiente
- de investigación y desarrollo e innovación
- a la formación
- a la contratación y empleo de trabajadores desfavorecidos y trabajadores con discapacidad
- destinadas a reparar los perjuicios causados por determinados desastres naturales
- de carácter social para el transporte en favor de residentes en regiones alejadas
- para infraestructuras de banda ancha

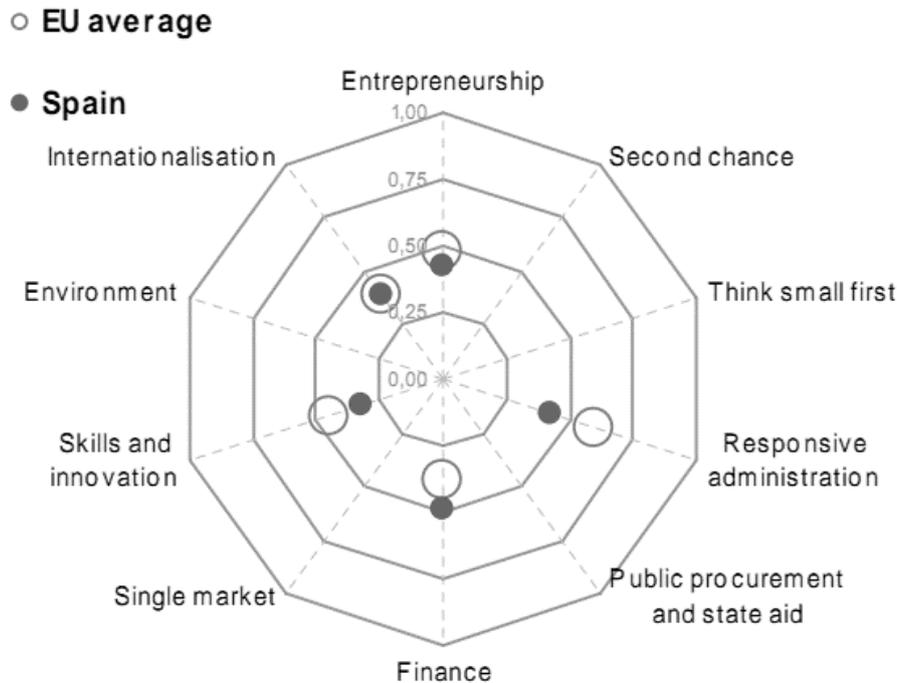
- a la cultura y la conservación del patrimonio
- a infraestructuras deportivas y recreativas multifuncionales y
- a infraestructuras locales

### **3.1.2. Desarrollo económico en la CEE**

La importancia de las empresas PYME en Europa se observa en el sentido que estas representan aproximadamente el 99% (23 millones) (Comisión Europea de la UE, 2005) del total de empresas y proporcionan trabajo al alrededor de 75 millones de personas. Las PYME son uno de los núcleos fundamentales para la innovación. En Alemania las empresas pequeñas y medianas dan aproximadamente el 99,7% de todo el IVA de las empresas registradas, además constituyen aproximadamente el 65,8% de todas las contribuciones del seguro social, donde además se genera el 37,5% de los ingresos totales y donde el 83,0% de todos los aprendices reciben formación (Günterberg & Pahnke, 2012). Sin embargo, estos valores de apariencia exitosa tienen grandes críticos.

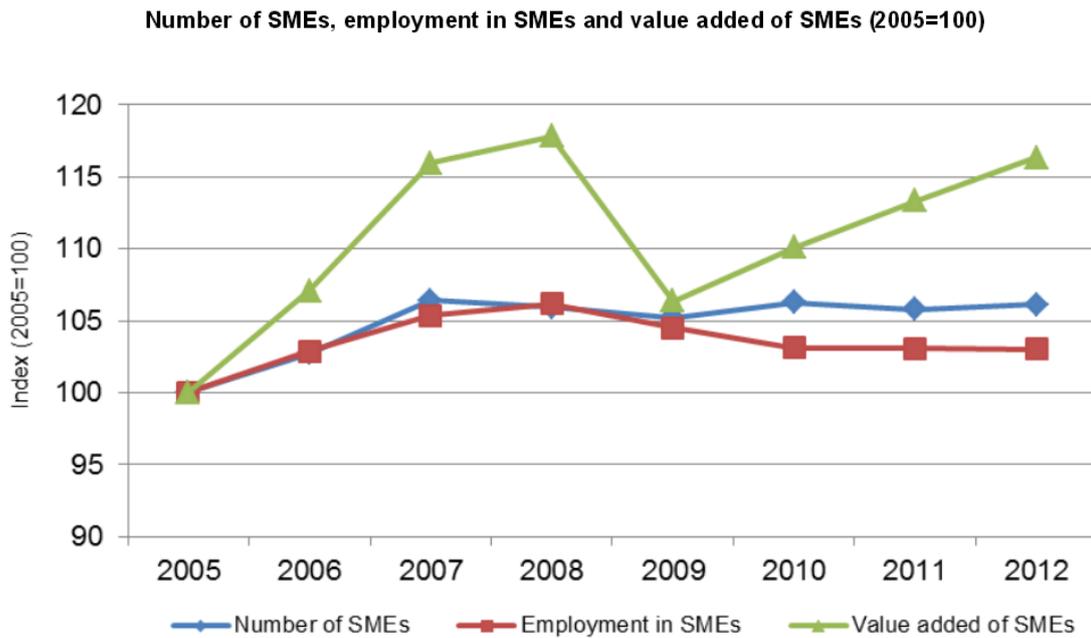
Es así que debido a la restringida capacidad económica individual de las empresas PYME y de por ello su posible inflexibilidad para adaptarse rápidamente a los cambios del mercado, por lo que son las primeras en sufrir grandes crisis.

Tomemos por ejemplo a España que históricamente ha sido un país con muchas PYMEs:



**Gráfica 15 El perfil de la PYME española frente a la de la UE-27 (Fuente: Comisión Europea, [www.oecd.org](http://www.oecd.org) (2006))**

El 99,9% de las empresas españolas son según la OECD de carácter PYME, dando empleo al 78,7% de total de empleados. El estudio más reciente de la OECD muestra que después de la crisis del 2009, la importancia de las PYME ha ido creciendo poco a poco (Fuente: Eurostat/National Statistics Offices of Member States/Cambridge Econometrics/Ecorys):

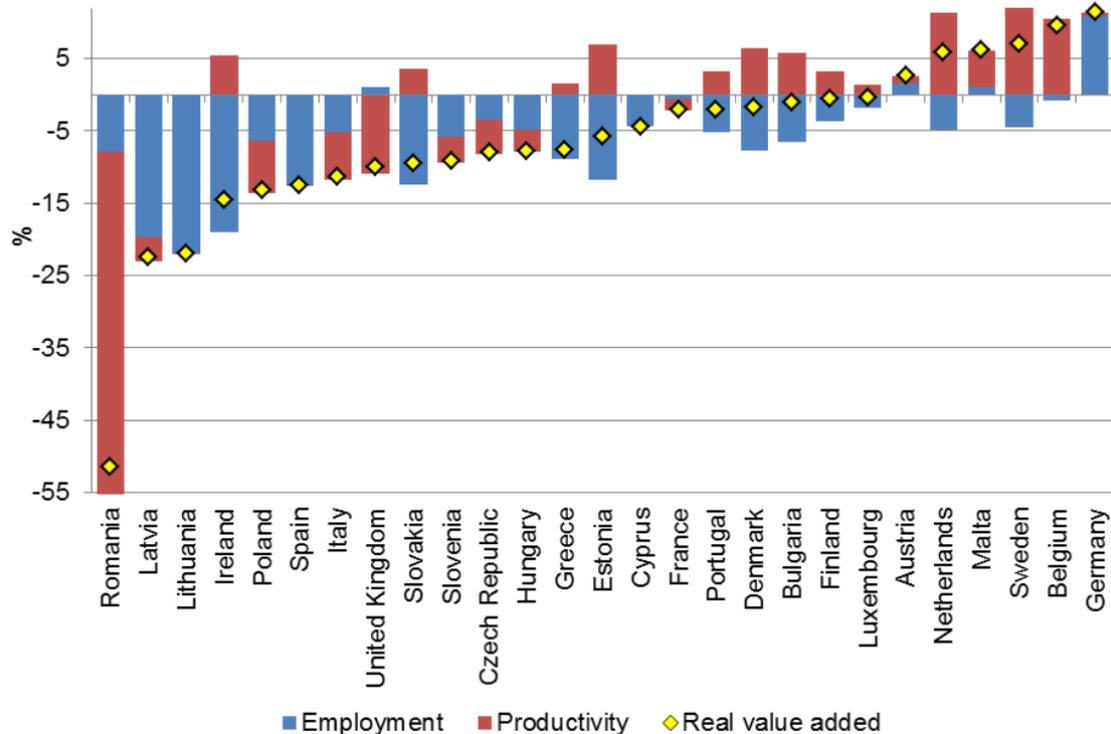


**Gráfica 16 Número de PYMES, empleados y valor agregado de PYMES**

Un análisis a nivel por países muestra que ese desarrollo no es homogéneo. Es así como según la OECD desde el 2011 las PYME en Austria y Alemania han superado sus niveles de valor agregados y de empleo que tenían en el 2008. De otro lado las PYME en Bélgica, Finlandia, Francia y Luxemburgo, ha llegado a los valores de aquel año y los otros 20 estados miembros, entre ellos España, las PYME han sido hasta ahora incapaces de recuperarse a los niveles previos a la crisis (Wymenga, Spanikova, Barker, Konings, & Canton, 2012).

En efecto un problema radica aparentemente en la diversificación de las PYMES, que da a países como Alemania, Inglaterra, Francia e Italia una gran ventaja en la concentración de áreas específicas. En Alemania hay una gran concentración de PYMES que se han especializado en tecnología avanzada y en Inglaterra,

Francia e Italia hay una gran concentración de PYMES especializadas en servicios de conocimiento, asesoría y consulta.



**Gráfica 17 Crecimiento anual de empleos, productividad y valor agregado por países del 2008 al 2011 (Fuente: (Wymenga, Spanikova, Barker, Konings, & Canton, 2012))**

Finalmente tenemos el SBA<sup>8</sup> que fue puesto en vigencia en su forma actual en 2008. Esta ley muestra la importancia que la CE da al desarrollo de las PYMEs y establece un campo de acción claro para su fomento y observación. La SBA establece también los argumentos necesarios para poder medir una política en este sector y señala 10 principios básicos para ello:

<sup>8</sup> SBA: Small Business Act for Europe

- El espíritu empresarial
- Una segunda oportunidad
- Pensar primero a pequeña escala
- Administración responsable
- Ayudas estatales y la contratación pública
- Acceso a la financiación
- Mercado Único
- Habilidades e innovación
- Medio ambiente
- Internacionalización

Basados en estos principios se elaboran estadísticas y de ellas posteriormente ajustes a las políticas de apoyo de la industria PYMES (Saisana, 2012).

Es así que estos análisis y estudios ponen un mayor énfasis en la búsqueda de información en los marcos de desarrollo de las empresas PYMES para establecer y predecir su desarrollo, si bien no desde el punto de vista clásico económico, si desde el punto de vista socioeconómico, que es la meta de esta investigación.

### **3.1.3. Demografía empresarial -**

#### **Tendencias en Europa**

En un estudio reciente del Deutsche Bank (Vetter & Köhler, 2014), se define que el número de empresas activas no dice mucho sobre el estado del sector empresarial. En particular, cualquier estructura económica con muchas pequeñas empresas también puede indicar que un proceso de consolidación aún no se ha alcanzado y por consiguiente es difícil de alcanzar. Si un alto nivel de competencia garantiza que las empresas no rentables salgan del mercado o a través de fusiones estas puedan mejorar su economía, una reducción en el número de empresas puede conducir a ganancias de eficiencia global económicas.

En las últimas crisis económicas, dice Vetter, se ha comprobado que las grandes empresas y las micro-empresas han sobrepasado los malos tiempos en mejor forma, siendo a las PYMEs las que más han sufrido en esos tiempos. Según Eurostat<sup>9</sup> emplea una micro empresa en la UE sólo dos personas en promedio. Si bien, estas en los tiempos de abundancia económica no crecen o crecen muy lentamente, en la misma forma en tiempos de crisis no decrecen

---

<sup>9</sup> Oficina de estadística de la Unión Europea

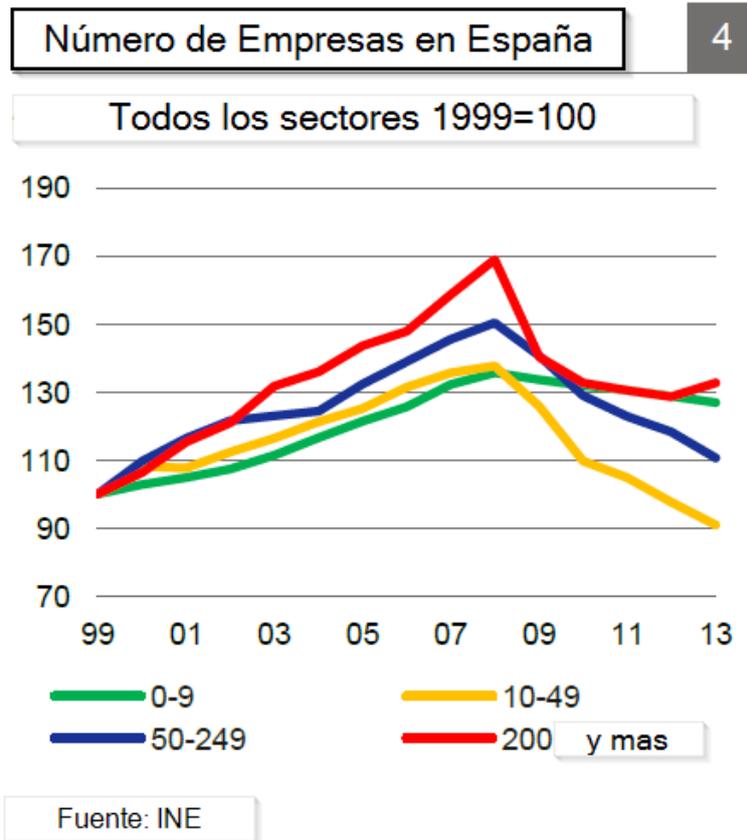
rápidamente. Más del 30% de estas empresas no tiene otros empleados y alrededor del 70% tienen menos de cinco.

El estudio hace una mención especial a la situación española, Desde mediados de los años 90 España pasó por un desarrollo altamente expansivo, que fue impulsado principalmente por el sector de la construcción. En comparación con el nivel de 1999, el número total de empresas, según la oficina de estadística europea, se incrementó en aproximadamente un 35% de su máximo de 2008. Se observó una expansión particularmente en las grandes empresas con 200 o más empleados, el número había aumentado a 2008 en casi un 70%. Desde entonces, sin embargo, se observa una disminución substancial de aproximadamente el 10% en casi todas las categorías de empresas. El papel especial del sector de la construcción para el desarrollo económico general en España se observa también en la demografía de las empresas de una manera impresionante.

En general, oficina de estadística europea observa un desarrollo muy diferente entre las PYMEs y las grandes empresas en España. Al final de la recesión se espera un cambio de tendencia en el sector empresarial. El cambio observado recientemente de que las pequeñas y medianas empresas desaparecen y su función es en parte asumida por las grandes empresas, debido a su actual situación financiera, continuará al menos a corto plazo.

En forma contraria se observa en Alemania, que las pequeñas, medianas y grandes empresas se han incrementado de manera significativa numéricamente, a pesar de la crisis en la zona euro, lo que explica la evolución positiva en el mercado laboral. Con mayor claridad dice Vetter, se observan las diferencias en el sector manufacturero, donde las microempresas disminuyeron, mientras que las empresas industriales, especialmente las medianas y grandes muestran un desarrollo positivo. Probablemente se debe esto principalmente al hecho de que las empresas con el aumento del tamaño han tomado una orientación más fuerte hacia la exportación, beneficiándose de un mayor crecimiento fuera de la zona europea.

Las pequeñas y medianas empresas han sido los grandes perdedores en los países euro-crisis en los últimos años. En España, Italia e Irlanda, su número se redujo significativamente, más fuerte que la de las grandes empresas o microempresas con menos de 10 empleados. En los países donde los choques económicos sufridos solamente han sido a corto plazo (por ejemplo, Alemania o el Reino Unido), no se observaron tales cambios.



Gráfica 18 Número de empresas en España por año y clase

### 3.1.4. Análisis en función del tiempo

Existen varios métodos para analizar series en función del tiempo. En Schlittgen (2012) se describen los diferentes métodos para este tipo de análisis, que según el objetivo buscado, utilizan diferentes estrategias. En este estudio tres definiciones son importantes:

- Series de tiempo: Es una sucesión  $y_1, \dots, y_N$  de realizaciones de una sección de un proceso estocástico.

- Proceso estocástico: Es una sucesión de variables casuales ( $Y_t$ ). El índice  $t, t \in N, N_0$  se considerada en relación al tiempo.
- Proceso normal también llamado proceso Gauss. Es un proceso estocástico en el cual las variables están distribuidas normalmente.

Un factor típico en el desarrollo de series de tiempo, crucial en el análisis del desarrollo de empresas PYME, es la falta de una homogeneidad a través del tiempo. En algunos casos, y en especial de la compañía objeto de evaluación, es necesario tener en cuenta la temporalidad y – llegado el caso – hacer una corrección de la misma. Igualmente tenemos que tener en cuenta la periodicidad de las observaciones del desarrollo económico, que por la característica de los datos obtenidos tiene que ser obvia. Es decir, si se hace una corrección de temporalidad en las observaciones del desarrollo de la red social, esta misma corrección tendrá que ser tomada en cuenta en las observaciones del desarrollo económico de la misma.

Para encontrar si hay o no una correlación entre el desarrollo económico de la empresa y la evolución o comportamiento de la red social, se estudiará la periodicidad de cada una de estas series de tiempo.

### **3.2. Conclusiones**

En la literatura actual no se encuentra un análisis específico sobre las influencias de las redes sociales en el desarrollo económico de empresas PYME. En el marco económico de desarrollo de la Comunidad Económica Europea sin embargo este tipo de empresas juegan un papel muy importante. Si las redes sociales han sido de vital importancia en el transcurso del tiempo para que gobiernos, regímenes, instituciones y grandes empresas hayan llegado crecido y se hayan desarrollado, se establece la clara pregunta si ese efecto, probablemente en una escala menor, también se pueda observar en las empresas PYMEs.

La teoría del análisis económico en especial la regresión lineal múltiple se presta para buscar una relación de tipo matemático entre el desarrollo de la red social y el desarrollo económico de tales empresas. Debido a la variedad tanto de tamaño como de forma de trabajo de las empresas de este contexto, es necesario tratar de mantener la investigación en un nivel general para darle practicabilidad a los resultados encontrados. Mas sin embargo no se podrá omitir sentencias y evaluaciones específicas de la empresa ejemplo y de uso en esta investigación.

## CAPÍTULO IV. METODOLOGÍA

### 4.1. Introducción

Se trata de un análisis con multivariantes, en cual se definen como variables dependientes todos aquellos valores obtenidos del sistema mediante extracción de datos del sistema contabilidad y control de gestión de la empresa (SAP R/3 ERP 6). En este caso son tres variables dependientes por mes y año: Costos (C), Ingresos (I) y Balance general (B) como resultado de la operación Ingresos - Costos. Estos datos son normalizados (5.3) para efectos de comparación como más adelante se describe en el caso de los datos obtenidos del análisis de la red social.

Una cuarta variable dependiente que se puede introducir es el valor concerniente a la medida específica del desarrollo de la empresa en el contexto de la industria donde se mueve. En este caso es el número de pasajeros por mes y año. En este estudio no se usará.

Las variables independientes son aquellas obtenidas del análisis de la red social y son descritas más adelante (5.2).

La ecuación fundamental del análisis es la regresión lineal múltiple:

$$Y = \alpha + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon \quad (9)$$

donde  $Y$  es una o el conjunto de variables dependientes y  $X$  las variables independientes.  $\alpha, \beta_x$  son los coeficientes de regresión y  $\varepsilon$  un coeficiente de error.

## **4.2. Herramientas de software**

### **4.2.1. El lenguaje R**

El lenguaje R según W. N. Venables (2013) es derivado de los trabajos que hicieron los autores en 1990 entorno a los sistemas S y S-Plus durante su estadía en la Universidad de Adelaide. Debido a que R es un software libre y abierto, se presta especialmente para que la comunidad lo pueda usar y extender. Además es muy útil especialmente porque existen actualmente más de 4800<sup>10</sup> paquetes con funciones ya programadas listas para ser usadas en cualquier proyecto. En especial se relatan los siguientes paquetes que son usados en esta investigación:

- Igraph (Csardi & Nepusz, 2006): herramienta para analizar redes sociales
- Multiplex (Ostoic, 2013): para analizar redes sociales múltiples con algebra

---

<sup>10</sup> <http://www.cran.r-project.org/>

- Sna (Butts, 2013): herramientas para análisis de redes sociales
- Snort (Dubossarsky & Norrie, 2012): análisis de redes sociales en tablas relacionales
- Erer (Sun, 2013): investigación empírica en economías
- MicEcon (Henningsen, 2012): modelaje y análisis de microeconomías
- AER (Kleiber & Zeileis, 2008): econometría aplicada

En el “Apéndice – Programas en R” están los programas escritos por mí para obtener la información necesaria para hacer la investigación.

En algunos documentos se menciona este lenguaje como un sistema estadístico, sin embargo, los autores lo definen como un entorno en el cual se han implementado en forma fácil técnicas de estadísticas.

Además de la herramienta Lenguaje R, se utiliza en esta investigación en forma intensiva y en diferentes tareas el software Excel de Microsoft en la versión 2013.

#### **4.2.2. Microsoft Excel**

No hay prácticamente una herramienta parecida y con tanta capacidad para hacer análisis de todo tipo que la tabla de cálculo de

Microsoft Excel. El origen de esta herramienta de software se remonta al año 1985. Por su facilidad de uso y cantidad de posibilidades con la ayuda de VBA<sup>11</sup> para automatizar o crear nuevas funciones, Excel es utilizado por muchos profesionales. Su integración en software ERP<sup>12</sup> tales como SAP R/3, Navision o Oracle financieros hace que su uso en casi todas las compañías sea diaria. Además, tanto SPSS, el Lenguaje R y otras herramientas como UCINET reconocen la importancia de esta tabla de cálculo al ofrecer desde un comienzo una interface desde donde se pueden usar las funciones de Excel o desde él se pueden usar los datos y otras funciones de las mismas herramientas.

En el transcurso del desarrollo y evolución de este software, ha habido una gran cantidad de software similar, que desde el punto de vista de atributos muchas veces han superado a Excel tanto en capacidad como en estabilidad. Sin embargo, el mercado por razones que no son cuestión de este estudio ha rechazado parcialmente estas herramientas. Entre ellas están por ejemplo Lotus-1-2-3, VisiCalc, Quattro Pro o la misma Multiplan originaria igualmente de la casa Microsoft como lo es Excel. En el mercado actual existen varios paquetes para diferentes sistemas operativos de tablas de cálculo, que soportan todas sin excepción el formato de Excel. Por ejemplo Numbers

---

<sup>11</sup> Visual Basic for Applications

<sup>12</sup> sistemas de planificación de recursos empresariales

de Apple para su sistema operativo, StarOffice Calc, Calc, Gnumeric y Kspread para los sistemas Linux (Colaboradores de Wikipedia, 2013).

### **4.2.3. UCINET**

Está en la lista de software usados para análisis de redes sociales, que abarca en Wikipedia<sup>13</sup> más de 70 al día de hoy (12.09.2013)<sup>14</sup> y más de 80 en Huisman & van Duijn (2011), todos con diferentes especialidades y características. Sin embargo en Wikipedia se mezcla la definición entre el análisis de redes sociales desde el punto de vista clásico y el que entiende como red social solo aquella que es definida por su actividad en internet (como lo son Facebook, twitter o google+<sup>15</sup> como ejemplos de innumerables redes en internet de intercambio de información de todo tipo) y que son llamadas redes sociales por su carácter privado. Es así como una herramienta de software como lo es WolframAlpha está diseñada para dar respuestas con respecto a las actividades en dichas redes (Nelson, 2013).

UCINET (Borgatti, Everett, & Freeman, 2012) así como otras herramientas de software es una herramienta especializada para el

---

<sup>13</sup> Sitio de Web que se maneja como una enciclopedia universal, cuya administración y control no está supervisada por una entidad central o gubernamental

<sup>14</sup> [http://en.wikipedia.org/wiki/Social\\_network\\_analysis\\_software](http://en.wikipedia.org/wiki/Social_network_analysis_software)

<sup>15</sup> Sitios de Web donde los usuarios pueden intercambiar información de diferente tipo tales como ideas, fotos, gráficos, documentos etc.

análisis de redes sociales en sentido tradicional. La definición más concreta desde mi punto de vista que describe a UCINET es esta: "Un paquete completo para el análisis de datos de redes sociales, así como de otras redes de tipo 1-mode y 2-mode. Puede manejar un máximo de 32.767 nodos (con algunas excepciones), aunque en términos prácticos muchos procedimientos se prolongan demasiado cuando la cantidad de nodos está entre 5.000 y 10.000. Los métodos de análisis de redes sociales incluyen medidas de centralidad, identificación de grupos, análisis de papeles, teoría de grafos básica y el análisis estadístico basado en permutaciones. Además, el paquete tiene rutinas de análisis de matriz fuertes, tales como álgebra matricial y estadística multivariante". Esta definición se encuentra en innumerables fuentes en internet sin poder establecer cuál es la original, ya que esta no se encuentra en el sitio de Web del autor.

UCINET es según sus autores una colección de herramientas de análisis que comenzó a desarrollarse a principios de los años 80. La versión actual está adaptada para correr en sistemas de 64 bits, dando así un gran paso en el rendimiento durante el cálculo de operaciones complejas.

#### **4.2.4. MemBrain**

MemBrain es un editor gráfico de redes neuronales, que soporta redes neuronales artificiales de tamaño y arquitectura

arbitraria (Jetter, 2014). Con este paquete de software es posible diseñar redes neuronales de diferentes tipos, que ayudan a la búsqueda de posibles redes que se puedan adaptar fácilmente a diferentes entornos. El método gráfico de diseño además ayuda a entender el funcionamiento básico. El producto tiene además la posibilidad de producir código en el lenguaje C que puede ser usado por otros programas para el análisis complejo de situaciones. Esta posibilidad no es usada, ya que no se utilizará C en ningún momento del trabajo de investigación.

En la descripción del programa, el autor define el producto como “un software que originalmente había sido desarrollado para ofrecer la máxima flexibilidad para el desarrollo y el estudio de las redes neuronales artificiales combinados con una interfaz gráfica de usuario intuitiva.”

Una de las grandes ventajas de este software es la capacidad sencilla de importar datos para el entrenamiento de una red y posteriormente con otros datos verificar su efectividad.

### 4.3. Construcción de la red social

#### 4.3.1. Proceso de adquisición de datos del sistema de correo electrónico

Durante la búsqueda de métodos de adquisición de datos provenientes del correo electrónico, me encontré con gran cantidad de dificultades. Si bien los estándares (Crocker, 1982. Postel, 1982. Klensin, Freed, Rose, Stefferud, & Crocker, 1994. Klensin (Ed.), 2001. Resnic (Ed.), 2001) están bien definidos, cada sistema utiliza una forma diferente para exportar los datos recibidos.

Siguiendo la metodología de Cain (2012) y Basso (2011) el sistema de correo usado en esta investigación ha sido Microsoft Exchange 2012 y Microsoft Outlook 2013, que es el más popular sistema de administración de correo electrónico<sup>16</sup>. Este sistema permite la posibilidad de exportación de correos en el formato de la Tabla 6. Sin embargo, observamos que falta un dato importante para el análisis: la fecha de envío del correo electrónico.

De	(Nombre)
De	(Dirección)
De	(Tipo)
A	(Nombre)
A	(Dirección)

<sup>16</sup> <http://www.peterdehaas.net/ibm/> Computer Profile, Mayo 2014

A	(Tipo)
CC	(Nombre)
CC	(Dirección)
CC	(Tipo)
BCC	(Nombre)
BCC	(Dirección)
BCC	(Tipo)

**Tabla 6 Formato de exportación desde Microsoft Outlook**

Nos damos cuenta que falta un valor importante para el análisis: la fecha de recibimiento o envío del correo electrónico.

Otra posibilidad que ofrece Microsoft Outlook para exportar información es utilizar la función de cortar, copiar, pegar de una tabla preparada con las columnas requeridas.

Esta función, que en el primer momento parece práctica, y que puede ser para algunos casos suficiente, sin embargo muestra una debilidad contundente para el análisis exacto de los datos: no se pueden extraer datos de la dirección de correo electrónico en el formato <nombre>@<dominio>.<región>. Solo es posible ver el alias de la dirección y de esta forma no es posible hacer una correlación exacta entre el que manda un correo y el que la recibe. Por ejemplo, si en mis contactos tengo guardado un contacto con el nombre "Carlos Aguilera (UNED)" con dirección c.aguilera@dell.com pero esa persona cuando manda su correo llega con el nombre de "Aguilera, Carlos" <c.aguilera@dell.com>, en MS-Outlook no habrá una correspondencia entre esos dos correos, ya que cuando se manda un correo se envía a

"Carlos Aguilera (UNED)" pero se recibe de "Aguilera, Carlos". Esa comodidad en MS-Outlook es un problema para la investigación. Así, es necesario buscar un método diferente para recolectar los correos en una forma que pueda ser analizada. Así es que es necesario buscar un método diferente para recolectar los correos en una forma que pueda ser analizada.

El uso de otro cliente puede ser una alternativa: Por ejemplo, Outlook Express. La función de exportación es idéntica que MS-Outlook, pero tiene una forma de guardar correos que no posee MS-Outlook. Es posible guardarlos en el formato EML. Este formato es legible y se ve por ejemplo de la siguiente forma:

```
From: "Das Team von Microsoft Outlook Express" <msoe@microsoft.com>
To: "Neuer Outlook Express-Benutzer"
Subject: Willkommen
Date: Sat, 30 Apr 2011 00:03:41 +0200
MIME-Version: 1.0
Content-Type: text/html;
    charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable
X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2900.5994
```

Este formato nos da la información requerida para el análisis. Pero implicaría que cada uno de los correos tendría que ser guardado en forma individual. El uso de un script en Visual Basic en Outlook no es nada fácil y toma mucho tiempo. Buscando alternativas y habiendo probado otros clientes tales como Mozilla Thunderbird, Lotus Notes entre los tradicionales u otros como Eudora, KMail, SeaMonkey Mail o Pegasus Mail entre los exóticos, encontré un medio práctico que a la

vez no molesta al usuario en su vida cotidiana y ofrece una forma fácil y útil para la extracción regular y permanente de correos electrónicos. Se trata de acceder a un servicio de archivo tal como MailStore (<http://www.mailstore.com/>) donde es posible exportar todos los correos en formato eml. Otras alternativas, tales como sistemas basados en Perl<sup>17</sup> o en el mismo R<sup>17</sup> pueden ser programados y usados para extraer los datos igualmente.

Una vez la exportación de correos en formato eml es completado, necesitamos un programa que extraiga de todos esos correos la información necesaria para el análisis.

El formato que encontramos es del siguiente tipo:

```
From: "Nombre completo" <correo@dominio.región>
To: "Nombre completo" <correo@dominio.región>
CC: "Nombre completo" <correo@dominio.región>
BCC: "Nombre completo" <correo@dominio.región>
Subject: Vogelschlag
Thread-Topic: Vogelschlag
Thread-Index: AcS1EjqqCFzgHyZ1Q5Oz1JKf9qCg5w==
Date: Mon, 18 Oct 2004 12:58:57 +0000
Message-ID: <E076DCA8200C2A48B84FE2D35D5CC58F894317@ntsrv-
mail.fmo.de>
Content-Language: de-DE
X-MS-Has-Attach: yes
X-MS-TNEF-Correlator:
Content-Type: multipart/mixed;
    boundary="_004_E076DCA8200C2A48B84FE2D35D5CC58F894317ntsrm
ailfmode_"
MIME-Version: 1.0
X-MailStore-Folder-UTF7: archivo-donde-se-encuentra-el-correo-
original
X-MailStore-Message-ID:
<e076dca8200c2a48b84fe2d35d5cc58f894317@ntsrv-mail.fmo.de>
```

---

<sup>17</sup> Lenguajes de Programación comunes



Es por eso que en el programa de extracción utiliza una base de datos adyacente que iguala todos estos correos en uno solo. En mi caso por asunto de simplicidad son todos generalizados al correo corto de 3 letras.

La base de datos en donde se archivan los datos esenciales tiene la forma siguiente:

Nombre del campo	Tipo del campo
SERIALNR	<Valor automático>
EMAILID	Texto
EMAILDATE	Fecha/Hora
EMAILFROM	Texto
EMAILTO	Texto
EMAILCAT	Texto

**Gráfica 19 Descripción de los registros de la base de datos**

Se observa que no se tiene registro extra o especial para CC o BCC. Durante el proceso decisivo de diseño de la base de datos, así como el de la forma del análisis se probaron diferentes métodos. La primera aproximación (Rodríguez F. , 2010) encuentra la dificultad de diferenciar o dar un valor a la conexión entre los nodos si este es conectado en forma directa (TO:), en forma indirecta (CC:) o en forma ciega (BCC:). Esa primera aproximación da cuenta de que no hacer una diferenciación puede conducir a una mala interpretación del valor de la red. Es decir que una conexión directa tiene que tener un valor mayor que una indirecta por su causa. En una conversación directa de los actores se espera una actividad más estrecha entre los mismos,

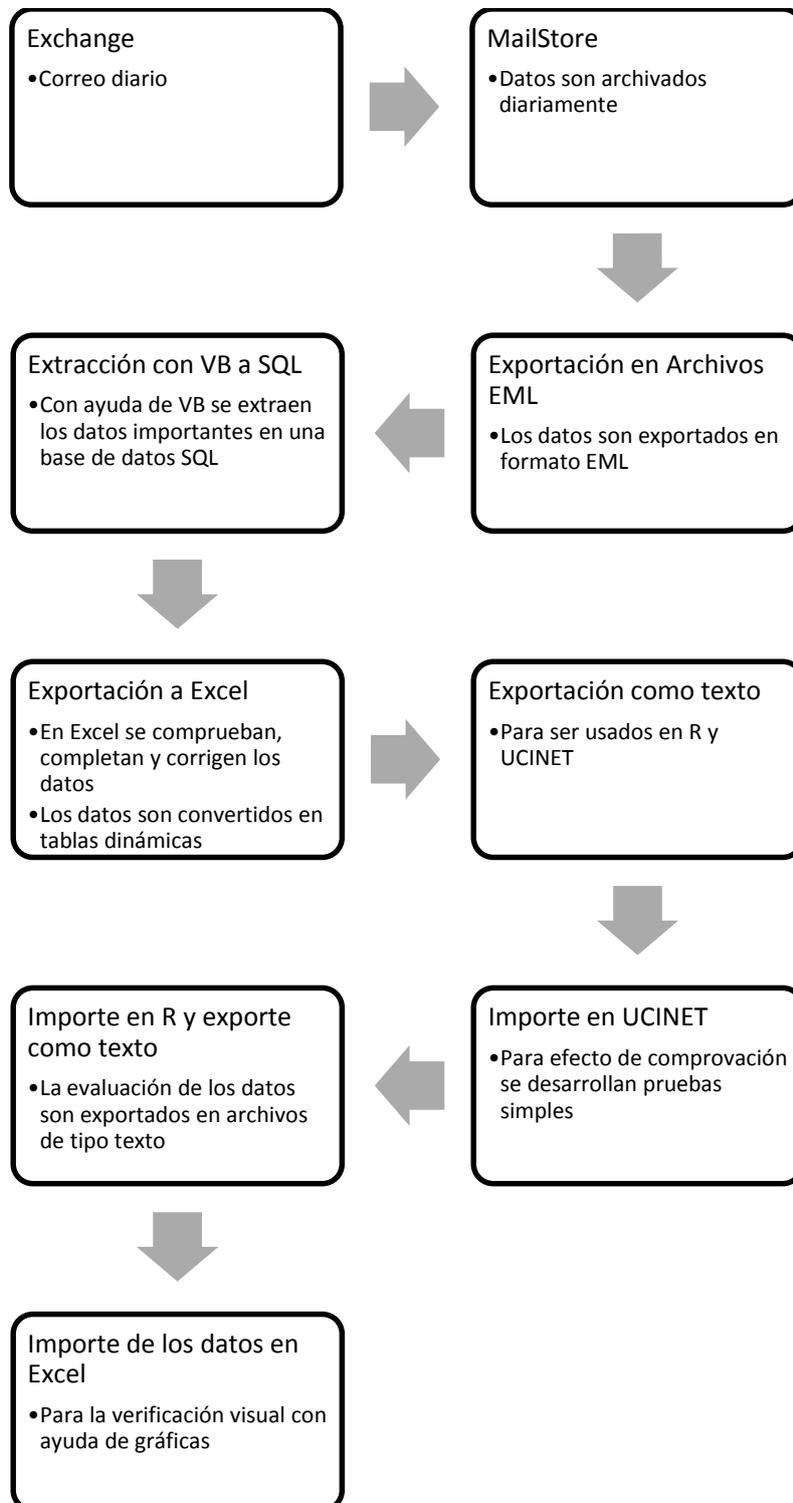
siendo aquellos que son pasivos en la interacción de un valor intrínseco menor. Es por eso que se introduce un nuevo registro en la base de datos que da valor a eso. En el registro EMAILCAT se guardan 3 valores diferentes: 1 para una comunicación directa (TO:) y 2 para una comunicación indirecta (CC: o BCC:). En el análisis no hago diferencia alguna entre BCC: y CC: en asunto de valor.

En el análisis posterior se introduce un nuevo registro llamado EMAILCLASS. Este registro contiene un valor diferenciado de la comunicación y puede tener los siguientes valores:

- PRIVAT (carácter privado o personal)
- COMPANY (compañeros de la misma empresa)
- JOURNALISM (periodistas y similares)
- POLITIC (políticos no en gremios dueños de la empresa)
- SHAREHOLDER (políticos en gremios dueños de la empresa)
- REGIONAL (representantes de industria regional)
- SAMEINDUSTRY (representantes de industria aeronáutica)
- OTHERS (identificados, pero no clasificables)
- UNKNOW (no identificados)

En función del análisis del dominio o región y conocimiento interno se adjudica esta clasificación. Esta acción tiene que hacerse en forma manual, ya que todo automatismo puede conducir a clasificaciones erróneas.

Proceso de adquisición de los datos:



**Gráfica 20 Proceso de adquisición de datos**

Estos son los datos que contiene la base de datos, una vez ha sido llenada con la información disponible:

Característica	Valores
Cantidad de registros	365.602
Cantidad de emisores únicos	8.856
Cantidad de destinatarios únicos	16.058
Cantidad de dominios únicos emisores	4.992
Cantidad de dominios únicos destinatarios	6.272
Cantidad de correos cuyo emisor y destinatarios son internos	180.326

**Tabla 7 Características del contenido de la base de datos**

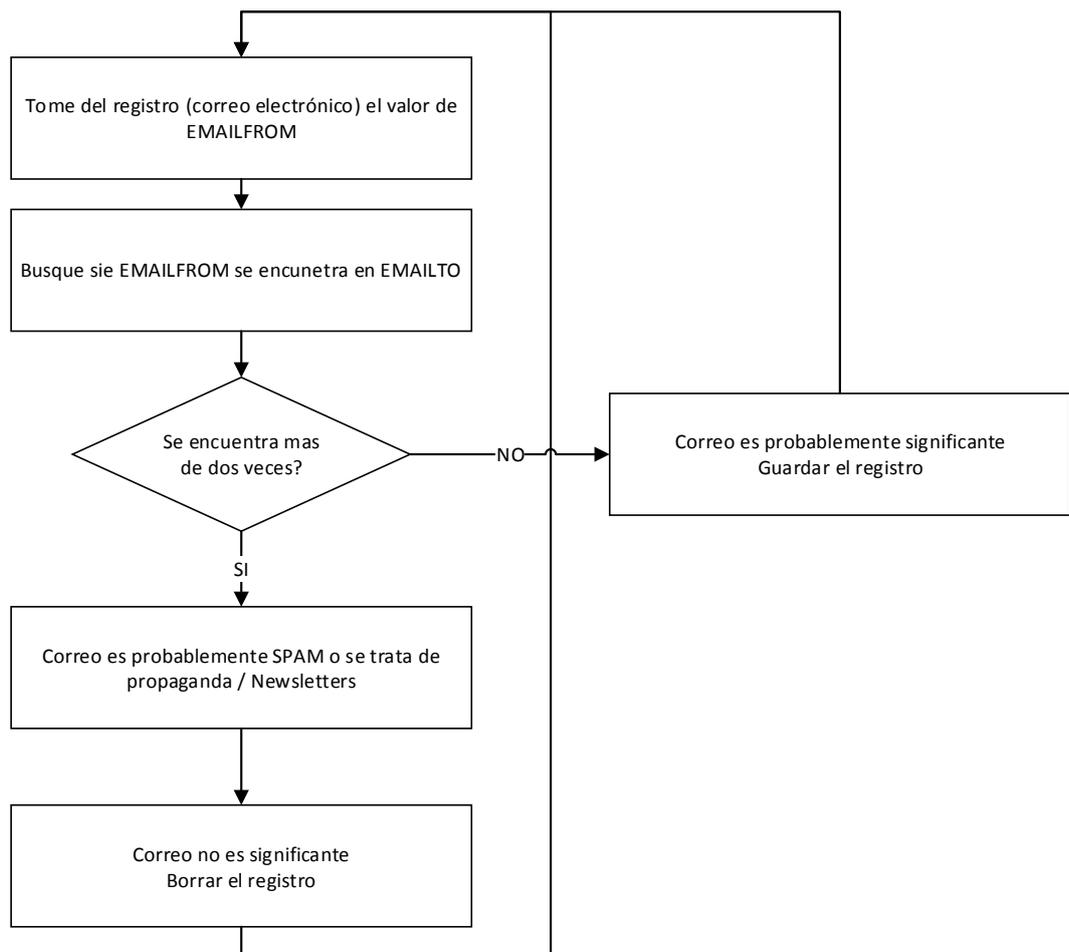
#### **4.4. Propuesta de un método automático**

La idea fundamental es describir un método práctico y aplicable para recolectar datos de correos electrónicos.

La recolección de datos se podría automatizar con ayuda de macros que exporten en forma mensual una parte de los correos controlados en el servidor de correos o en el Gateway que recibe inicialmente el correo y lo analiza y clasifica si contiene virus o es spam, de tal forma que no se violen leyes vigentes. De esta forma no solo se evitan posibles problemas en el uso del mismo, sino que igualmente el tamaño de la base de datos es mucho menor, especialmente por el hecho de no guardar los posibles documentos que vayan en el correo mismo.

#### 4.4.1. SPAM y otros correos no deseados

De la Información contenida en EMAILFROM y EMAILTO se pueden extraer datos importantes para el análisis, tal como el dominio, región o nombre del proveedor. Igualmente haciendo una comparación entre estos dos campos, es posible filtrar toda la información no significativa para la detección de una red social tales como el correo spam, newsletters, propaganda, correo a grupos etc.. El método recomendado es mostrado en **Gráfica 21**.



**Gráfica 21 Diagrama de flujo de datos para borrar correos SPAM**

Los siguientes problemas fueron corregidos con diferentes métodos:

- Mensajes enviados por una persona pero que son recibidos por varios. Para esto se escribió un pequeño macro en Excel que se encarga de generar un registro por cada receptor, tanto para el destinatario como para el originador.

- Las direcciones de una misma persona pueden ser dadas de diferentes formas, con o sin título, con o sin nombre, la dirección de correo electrónico con o sin designación, apellido o dirección incompleta, sobre correo general o de empresa. La corrección de esta anomalía requiere tiempo de análisis tanto visual como funcional y se hace con ayuda de tablas dinámicas.

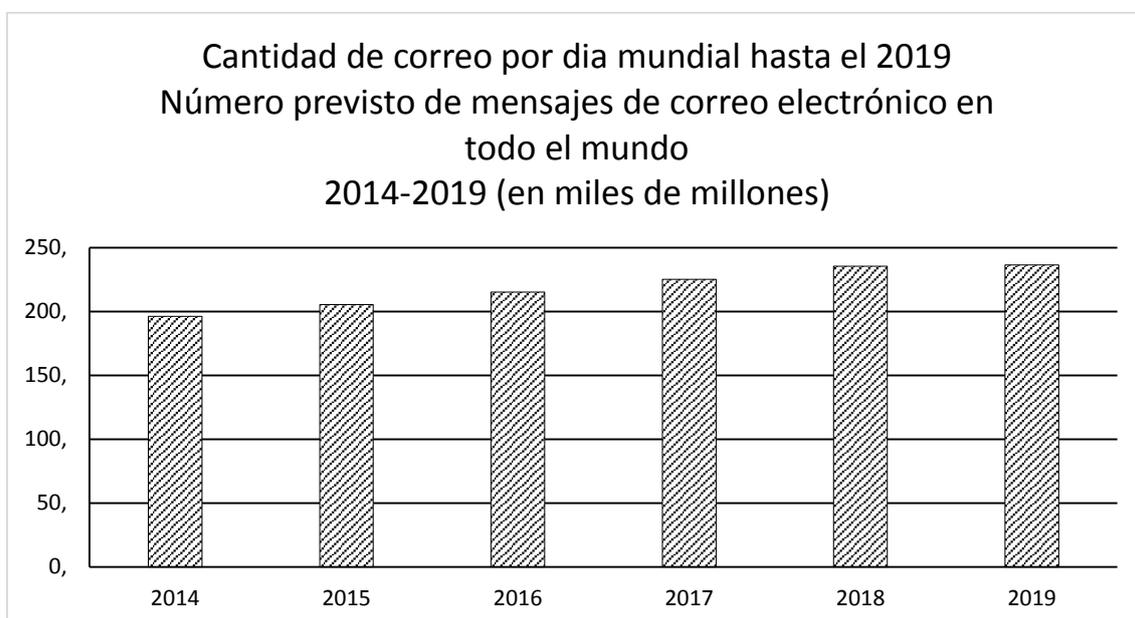
- El monto de correo llamado spam puede ser significativo, pero debido a su calidad única o a la existencia de un máximo dos (por los filtros anti-spam), eliminando los correos cuyo originador solo aparece 1 o 2 veces en el espacio analizado, se disminuye el nivel de ellos.

- Por último un correo que aparece muchas veces pero que para el análisis de la red tiene poco significado son los llamados "Newsletters". Estos son filtrados en una forma sencilla: todo aquel

originador que no existe como destinatario es excluido de la base de datos. El diagrama de flujo de datos de la **Gráfica 21** elimina éstos.

#### 4.4.2. Desarrollo a través de tiempo

El uso de correo electrónico según el The Radicati Group (en © Statista 2016), avanza continuamente como se puede apreciar en la **Gráfica 22** Pronóstico del uso de correo electrónico hasta el 2019.



**Gráfica 22 Pronóstico del uso de correo electrónico hasta el 2019**

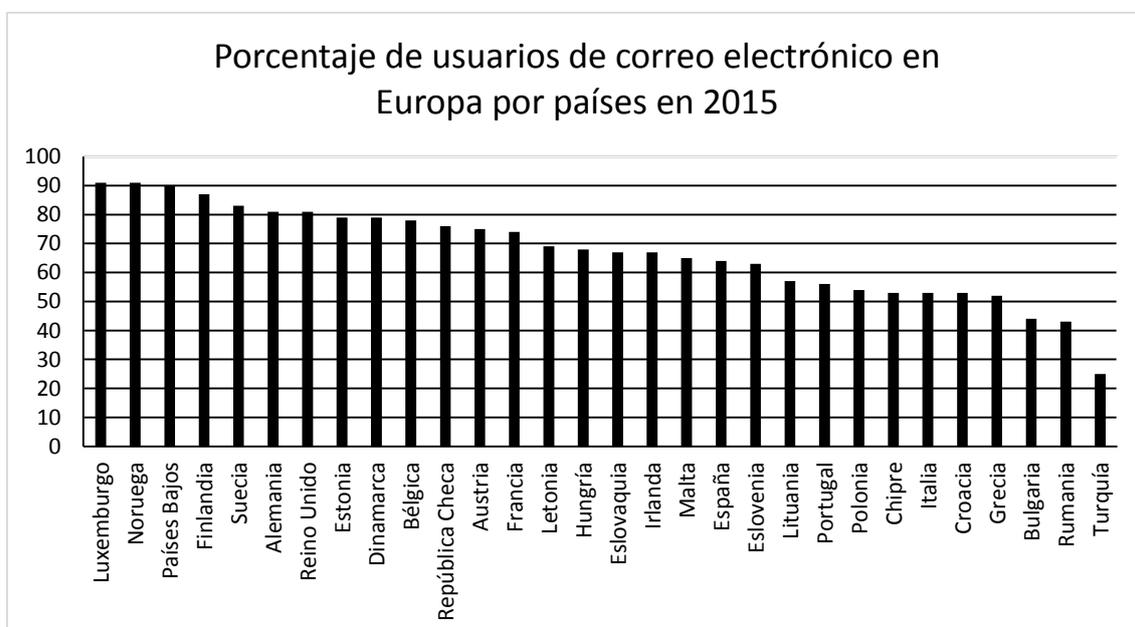
Si bien en “Business User Survey 2009” de la oficina nacional de estadística de Gran Bretaña dice que el 81% del correo electrónico es SPAM, el monto de correo electrónico sigue subiendo.

El crecimiento de la cantidad de correos por año relevantes para la investigación ha sido continuo y permanente, como se constata en el transcurso del estudio:

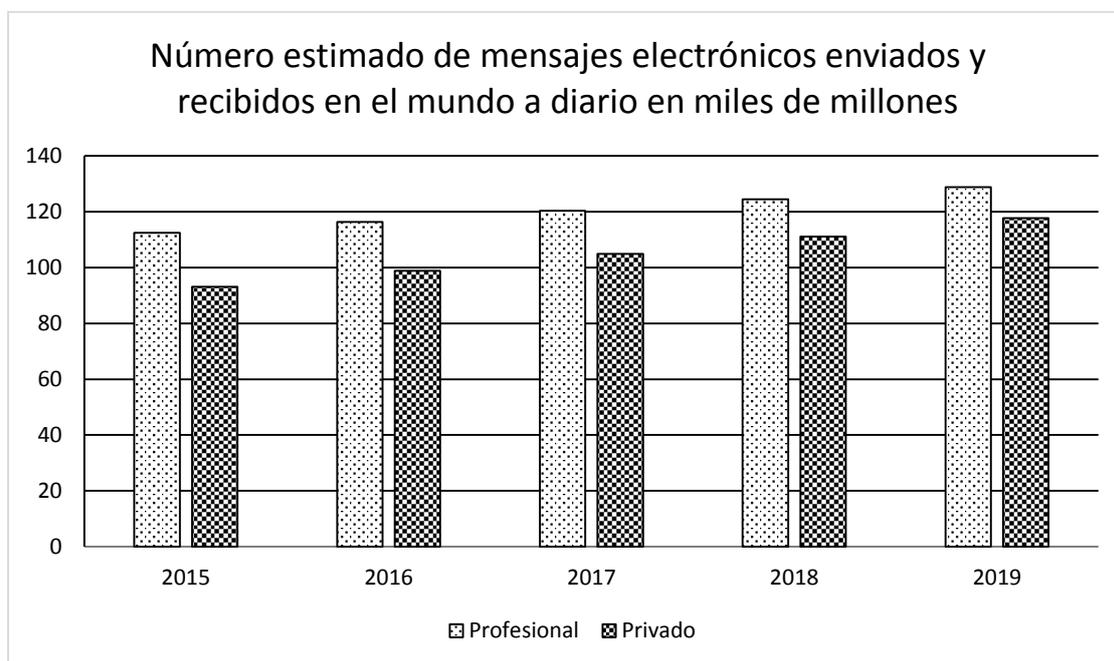
Año	#
2002	568
2003	681
2004	770
2005	1.152
2006	1.282
2007	1.456
2008	1.656
2009	1.990
2010	2.471
2011	2.591
2012	2.178
2013	2.059

**Tabla 8 Correos significantes por año**

Si proporcionamos estos valores a todos los empleados que tienen acceso a correo electrónico, el número sería mucho mayor. Según la oficina de estadística el porcentaje de usuarios de Email en Europa no es igual en todos los países, sin embargo, su uso es significativo.



**Gráfica 23 Porcentaje de usuarios de correo electrónico en Europa por países en 2015**



**Gráfica 24 Número estimado de mensajes electrónicos enviados y recibidos en el mundo a diario en miles de millones**

Las observaciones y predicciones de las diferentes oficinas de estadística a nivel Europeo, pero en especial de Statista, que es uno de los mayores portales de estadísticas del mundo<sup>18</sup>, ven un continuo crecimiento de este medio de comunicación.

<sup>18</sup> Hamburger Abendblatt, 25. März 2014

## **CAPÍTULO V. ANÁLISIS EMPÍRICO**

### **5.1. Análisis de los datos con ayuda de la teoría de grafos**

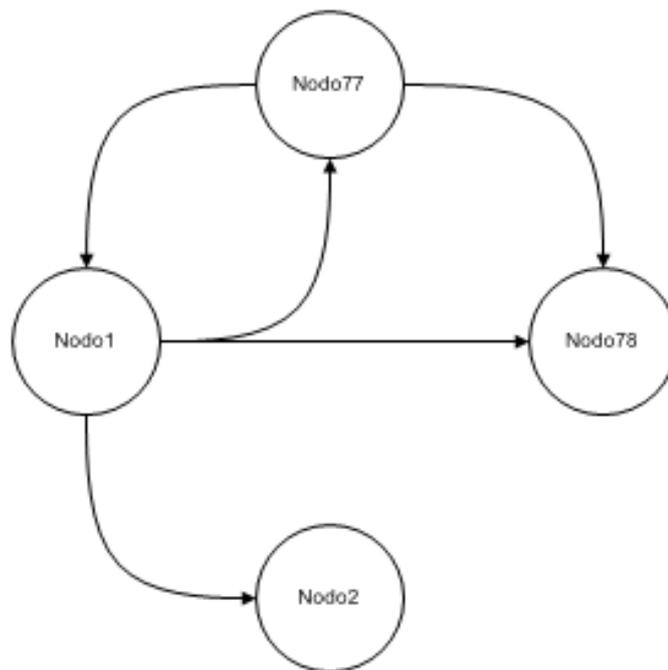
Las conexiones entre los actores en la red social se pueden mostrar en un grafo. Este grafo es ponderado (Schmidtke, 2006) tiene como vértices a los nodos de la red, las aristas corresponden a la comunicación entre los nodos y la valuación de la misma es determinada por la cantidad de correos intercambiados que son individualmente ponderado según si son de forma directa (TO:) o indirecta (CC: o BCC:).

El grafo por la particularidad de la adquisición de los datos es del tipo conexo (Schmidtke, 2006) ya que existe para cualquier nodo un camino para llegar a otro nodo. Otra particularidad del grafo obtenido es que se trata de un grafo dirigido o dígrafo (Weisstein E. W., 2012) es decir no todos los nodos tienen una conexión bilateral. Esto se observa especialmente cuando se filtran los datos en aquellos que no son influyentes a la red, tales como son Newsletters, Spam y Publicidad.

Viendo un ejemplo complejo de un intercambio de correo electrónico, se puede diseñar un grafo. Tomemos por ejemplo estos 2 correos electrónicos:

```
From: "Nodo77" <Nodo77@dominio2.zona3>
To: <Nodo1@dominio1.zona1>
CC: "Nodo78" <Nodo78@ dominio2.zona3>
From: "Nodo1"<Nodo1@dominio1.zona1>
To: "Nodo77" <Nodo77@dominio2.zona3>
CC: "Nodo78" <Nodo78@ dominio2.zona3>, "Nodo2"<Nodo2@dominio1.zona1>
```

El grafo correspondiente se puede visualizar de esta manera:



**Gráfica 25 Dígrafo ponderado o red a partir de dos correos electrónicos**

Como podemos ver tenemos un dígrafo ponderado que en la literatura es llamado formalmente una red (Weisstein E. W., 2012).

Este grafo en una aproximación inicial lo podemos medir con ayuda de matrices.

El grafo en la Gráfica 25 lo podemos mostrar con dos matrices. Una llamada matriz de incidencia que representa la interconexión de los nodos y aristas y una matriz de adyacencia en donde si hay una arista entre dos vértices entonces el elemento es 1 y de lo contrario es 0.

El dígrafo en cuestión lo llamamos  $D(V,A)$  cuyos conjuntos están definidos como  $V = \{N1, N2, N77, N78\}$

$$A = \{\{N1, N2\}, \{N1, N77\}, \{N1, N78\}, \{N77, N1\}, \{N77, N78\}\}$$

La matriz de adyacencia

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

En esta matriz tanto las columnas como las filas representan los nodos, siendo el valor solo 1 o 0 si existe una conexión o no.

La matriz de incidencia

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

En esta matriz las columnas representas las aristas y las filas los nodos, dándose el valor de 1 si el nodo usa esa arista y cero por el contrario.

La matriz de incidencia sin embargo para nuestro análisis, debido a su característica no cuadrada no podemos usarla, por eso me concentro solo en la matriz de adyacencia que en el transcurso de la investigación la usaremos en formas diferentes. En principio solo una matriz cuadrada puede darnos valores lineales que podemos observar en transcurso del tiempo. El Determinante de la matriz puede ser un valor lineal que podemos analizar al igual que su valor inverso (Weisstein E. W., 2012).

Estos valores son fáciles de calcular con ayuda de Excel, ya que existen las funciones MDET()<sup>19</sup> y MINV()<sup>20</sup>.

Como se dijo en un principio los grafos a usar son ponderados. Una forma de darle valor a las aristas es tener en cuenta la cantidad de correos enviados de un nodo a otro y darle un valor de ponderación a cada uno de los valores individuales. En este caso sin tener una razón diferente a la de buscar un significado, decido darle el valor de 1 a las conexiones directas y de 0,5 a las conexiones indirectas.

Veamos como se ve la matriz en este caso:

---

<sup>19</sup> Determinante de la matriz

<sup>20</sup> Inverso de la matriz

$$\begin{pmatrix} 0 & 0,5 & 2 & 0,5 \\ 0,5 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0,5 \\ 0,5 & 0 & 0,5 & 0 \end{pmatrix}$$

El Determinante de esta matriz sería 0,0625 y el valor inverso 0 ya que hay muy poca interacción entre los nodos.

Veamos un ejemplo más complejo, basado en los datos adquiridos en un mes específico:

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17
N1	0	20,5	12,5	3	7,5	22,5	7	8	0	11	22	14	13	12	34	2	1
N2	72	0	7	91	12	17	47	24	44	3	33	74	97	99	99	38	42
N3	3	26	0	15	82	82	16	30	50	51	13	34	79	18	86	61	27
N4	4	96	92	0	68	71	98	91	98	36	89	77	29	55	86	14	4
N5	42	76	51	41	0	67	12	8	64	43	36	93	73	54	21	31	47
N6	76	83	6	61	8	0	68	92	68	44	17	90	11	30	7	20	17
N7	76	28	52	97	30	24	0	75	90	17	61	51	59	63	75	52	17
N8	73	12	14	77	27	40	1	0	11	88	72	39	36	98	28	62	88
N9	27	28	33	3	26	51	92	18	0	61	46	70	18	12	52	41	73
N10	30	4	62	3	47	33	37	85	10	0	57	8	23	15	15	73	87
N11	11	26	36	36	73	19	68	14	42	60	0	60	61	48	90	50	48
N12	45	42	50	36	34	7	4	94	55	72	6	0	59	64	6	40	44
N13	60	27	90	7	84	96	88	33	37	8	11	95	0	73	75	20	39
N14	21	71	91	59	59	40	27	41	49	50	20	49	44	0	33	62	84
N15	52	85	89	33	20	35	37	95	26	39	73	70	1	31	0	91	44
N16	16	32	7	35	68	23	70	56	19	16	73	25	100	73	21	0	24
N17	85	41	39	2	46	61	78	14	32	96	28	59	87	86	92	78	0

El Determinante de esta matriz es 3,93208E+32 y el valor inverso 0,024020447

Para otro mes:

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15
N1	0	20,5	12,5	3	7,5	22,5	7	8	0	11	22	14	13	12	34
N2	73	0	14	95	13	90	30	12	58	51	16	38	3	67	57
N3	5	79	0	28	1	82	23	7	23	24	84	54	79	84	32
N4	44	52	7	0	36	28	80	33	65	74	68	94	74	82	46
N5	48	82	17	51	0	49	25	28	48	52	42	82	22	11	92
N6	26	71	10	58	9	0	64	12	69	45	30	23	15	46	66
N7	30	19	57	74	15	36	0	91	23	79	96	35	88	22	79
N8	66	20	33	62	100	24	3	0	70	18	13	59	81	89	52
N9	50	84	77	69	93	7	91	61	0	10	51	88	60	58	74
N10	33	4	53	80	10	26	27	85	94	0	43	98	71	86	68
N11	7	31	29	40	6	96	66	37	78	89	0	1	69	79	2
N12	30	32	76	54	78	89	16	40	20	78	50	0	77	14	30
N13	82	92	72	15	80	2	67	80	25	70	69	95	0	26	30
N14	63	57	64	69	19	93	29	26	58	80	75	5	19	0	21
N15	66	20	86	73	33	68	99	37	91	99	50	2	77	16	0

El Determinante de esta matriz es  $-4,34028E+27$  y el valor inverso  $-0,017986275$

Por ahora no nos interesa el significado de estos valores, más si el tener un valor específico de la red para los meses escogidos.

Con la ayuda de grafos se visualiza la evolución de las redes en el transcurso del tiempo y los valores de las matrices se adjuntan a otros valores típicos de las redes sociales, los cuales se mostrarán más adelante.

Ahora la pregunta lógica sería si es posible obtener más información partiendo de la teoría de grafos. Por ejemplo, el calcular el

diámetro o el diámetro promedio pueden ser valores interesantes en una investigación. Para la matriz en Gráfica 25 el diámetro o distancia mayor entre los nodos, tomando como valor de las aristas 1 y 0,5 de acuerdo a si la comunicación es directa o indirecta, ofrece un valor de 2 ( $N78 \Rightarrow N2$ ) y el de diámetro promedio es de 1,08824. Estos valores se obtendrán con ayuda del software UCINET (Borgatti, Everett, & Freeman, 2012) y calculados en forma general con ayuda del programa R.

## 5.2. La red Social

Una vez clasificados los integrantes de la red social, se filtran todos aquellos que a través de todo el periodo objeto de investigación (de 1.1.2002 al 31.12.2013) o solo son receptores o sólo han enviado mensajes. Por ejemplo, estos posibles candidatos no van a ser tenidos en cuenta:

EMAIL	DOMINIO	GRUPO	Nº ENVIADOS	Nº RECIBIDOS
UNK-001@google.com	GOOGLE.COM	UNKNOWN	1526	0
INT-120@fmo.de	FMO.DE	INTERNAL	815	0
INT-007@fmo.de	FMO.DE	INTERNAL	0	3996
PRI-250@t-online.de	T-ONLINE.DE	PRIVAT	0	227

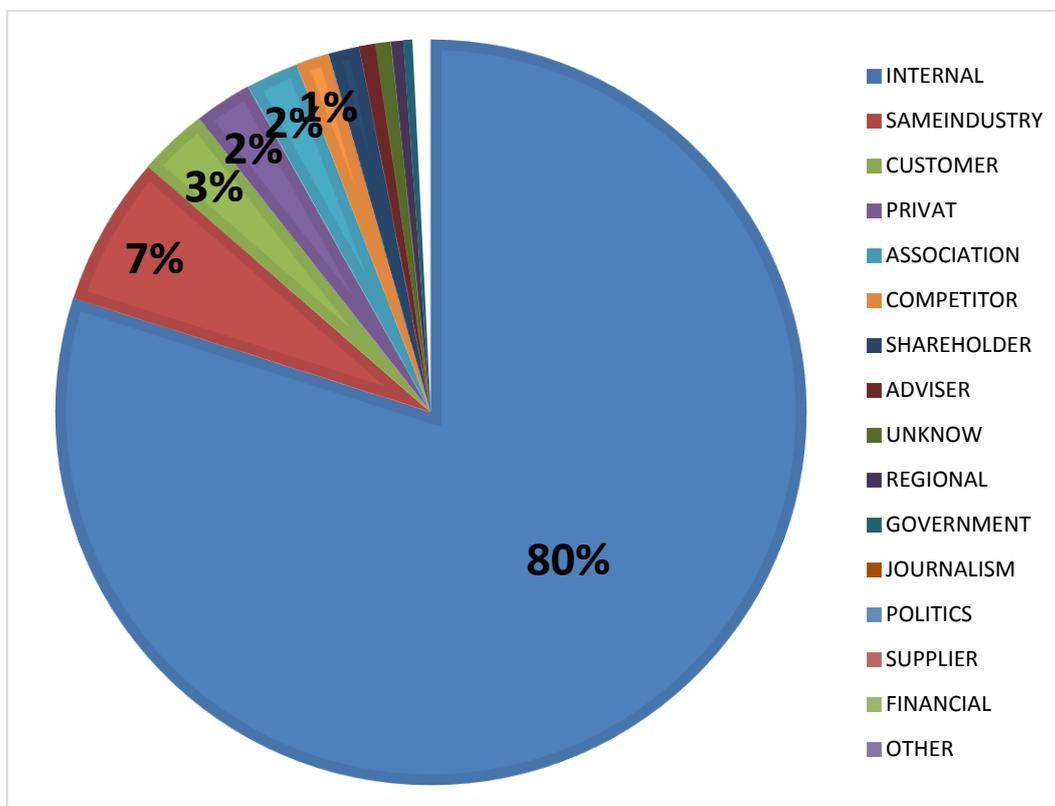
**Tabla 9 Ejemplo de miembros que no se tienen en cuenta en la evaluación**

UNK-001 es en este caso googlealerts-noreply@google.com, que es usado como abono de información en internet. Los otros son listas de correo o que envían "newsletters" o SPAM.

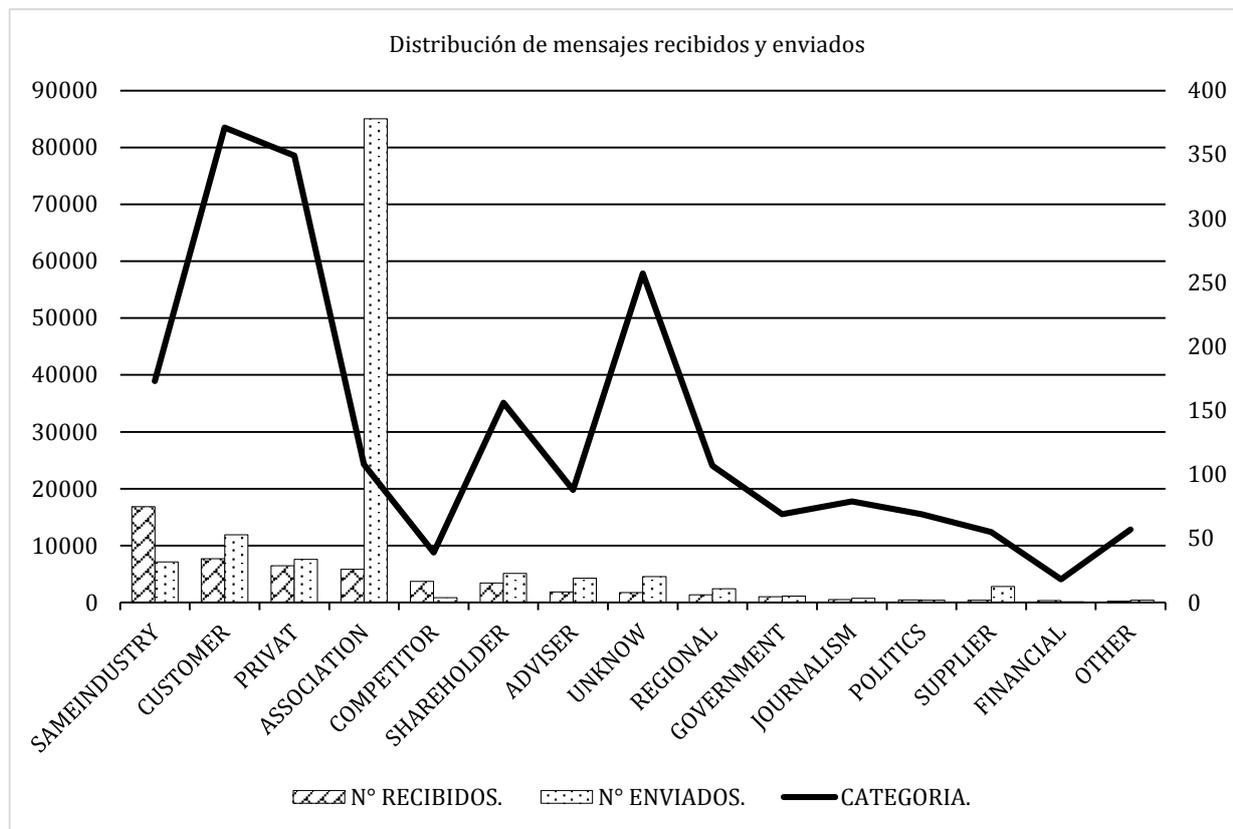
En la primera observación de los datos vemos una tendencia clara de que los miembros de un grupo específico juegan un papel importante en la red social. Cabe anotar que la clasificación de los grupos podría haber sido más pequeña. Es así como el grupo de consejería (ADVISER) y del sector financiero (FINANCIAL) se podrían catalogar como proveedor (SUPPLIER), los grupos de otros (OTHER), de carácter regional (REGIONAL), de los competidores (COMPETITOR) y pertenecientes a la misma industria (SAMEINDUSTRY) podrían sumarse en un conjunto para reducir la cantidad de grupos. Mas, sin embargo, creo que es de suma importancia conocer los grupos que pueden influenciar el desarrollo de una empresa en forma detallada. Los grupos de carácter político (POLITICS) y de gobierno (GOVERNMENT) están en este caso muy enlazados con el grupo de dueños de la empresa (SHAREHOLDER), ya que todas las decisiones importantes, en especial en este caso porque se trata de un aeropuerto regional, tienen influencia pública. Es por eso mismo que el grupo de periodistas (JOURNALISM) por el carácter público de la empresa pasa a tomar un carácter especial y por eso se debe tener en cuenta por separado. El último grupo escogido es el de las asociaciones (ASSOCIATION) en las cuales la compañía tiene representación ya que por su carácter y forma pueden influenciar el desarrollo de la empresa sustancialmente. En este caso son aquellas organizaciones que reglamentan el tráfico aéreo en una u otra forma. Entre ellas están la

International Civil Aviation Organization (ICAO), Airports Council International (ACI), International Air Transport Association (IATA), la Organización de Aeropuertos de habla alemana (ADV), la comunidad Económica Europea, Deutsche Flugsicherung (DFS) y EUROCONTROL entre otras.

Como se ve en los diagramas siguientes el porcentaje de mensajes e integrantes de la red es mayoritariamente de carácter puro interno. En efecto, no sería un buen indicio si el porcentaje de intercambio de correo electrónico de carácter interno hubiera sido menor. Se supone que el correo electrónico es un instrumento de gestión importante en la actualidad.



**Gráfica 26 Distribución por grupo de los integrantes de la red social**



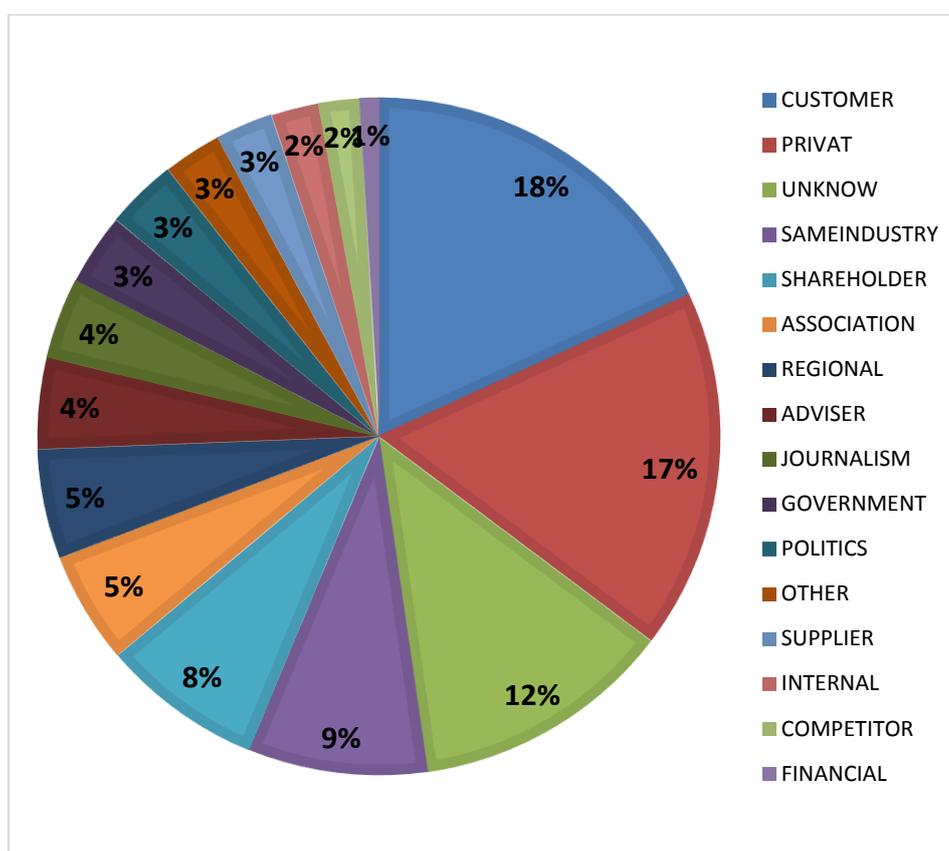
**Gráfica 27 Distribución de mensajes enviados y recibidos**

Como se describió al principio de la investigación, he dividido ahora la red en una de carácter puro interna y otra de carácter externa. La red social implicada o descrita con estos valores de carácter interno no es parte de la investigación, aunque puede tener un valor mucho más valioso. Ese paradigma se deja para una investigación futura o complementaria.

Una vez clasificados los mensajes solo de carácter externo, quedan 185.276 mensajes. Cabe anotar que es posible que por el diseño de la estructura de la investigación los mensajes que son enviados tanto a miembros de la empresa como a no miembros sean

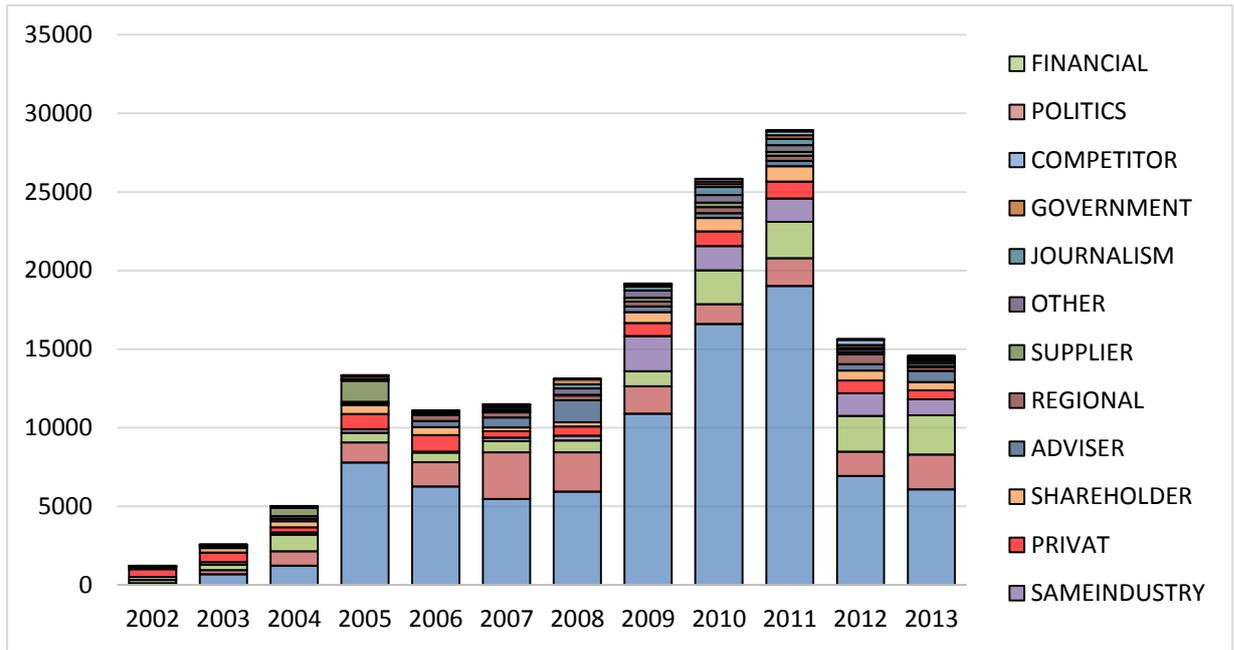
tomados en cuenta solo parcialmente. Para poder usar la información disponible fue necesario hacer un control de calidad exhausto de su contenido (Rodriguez F. , 2010).

De los 185.276 mensajes eliminando con el método descrito en la **Gráfica 21**, quedan para objeto de la evaluación 69.049 mensajes, distribuidos en estos grupos:



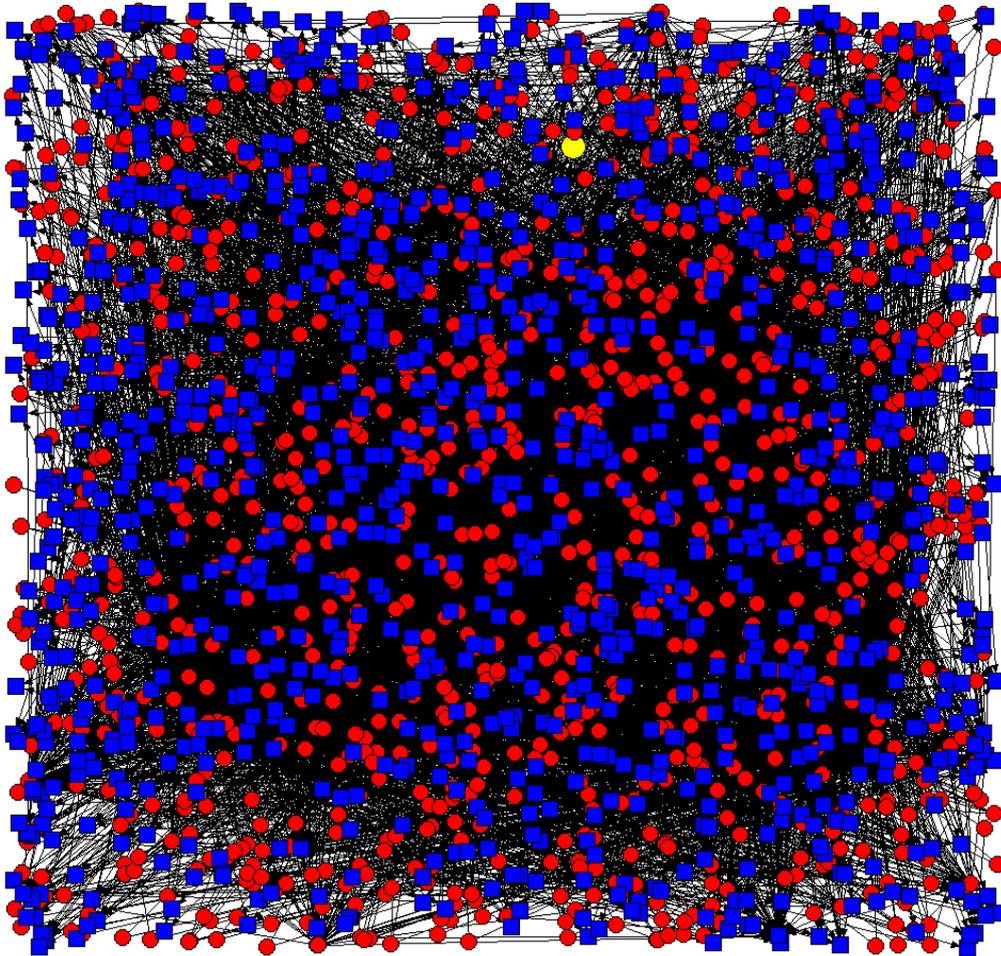
**Gráfica 28 Distribución de los integrantes de la red social según grupo (solo externos)**

Veamos cómo se desarrolla a través del tiempo:



**Gráfica 29 Cantidad de mensajes a través del tiempo por grupos y años**

### 5.3. Análisis de los Datos de la red social con UCINET y R



**Gráfica 30 Visualización de la red de modo 2 con UCINET basada en intercambio de correo electrónico por el director de la empresa**

En la Gráfica 30 vemos la primera visualización de todos los datos de uno de los actores de la red social motivo de investigación se trata de una red de modo dos donde los actores se representan con puntos rojos y a las organizaciones a que pertenecen en cuadrados azules. El actor principal es demarcado con un círculo amarillo. Esta

primera visualización demuestra la espera centralidad de la red, ya que su origen es a partir de un actor y no hay actor alguno que no tenga conexión directa o indirecta con él.

Para el análisis de los datos de la red social nos limitamos, como ya he dicho anteriormente (2.2.5) a valores globales de la red y no a valores particulares de un actor determinado. UCINET nos da en particular estos valores llamados indicadores estructurales de una red. Ellos son:

- Transitividad
- Densidad Distancia Geodésica
- Cantidad de componentes
- Grado
- Cantidad de conexiones
- Diámetro del grafo
- Cantidad de aristas
- Cantidad de vértices

Estos valores son estimados para cada mes, dando como resultado valores de la evolución de la red social en función del tiempo.

Transitividad según Gabor Csardi (2006) autor de esta función en R, es la probabilidad de que los vértices adyacentes estén conectados. Esto es a veces también llamado el coeficiente de agrupamiento.

Densidad según autor Carter T. Butts (2013) de esta función en R en el paquete igraph, es la densidad de los gráficos indicados ajustándolos según el tipo de grafo. En nuestro caso son siempre directos, es decir

que una conexión en un sentido no implica que el sentido contrario exista.

Grado es manejado en forma algo diferente en los paquetes sna e igraph en R. A los efectos de comprobar los valores más cercanos a los obtenidos en UCINET, utilizo la función básica de R de suma en una matriz y no uso las herramientas de los paquetes.

Los otros valores escogidos para ser analizados (cantidad de componentes, cantidad de conexiones, diámetro del grafo, cantidad de aristas y cantidad de vértices) junto con los calculados, son normalizados a una escala entre 1 y 2 para que puedan ser comparados.

Para ello se utiliza la formula

$$X_n = \frac{X_i - \min_{1,j} X}{\max_{1,j} X - \min_{1,j} X} + 1 \quad (10)$$

donde  $X_n$  es el valor normalizado entre 1 y 2 y  $X_i$  el valor original.

La razón de esta normalización es la de evitar el valor "0" que para posibles divisiones no sería posible.

Los datos recogidos tienen las siguientes características:

TRANSITIVIDAD	COMPONENTES	GRADO	DENSIDAD
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1st Qu.:1.297	1st Qu.:1.289	1st Qu.:1.093	1st Qu.:1.084
Median :1.427	Median :1.461	Median :1.191	Median :1.131
Mean :1.418	Mean :1.479	Mean :1.233	Mean :1.210
3rd Qu.:1.524	3rd Qu.:1.705	3rd Qu.:1.330	3rd Qu.:1.231
Max. :2.000	Max. :2.000	Max. :2.000	Max. :2.000

CONEXIONES	DIAMETRO	VERTICES	ARISTAS
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000
1st Qu.:1.087	1st Qu.:1.152	1st Qu.:1.245	1st Qu.:1.149
Median :1.228	Median :1.255	Median :1.439	Median :1.326
Mean :1.318	Mean :1.281	Mean :1.466	Mean :1.340
3rd Qu.:1.542	3rd Qu.:1.373	3rd Qu.:1.716	3rd Qu.:1.530
Max. :2.000	Max. :2.000	Max. :2.000	Max. :2.000

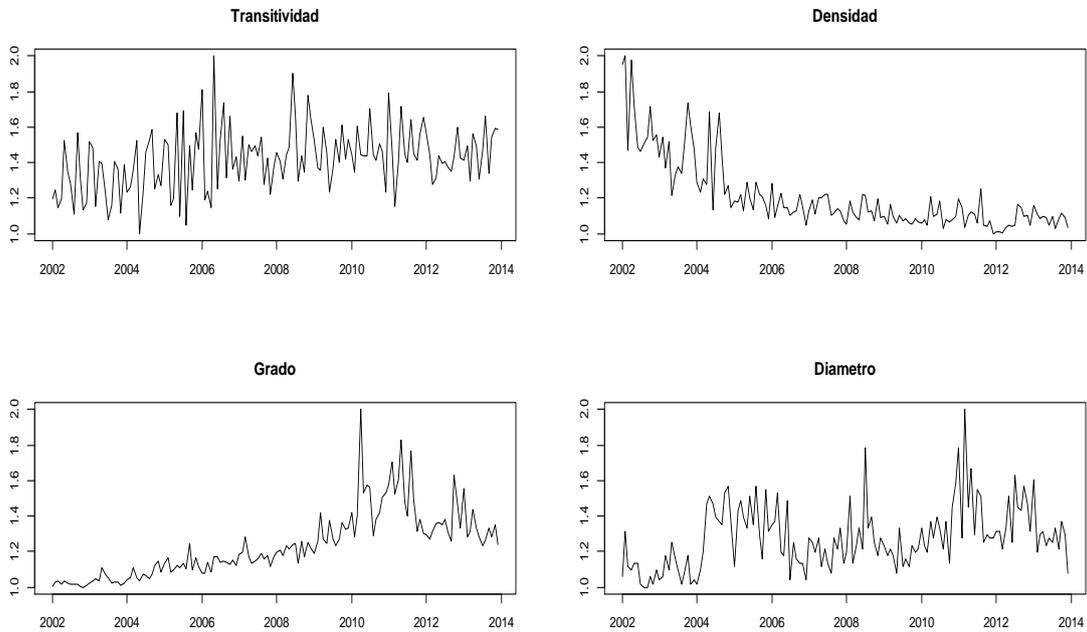
COVARIANZA	CORRELACION		
Min. :1.000	Min. :1.000		
1st Qu.:1.025	1st Qu.:1.048		
Median :1.030	Median :1.065		
Mean :1.056	Mean :1.119		
3rd Qu.:1.048	3rd Qu.:1.128		
Max. :2.000	Max. :2.000		

**Tabla 10 Resumen de los datos a ser analizados**

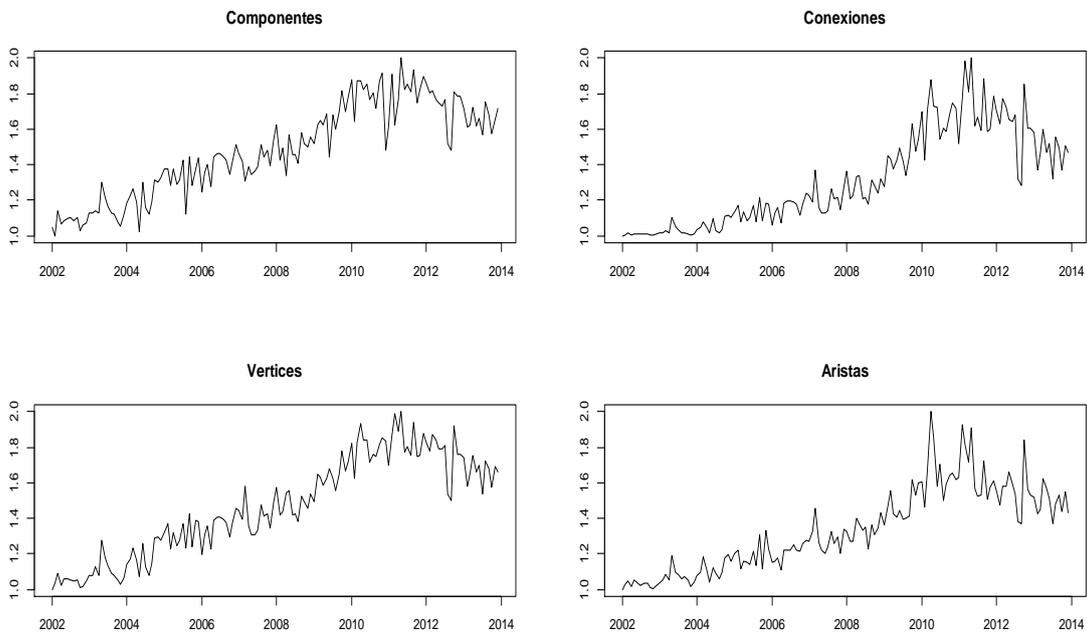
Se observa que la mediana de la covarianza es muy baja, lo mismo en la correlación. Esto lo podemos interpretar en el sentido de que los cambios o evolución de la red, desde el punto de vista mensual, son lentos. La similitud en el desarrollo de valores métricos lineales tales como cantidad de componentes, conexiones, vértices y aristas es obvia. Los otros valores son objeto de un análisis más intensivo para deducir su significado en el desarrollo de la red.

Veamos visualmente el comportamiento de cada valor en función del tiempo.

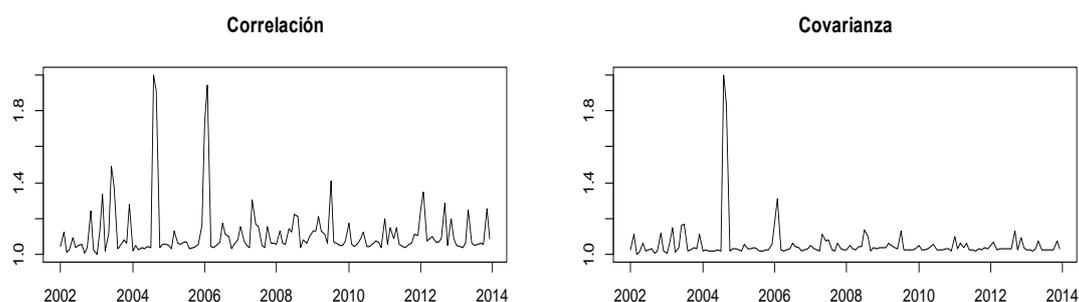
Estas son las gráficas obtenidas, visualizadas con R y comprobadas en Excel:



**Gráfica 31 Transitividad, Densidad, Grado y Diametro de la red normalizada**



**Gráfica 32 Componentes, Conexiones, Vértices y Aristas de la red normalizada**



**Gráfica 33 Correlación y Covarianza de la red normalizada**

Las gráficas confirman los valores numéricos analizados en la Tabla 10

Se observa que la evolución de los indicadores relativos a componentes, grado, número de conexiones, aristas y vértices es muy parecido por lo que se puede simplificar la búsqueda de un valor referencial a uno de estos valores. Por la característica de la red social asumo en el análisis solo el número de conexiones. De esta forma se tienen ahora como valores que se pueden tener en cuenta en el análisis de regresión las siguientes variables independientes:

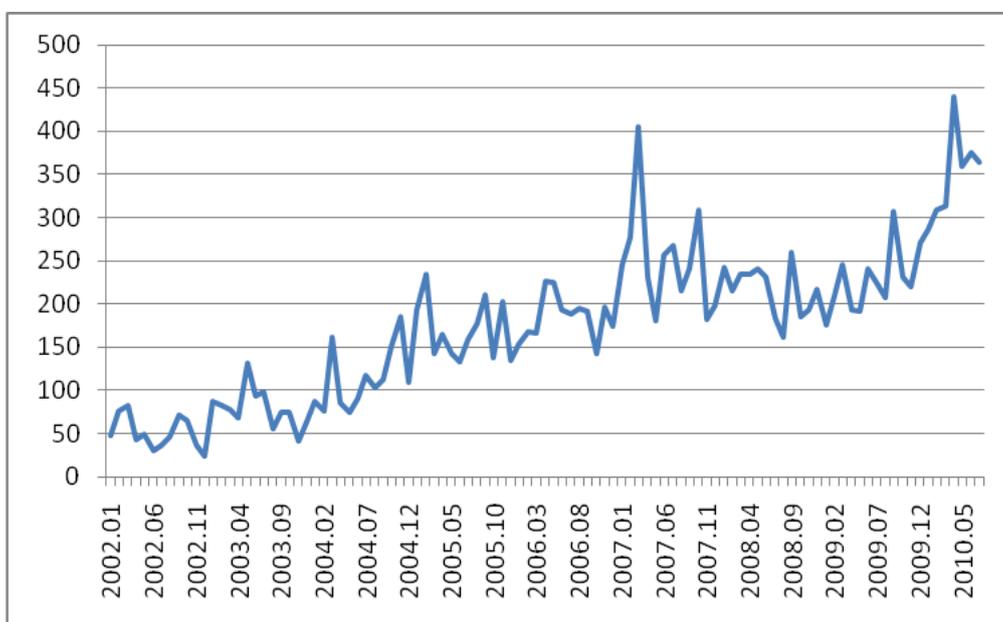
- Transitividad
- Densidad Distancia Geodésica
- Cantidad de componentes
- Grado de la red
- Diámetro del grafo

Estos valores se refieren únicamente a la red compuesta por los enlaces exteriores a los integrantes de la compañía. Un análisis de la red social interna y su influencia en el desarrollo de la compañía es parte de lo propuesto en el apartado 8.1 Conclusiones.

### 5.3.1. Análisis con UCINET

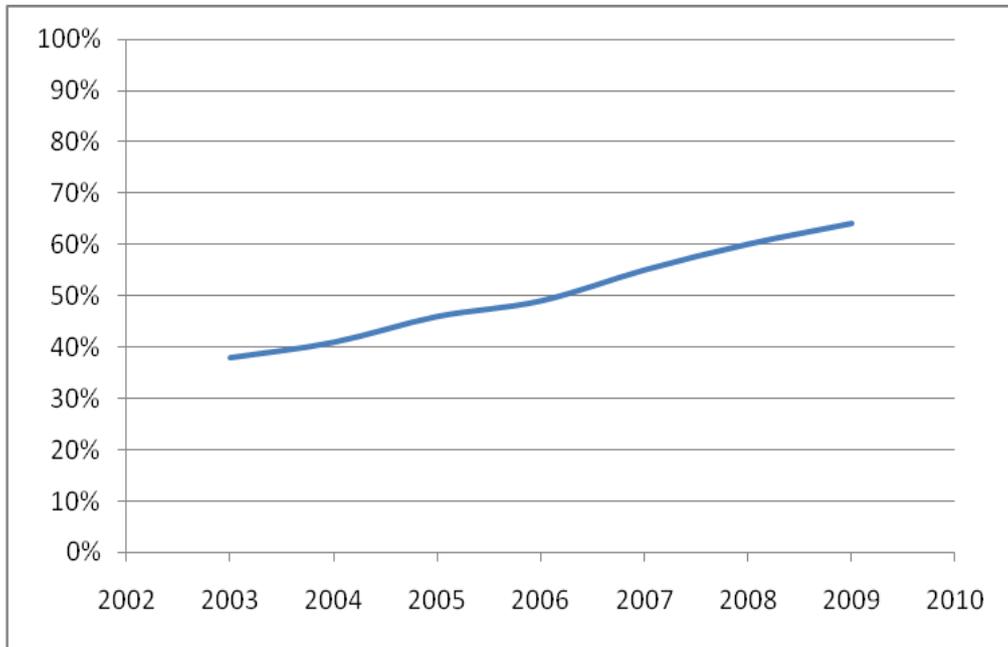
Un problema fundamental para efectuar un análisis de las redes es la adquisición de la base de datos, su cantidad y su calidad. La forma clásica de adquirir datos es el desarrollo de cuestionarios con preguntas u observaciones y según su diseño y validación de las respuestas se podrá o no hacer un análisis de la red y sacar conclusiones. El desarrollo de la tecnología nos da una oportunidad de efectuar una exploración del desarrollo de las redes en una forma automática, efectiva, precisa y prácticamente en *real-time*.

La distribución del monto de correo electrónico a través de los años da cuenta del desarrollo de la intensidad del mismo



**Gráfica 34 Cantidad de emails distribuidas por meses**

Comparado con el desarrollo del uso de correo electrónico en Europa, muestra el desarrollo del mismo en la empresa un desarrollo típico a nivel europeo:



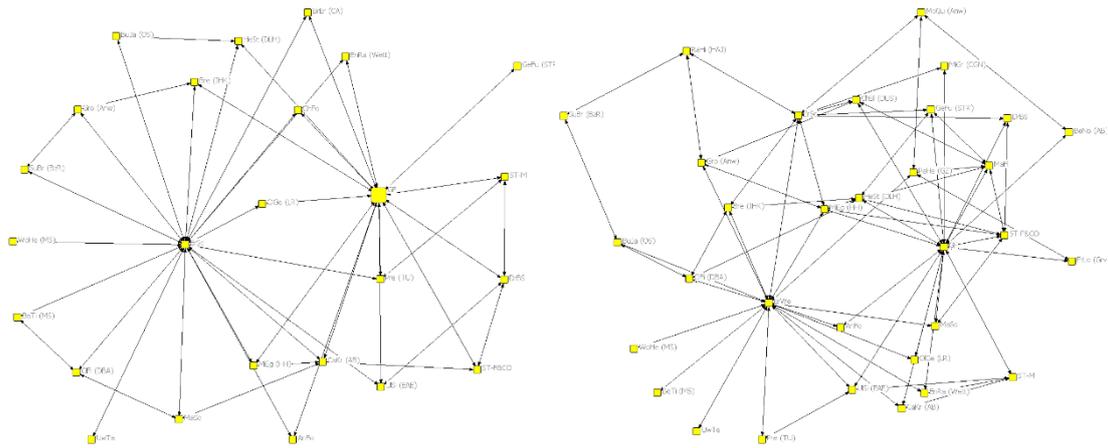
**Gráfica 35 Porcentaje de personas en el UE entre 16 y 74 años que usan regularmente email<sup>21</sup>**

Habiendo consolidado y comprobado los datos a ser utilizados, se concreta que los siguientes datos están disponibles para un análisis adecuado, de acuerdo al protocolo de importación del programa UCINET, utilizado para el análisis redes sociales.

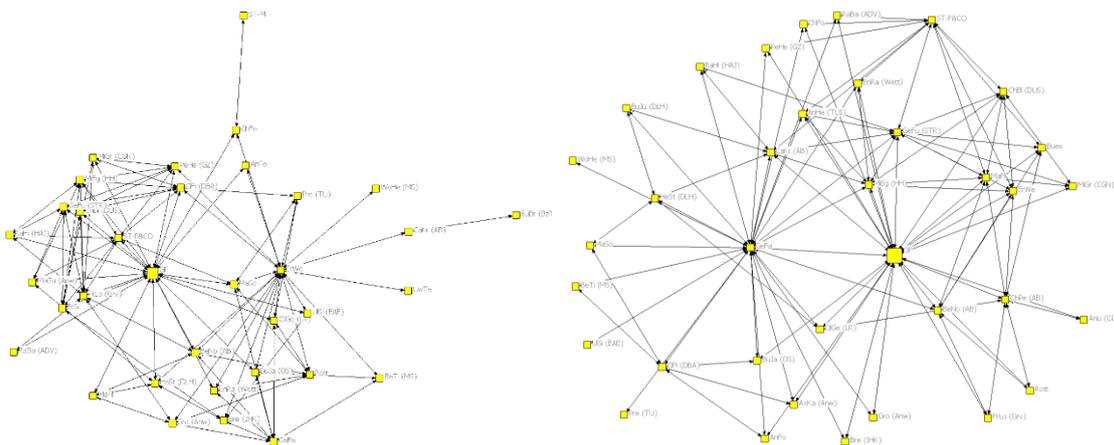
Con ayuda de UCINET visualizamos el desarrollo de la red social tomando como parámetros la cantidad de correo intercambiado

<sup>21</sup> Eurostat – Oficina de estadística de la Unión Europea  
<http://epp.eurostat.ec.europa.eu>





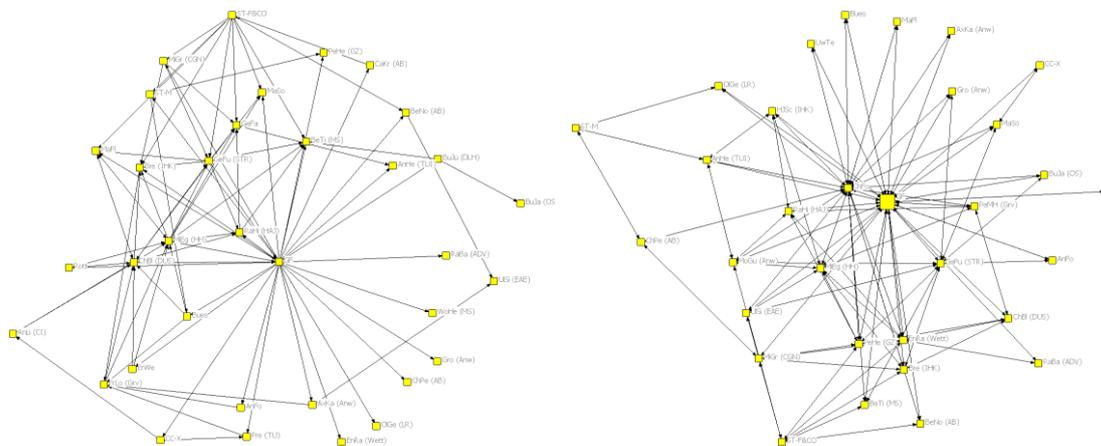
**Gráfica 37 Red social años 2004 y 2005**



**Gráfica 38 Red social años 2006 y 2007**

El año 2007 muestra una intensidad mayor que en los otros años. Esto se debe a los cambios en el ambiente económico con muchos riesgos de la compañía. En el año 2007 se inició uno de los programas más agresivos en cuestión de mejoramiento de procesos desde el punto de vista económico, llamado "TuneUp". Este plan de ahorro implicó una gran coordinación e intercambio con todas las identidades adscritas a la compañía, dando al final un efecto alto

positivo, cuyos resultados mayores se ven en la balanza positiva del año 2009 especialmente<sup>22</sup>



**Gráfica 39 Red social años 2008 y 2009**

Se observa un crecimiento en la complejidad y magnitud de la red. Los miembros de la red crecieron en una forma mayor en el año 2007. Esto concuerda también con el monto de emails mayor en ese mismo año (Gráfica 34).

Para el análisis se presta un Sociograma, en el cual se le pueden aplicar la teoría de grafos y matrices (Jansen 1999, pp. 85). Los actores son presentados como puntos (Gráficas anteriores) y la relación entre ellos es mostrada con líneas.

Para comparar el desarrollo de la red veamos algunas características de la misma en el transcurso del tiempo. Veamos el

<sup>22</sup> Página de web de la compañía: <http://www.fmo.de>

grado de centralidad de los actores según Linton Freeman, uno de los autores del software UCINET.

	2002	Degree	NrmDegree	Share
2	Bre (IHK)	101.000	7.282	0.377
3	BrEr (CA)	75.000	5.407	0.280
9	GF	24.000	1.730	0.090
4	BuJü (DLH)	21.000	1.514	0.078

**Tabla 11 Correo Electrónico 2002**

Los actores 2 y 3 tienen el mayor grado y podrían ser considerados como los más influyentes. El actor 2 pertenece a círculo de influencia de la compañía en el sector de dueños y los actores 3 y 4 son miembros de compañías aéreas. Una forma clara del comienzo de la red: Dueños y clientes son los primeros en formar parte de la red social.

	2003	Degree	NrmDegree	Share
11	GF	135.000	15.000	0.210
4	BrEr (CA)	113.000	12.556	0.175
9	ErSc	71.000	7.889	0.110
6	BuJü (DLH)	70.000	7.778	0.109
2	BeNo (AB)	42.000	4.667	0.065
15	MiEg (HH)	29.000	3.222	0.045

**Tabla 12 Correo Electrónico 2003**

Se observa como la influencia del actor 11, objeto de este estudio principal, toma forma definitiva. El contacto con los clientes mayores, como lo son las compañías aéreas y especialmente con el actor 4 es dominante. Esto es explicable, ya que las actuaciones e intenciones de la compañía en frenar la caída de la economía después de los hechos del 11 de septiembre 2001 en New York, se muestran

también en el desarrollo de la red social en torno a la gestión de gerencia (actor 11). El actor 4, gerente de una compañía de aviación, decide en este año cambiar su Home-Base, gracias a las ofertas del actor 11 como gerente del Aeropuerto. Pero también la parte de concepto privado, toma parte de la red. El actor 9, muy probablemente de carácter no de negocios directos parece tener una influencia en la red no discutible.

	2004	Degree	NrmDegree	Share
8	ErSc	182.000	12.281	0.211
13	GF	167.000	11.269	0.193
5	DrBS	48.000	3.239	0.056
21	ST-M	42.000	2.834	0.049
16	CaKr (AB)	37.000	2.497	0.043
19	Ulsi (EAE)	35.000	2.362	0.041
6	MiEg (HH)	33.000	2.227	0.038
4	BuJa (OS)	31.000	2.092	0.036
24	BeTi (MS)	29.000	1.957	0.034
12	OlGe (LR)	28.000	1.889	0.032
7	EnRa (Wett)	28.000	1.889	0.032
18	Pre (TU)	28.000	1.889	0.032
20	ST-F&CO	26.000	1.754	0.030
14	Gro (Anw)	24.000	1.619	0.028
22	SuBr (BzR)	23.000	1.552	0.027
15	HeSt (DLH)	19.000	1.282	0.022
10	ChFo	18.000	1.215	0.021
1	AnFo	14.000	0.945	0.016

**Tabla 13 Correo Electrónico 2004**

Lo ya observado en el año 2003 se confirma en el 2004. El actor 8 toma un puesto bastante influyente en el contexto. Cabe destacar que, si bien el correo interno de la compañía no es asunto de esta investigación, algunos integrantes como lo son los actores 20 y

21 asumen una posición bastante influyente en el desarrollo de la red en torno a la gestión. La intensidad en el intercambio con la administración regional (política) está ahora bastante presente (actores 4, 24, 12, 7 y 22).

	2005	Degree	NrmDegree	Share
10	ErWe	391.000	11.426	0.216
17	GF	284.000	8.299	0.157
12	MaFl	84.000	2.455	0.046
8	EnRa (Wett)	65.000	1.899	0.036
20	HeSt (DLH)	61.000	1.783	0.034
5	BuJa (OS)	61.000	1.783	0.034
7	MiEg (HH)	59.000	1.724	0.033
19	PeHe (GZ)	57.000	1.666	0.031
26	UlsI (EAE)	52.000	1.520	0.029
14	GeFu (STR)	52.000	1.520	0.029
16	OlGe (LR)	48.000	1.403	0.026
31	BeTi (MS)	48.000	1.403	0.026
23	MaSo	46.000	1.344	0.025
28	ST-M	43.000	1.257	0.024
4	Bre (IHK)	43.000	1.257	0.024
30	UwTe	42.000	1.227	0.023
3	ChBl (DUS)	39.000	1.140	0.021
22	CaKr (AB)	38.000	1.110	0.021
1	AnFo	37.000	1.081	0.020
21	RaHi (HAJ)	37.000	1.081	0.020
29	SuBr (BzR)	35.000	1.023	0.019
9	ErSc	34.000	0.994	0.019
11	ClFi (DBA)	31.000	0.906	0.017
27	ST-F&CO	26.000	0.760	0.014

**Tabla 14 Correo Electrónico 2005**

	2006	Degree	NrmDegree	Share
10	ErWe	402.000	24.130	0.201
19	GF	232.000	13.926	0.116
12	ClFi (DBA)	133.000	7.983	0.066
25	MaSo	89.000	5.342	0.044
30	ST-F&CO	88.000	5.282	0.044
22	HeSt (DLH)	80.000	4.802	0.040

16	GeFu (STR)	78.000	4.682	0.039
8	EnRa (Wett)	75.000	4.502	0.037
7	MiEg (HH)	66.000	3.962	0.033
15	FrLo (Grv)	62.000	3.721	0.031
5	Bre (IHK)	59.000	3.541	0.029
18	OlGe (LR)	56.000	3.361	0.028
1	AnFo	54.000	3.241	0.027
4	ChBl (DUS)	54.000	3.241	0.027
34	BeTi (MS)	52.000	3.121	0.026
6	BuJa (OS)	52.000	3.121	0.026
21	PeHe (GZ)	46.000	2.761	0.023
23	RaHi (HAJ)	40.000	2.401	0.020
9	ErSc	35.000	2.101	0.017
26	MoGu (Anw)	33.000	1.981	0.016
20	Gro (Anw)	28.000	1.681	0.014
3	BeNo (AB)	26.000	1.561	0.013
24	CaKr (AB)	24.000	1.441	0.012

**Tabla 15 Correo Electrónico 2006**

Los años 2005 y 2006 en sus características, a parte de un crecimiento constante y estable de la red, no muestran cambios sustanciales. El intercambio entre los miembros del sector, es decir con otros aeropuertos no es significativo en este momento. La influencia en la red de los miembros de carácter privado (actores 1 y 9) es permanente.

		Degree	NrmDegree	Share
14	GeFa	524.000	23.764	0.166
22	GF	452.000	20.499	0.143
10	ChPe (AB)	186.000	8.435	0.059
16	MaFl	162.000	7.347	0.051
28	CaKr (AB)	159.000	7.211	0.050
5	ChBl (DUS)	148.000	6.712	0.047
19	GeFu (STR)	146.000	6.621	0.046
11	MiEg (HH)	123.000	5.578	0.039
34	ST-F&CO	110.000	4.989	0.035
21	OlGe (LR)	87.000	3.946	0.027

4	BeNo (AB)	86.000	3.900	0.027
13	ErWe	82.000	3.719	0.026
9	Bues	76.000	3.447	0.024
32	Rott	69.000	3.129	0.022
8	BuJa (OS)	67.000	3.039	0.021
29	AnLi (CI)	67.000	3.039	0.021
1	AnFo	63.000	2.857	0.020
18	FrLo (Grv)	63.000	2.857	0.020
35	BeTi (MS)	63.000	2.857	0.020
30	MaSo	54.000	2.449	0.017
12	EnRa (Wett)	50.000	2.268	0.016
25	PeHe (GZ)	42.000	1.905	0.013
33	UlsI (EAE)	37.000	1.678	0.012
20	MiGr (CGN)	37.000	1.678	0.012
24	AnHe (TUI)	32.000	1.451	0.010
26	HeSt (DLH)	31.000	1.406	0.010
17	ChFo	21.000	0.952	0.007

**Tabla 16 Correo Electrónico 2007**

En el año 2007, vemos un crecimiento notable de la interactividad de los actores entre sí, especialmente se observa que los actores con acción decisiva, tales como los miembros del consejo de vigilancia, los gremios políticos (14, 21, 8 y 35) y los clientes más importantes (10, 28, 33, 24 y 26) usan con más intensidad la red.

El análisis de triadas en este año muestra un desarrollo claro:

Triad Census for dataset 2007	
	1
	-----
003	4068
012	0
102	2093
021D	0
021U	0
021C	0
111D	0
111U	0
030T	0
030C	0
201	848
120D	0
120U	0
120C	0
210	0
300	131

003 = A,B,C, the empty subgraph.  
 012 = A->B, C, subgraph with a single directed edge.  
 102 = A<->B, C, the subgraph with a mutual connection between two vertices.  
 021D = A<-B->C, the out-star.  
 021U = A->B<-C, the in-star.  
 021C = A->B->C, directed line.  
 111D = A<->B<-C.  
 111U = A<->B->C.  
 030T = A->B<-C, A->C.  
 030C = A<-B<-C, A->C.  
 201 = A<->B<->C.  
 120D = A<-B->C, A<->C.  
 120U = A->B<-C, A<->C.  
 120C = A->B->C, A<->C.  
 210 = A->B<->C, A<->C.  
 300 = A<->B<->C, A<->C, complete subgraph.

**Tabla 17 Censo de triadas año 2007**

La comunicación es en su mayor intensidad de forma bilateral (102, 201 y 300). Eso muestra una intensidad de comunicación y coordinación solo posible entre dos actores y menos en una comunicación de grupo como se observa en su desarrollo de las triadas para el año siguiente:

```

Triad Census for dataset 2008

          1
-----
003      4519
012      1568
102         72
021D     340
021U     317
021C     151
111D      24
111U      75
030T      46
030C        3
201         4
120D        1
120U         5
120C        11
210         4
300         0
    
```

**Tabla 18 Censo de tríadas año 2008**

El análisis de la centralidad en los años 2008 y 2009 coincide con el desarrollo de la empresa, consolidando los nexos logrados en los años anteriores. La red no tiene un crecimiento positivo y la intensidad es estable:

	2008	Degree	NrmDegree	Share
17	ChFo	411.000	14.864	0.160
22	GF	379.000	13.707	0.147
12	MiEg (HH)	215.000	7.776	0.084
10	CC-X	103.000	3.725	0.040
6	Bre (IHK)	97.000	3.508	0.038
1	AnFo	97.000	3.508	0.038
30	Pre (TU)	91.000	3.291	0.035
8	BuJa (OS)	78.000	2.821	0.030
33	ST-F&CO	70.000	2.532	0.027
4	BeNo (AB)	68.000	2.459	0.026
5	ChBl (DUS)	63.000	2.278	0.025
2	AxKa (Anw)	63.000	2.278	0.025
35	BeTi (MS)	61.000	2.206	0.024
7	BuJü (DLH)	52.000	1.881	0.020

34	ST-M	46.000	1.664	0.018
36	WoHe (MS)	44.000	1.591	0.017
21	OlGe (LR)	44.000	1.591	0.017
19	GeFu (STR)	43.000	1.555	0.017

**Tabla 19 Correo Electrónico 2008**

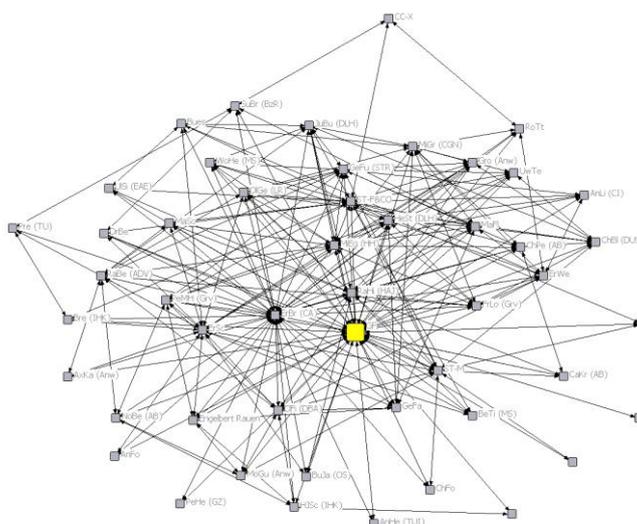
	2009	Degree	NrmDegree	Share
		-----	-----	-----
18	GF	391.000	23.798	0.196
14	ChFo	312.000	18.990	0.156
22	PeHe (GZ)	114.000	6.939	0.057
25	MoGu (Anw)	100.000	6.086	0.050
6	Bre (IHK)	90.000	5.478	0.045
11	MiEg (HH)	89.000	5.417	0.045
1	AnFo	72.000	4.382	0.036
9	CC-X	71.000	4.321	0.036
15	GeFu (STR)	69.000	4.200	0.034
26	PeMH (Grv)	56.000	3.408	0.028
23	RaHi (HAJ)	54.000	3.287	0.027
28	ST-F&CO	49.000	2.982	0.024
32		48.000	2.921	0.024
2	AxKa (Anw)	47.000	2.861	0.023
21	HJSc (IHK)	45.000	2.739	0.023
5	ChBl (DUS)	44.000	2.678	0.022
8	Bues	37.000	2.252	0.019
10	ChPe (AB)	36.000	2.191	0.018
17	OlGe (LR)	34.000	2.069	0.017

**Tabla 20 Correo Electrónico 2009**

En los últimos dos años el crecimiento económico de la empresa (Gráfica 43) no fue muy exitoso y en una forma paralela el crecimiento de la red y de las relaciones entre sus actores no creció sustancialmente. Al contrario de lo esperado y de lo mostrado en las estadísticas de la UE (UE-27)<sup>21</sup>, el crecimiento de la red no es significativo, más aun, en su complejidad se observa un decrecimiento

en el año 2009. ¿A qué se debe este fenómeno? Mi aclaración es que se ha llegado a un punto de saturación o de nivel estable en cuestión de finanzas. La compañía se encuentra a principios del 2009 en una situación segura. Y la crisis económica que se verá a finales del 2009 no es prevista. Un análisis de los efectos de la crisis económica a la red social y del posterior efecto de la erupción del volcán en Islandia a las actividades del aeropuerto, podrán ser estudiadas en un futuro.

Si analizamos la red completa, tomando todos los datos disponibles obtenemos un grafo parecido al del año 2007:



**Gráfica 40 Red social desde el año 2002 hasta el 2010**

La centralidad total muestra un papel significativo de pocos actores:

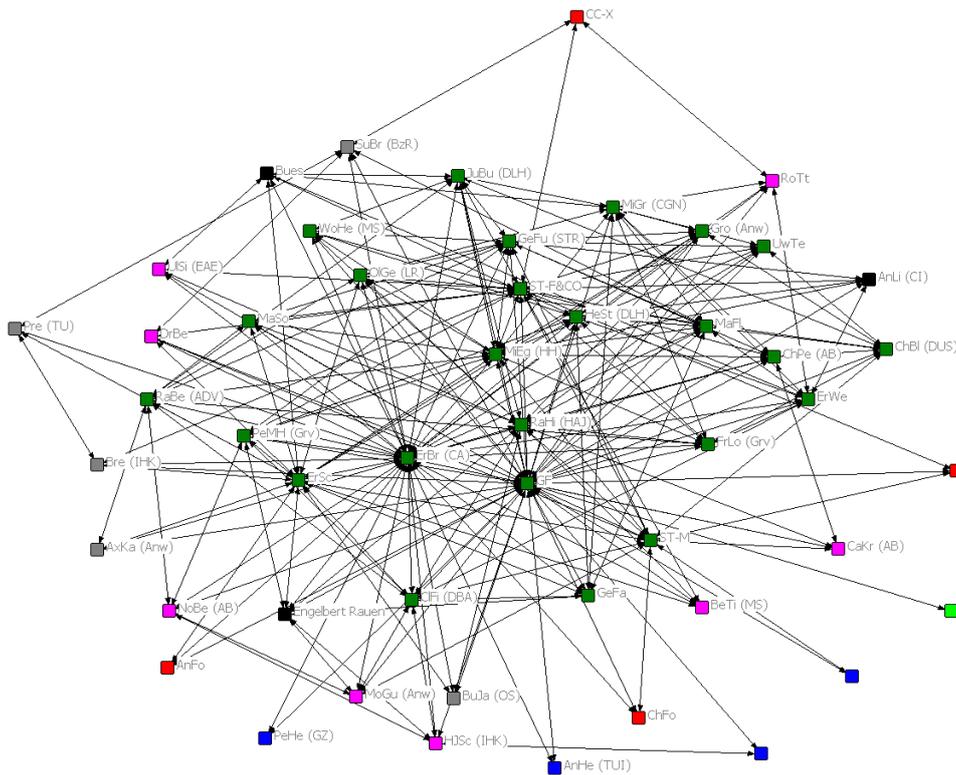
	2010	Degree	NrmDegree	Share
17	ErBr (CA)	2563.000	19.590	0.166
21	GF	2253.000	17.221	0.146
34	OlGe (LR)	671.000	5.129	0.043

30	MiEg (HH)	557.000	4.257	0.036
23	GeFu (STR)	527.000	4.028	0.034
41	ST-F&CO	508.000	3.883	0.033
1	AnFo	433.000	3.310	0.028
19	ErWe	390.000	2.981	0.025
7	BuJa (OS)	343.000	2.622	0.022
11	ChBl (DUS)	340.000	2.599	0.022
28	MaFl	338.000	2.584	0.022
5	BeTi (MS)	335.000	2.561	0.022
36	PeMH (Grv)	334.000	2.553	0.022
13	ChPe (AB)	331.000	2.530	0.021
8	Bues	313.000	2.392	0.020
29	MaSo	299.000	2.285	0.019
31	MiGr (CGN)	298.000	2.278	0.019
27	JuBu (DLH)	286.000	2.186	0.019

**Tabla 21 Correo Electrónico 2010**

pero un valor importante de actores intermedios para el mantenimiento de la red social a través de los años de la red misma.

Una vez analizada la red, veamos qué papel juega el desarrollo económico. Cuando la situación de la empresa fue crítica la intensidad y efectividad de la red se mostró. En el análisis no podemos medir factores tales como similitud/ disimilitud (similarity/dissimilarity) o jerarquía, ya que los datos obtenidos son unilaterales, es decir son el producto de observación de un solo actor. Sin embargo, podemos con ayuda de una herramienta analítica llamada "K-core" ver la asociación de los miembros de la red en una forma visual:



**Gráfica 41 Red social sobre todos los tiempos visualizada con K-cores**

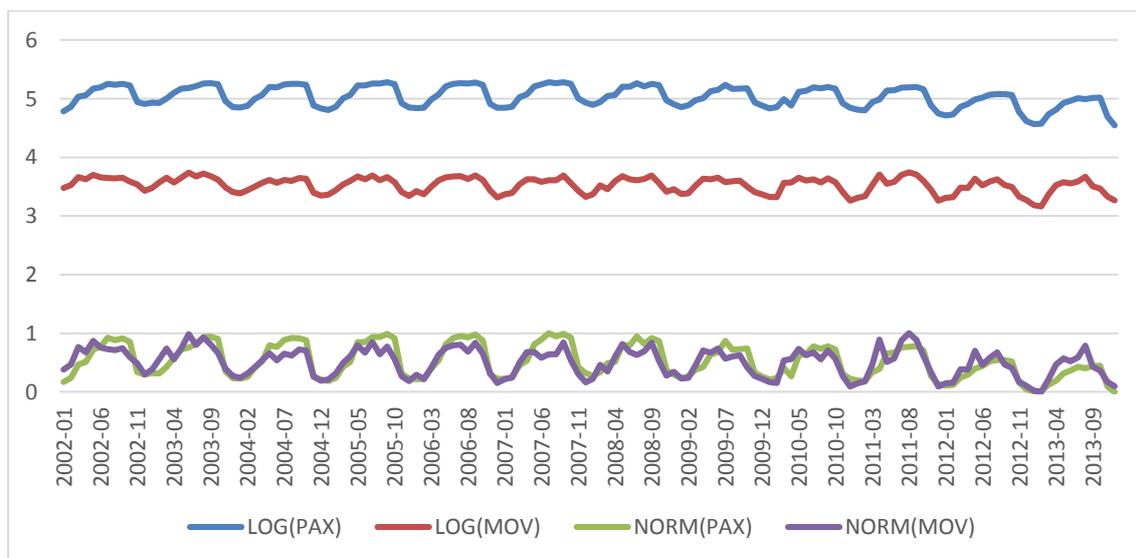
Este grafo fue creado mediante la herramienta de análisis K-core que está integrada en NetDraw (UCINET). Un K-core es un conjunto de nodos (actores) que están más estrechamente conectados entre sí que con otros actores en otros K-cores. Es decir, un K-core es una definición de un "grupo" o "sub-estructura" en un grafo.

Se observa la lógica distribución de la red, lo cual confirma las observaciones dadas en el transcurso de la investigación.

## 5.4. Análisis de los datos económicos de la empresa

Cada empresa tiene un factor especial para medirse a sí misma y compararse con la competencia. En la industria aeroportuaria los valores principales para compararse son la cantidad de pasajeros y la cantidad de movimientos aéreos que ha tenido el aeropuerto por semana, mes y año. Estos valores son rendidos en forma continua a diferentes organizaciones como la oficina de estadística, la oficina de coordinación aérea y las diferentes asociaciones aeronáuticas (ICAO, ACI, IATA, ADV etc.).

Los valores de estos movimientos los podemos visualizar tanto tomando el algoritmo del valor dado como normalizándolos según la fórmula (4).

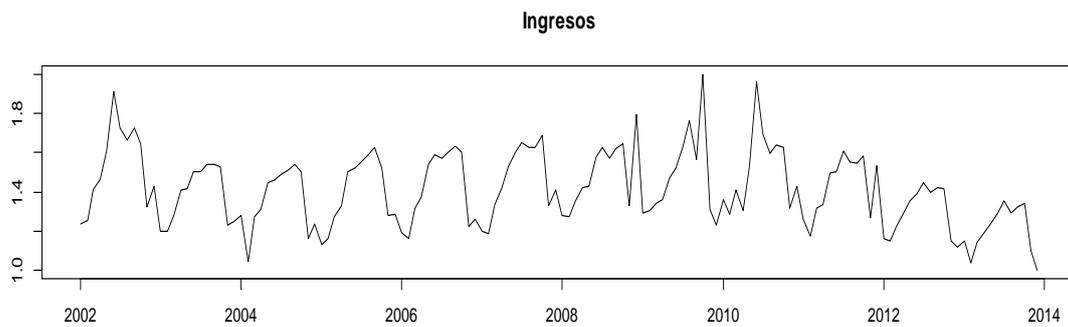


**Gráfica 42 Curva del desarrollo de pasajeros y movimientos**

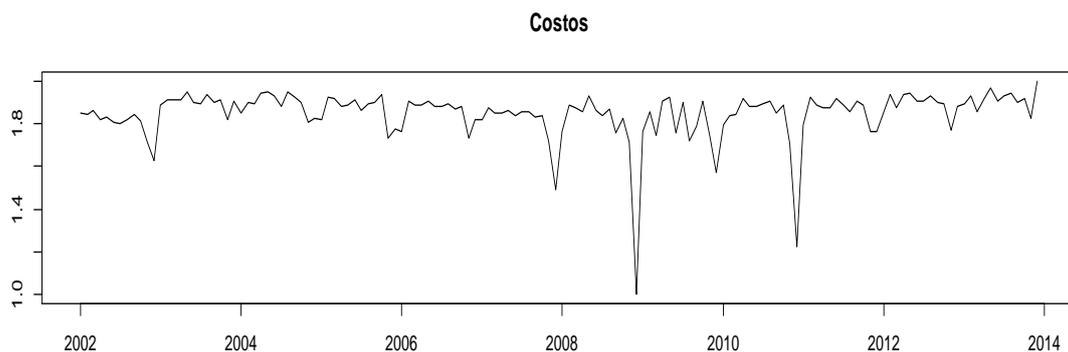
Se observa un efecto típico de evolución por temporadas. Ese mismo efecto se observa más adelante tanto en el balance general como en la curva de ingresos (Gráfica 45 y Gráfica 43).

### 5.4.1. Indicadores económicos / desarrollo en función del tiempo

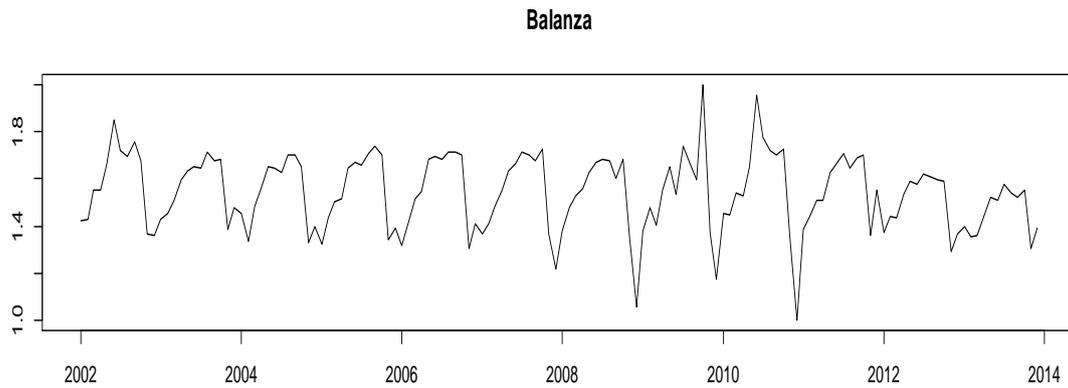
De la misma forma visualizamos los valores económicos obtenidos y normalizados:



**Gráfica 43 Curva de ingresos normalizada**



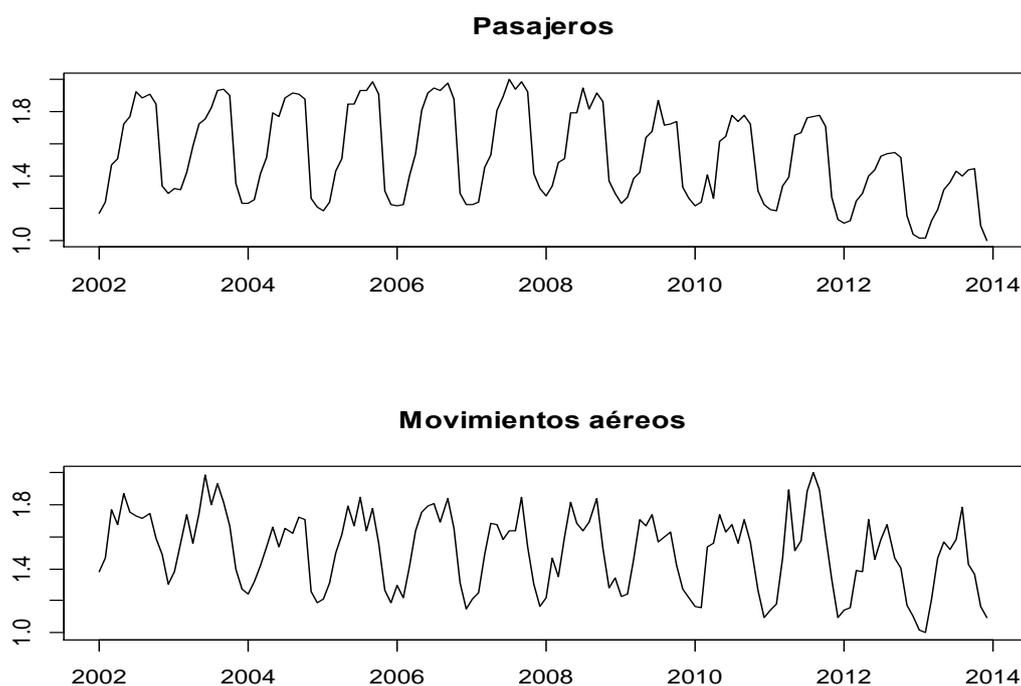
**Gráfica 44 Curva de costos normalizada**



**Gráfica 45 Curva del balance general normalizada**

Se observa una evolución por temporadas de los ingresos, pero un desarrollo casi lineal de los costos. Si vemos el desarrollo de los pasajeros y movimientos aéreos, se establece que el desarrollo económico está directamente relacionado con la cantidad de pasajeros que usan el aeropuerto. Es decir que los costos por su continuidad afectan muy poco un cambio en la curva del balance general y otros ingresos, tales como ingresos por arriendo de inmuebles, asesorías y aparcamientos.

Dicha evolución muestra unos efectos especiales en los años 2009 y 2011, que se observan en las gráficas tanto de ingresos, costos como balance general. Estos efectos también se observan en otra forma en la evolución de la cantidad de pasajeros a mediados del 2011 y a principios del 2010 y en la curva de movimientos aéreos, en una caída fuerte a mediados del 2011.

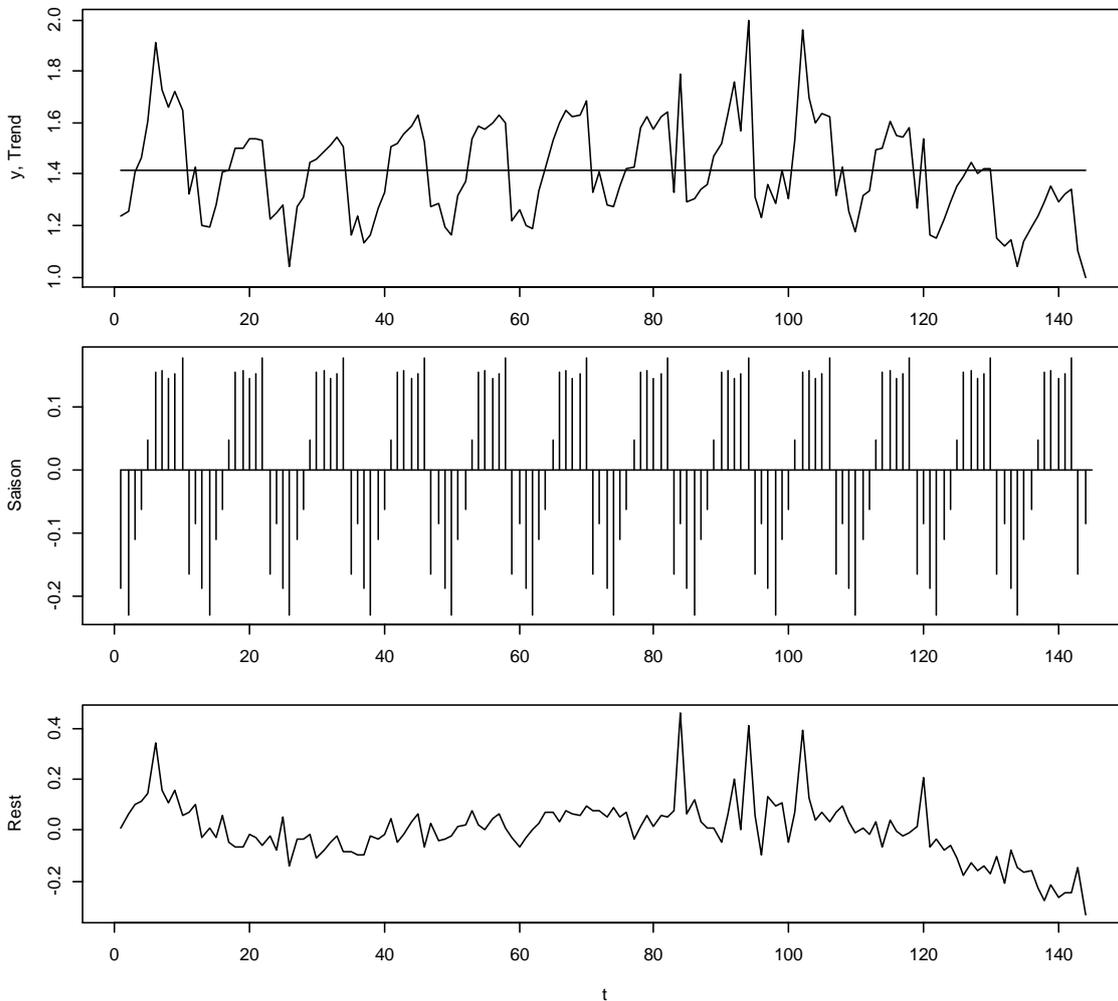


**Gráfica 46 Curva de cantidad de pasajeros y movimientos aeronáuticos**

Estos efectos de carácter empírico y determinista son explicables debido a los cambios radicales ocurridos en la industria, que afectaron fuertemente el desarrollo económico. Estos son en especial la decisión del mayor cliente de la empresa (una aerolínea) de consolidar algunas rutas en el año 2019 por lo que paulatinamente se ha ido retirando completamente de esta localización.

Los valores son influenciados por temporadas. Para poder analizar el cambio real de los valores utilizo la función de descomposición de una serie de tiempo en tendencia, temporada e irregularidad según el método de regresión lineal propuesta en

Schlittgen (2012) llamada en R `simpledecomp()`. El resultado aplicado al balance general es este:



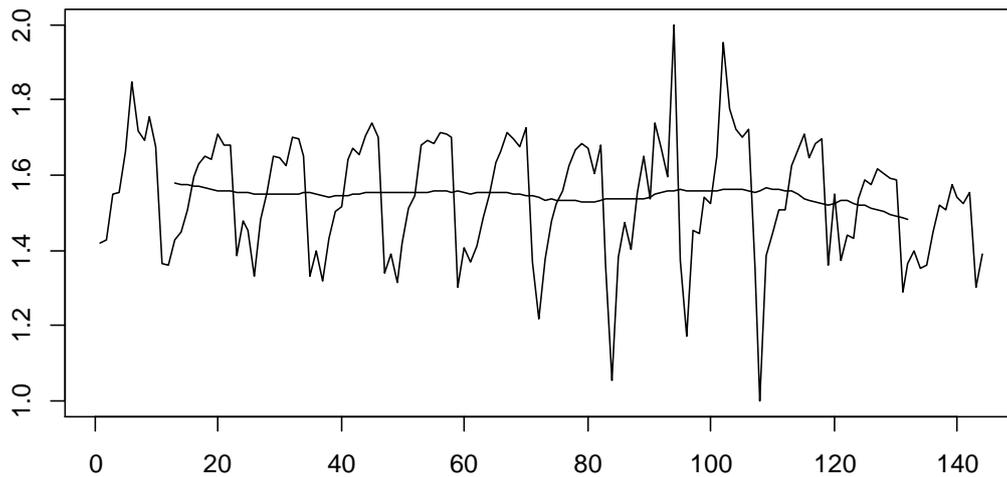
**Gráfica 47** Curva de tendencia, temporada y restos del balance general

Un análisis del balance general tomando el promedio de 24 meses muestra un comportamiento más estable pero con tendencia cada vez a un mayor decrecimiento. La fórmula para determinar el promedio es la propuesta por Schlittgen (2012):

$$\sum_{t=s+1}^{s+d} y_t / d \quad (11)$$

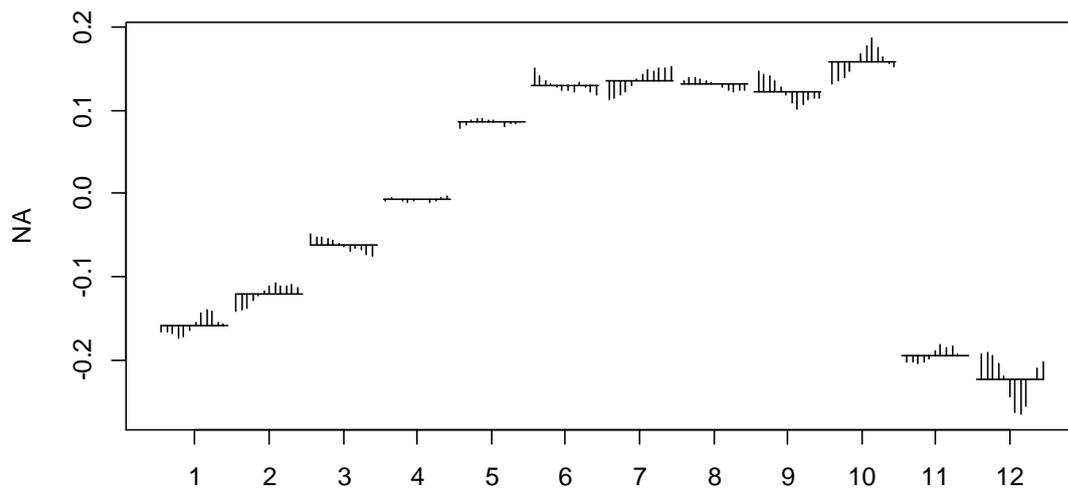
donde  $t$  es el período,  $d$  la cantidad de periodos a ser considerados (en este caso 24), e  $y$  el valor en el periodo determinado.

El resultado es aquí visualizado:



**Gráfica 48 Balance general y el promedio en periodos de 24**

Una observación del comportamiento por meses y año lo podemos visualizar en R de esta forma:

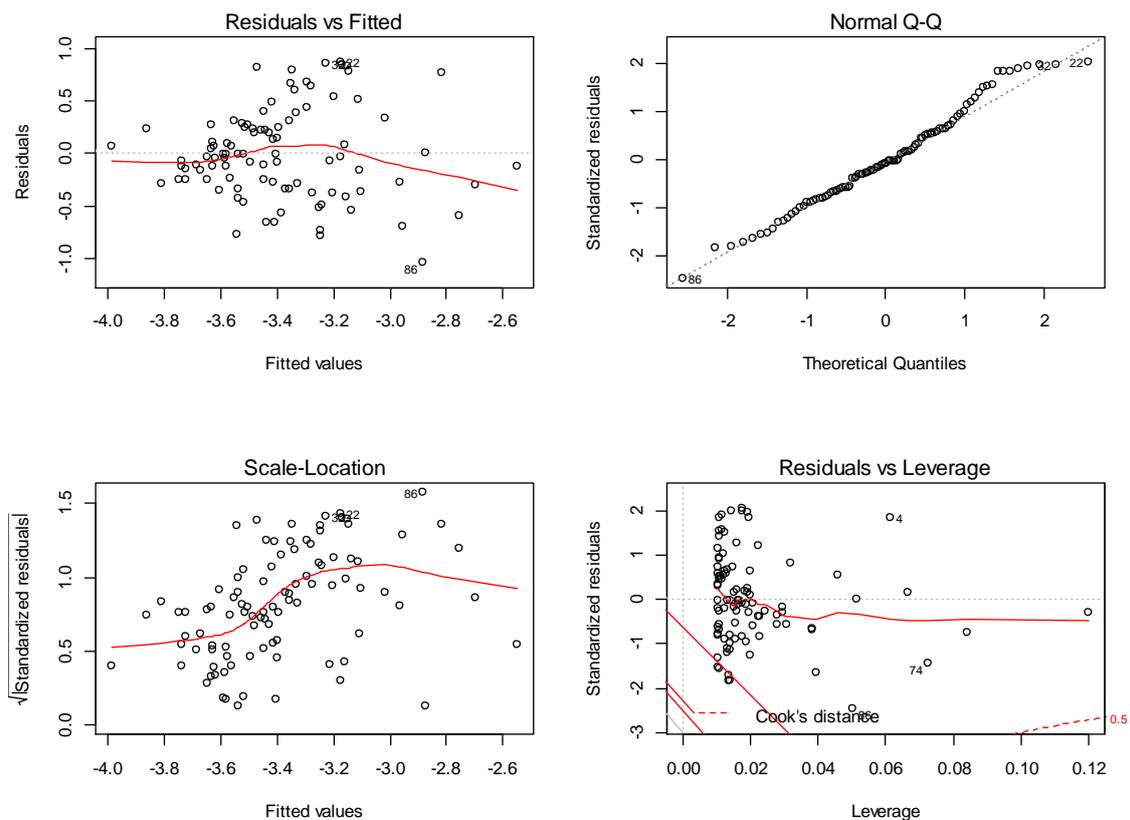


**Gráfica 49 Desarrollo por mes del balance general**

# CAPITULO VI. PRIMERA APROXIMACIÓN A UNA CORRELACIÓN

## 6.1. Introducción

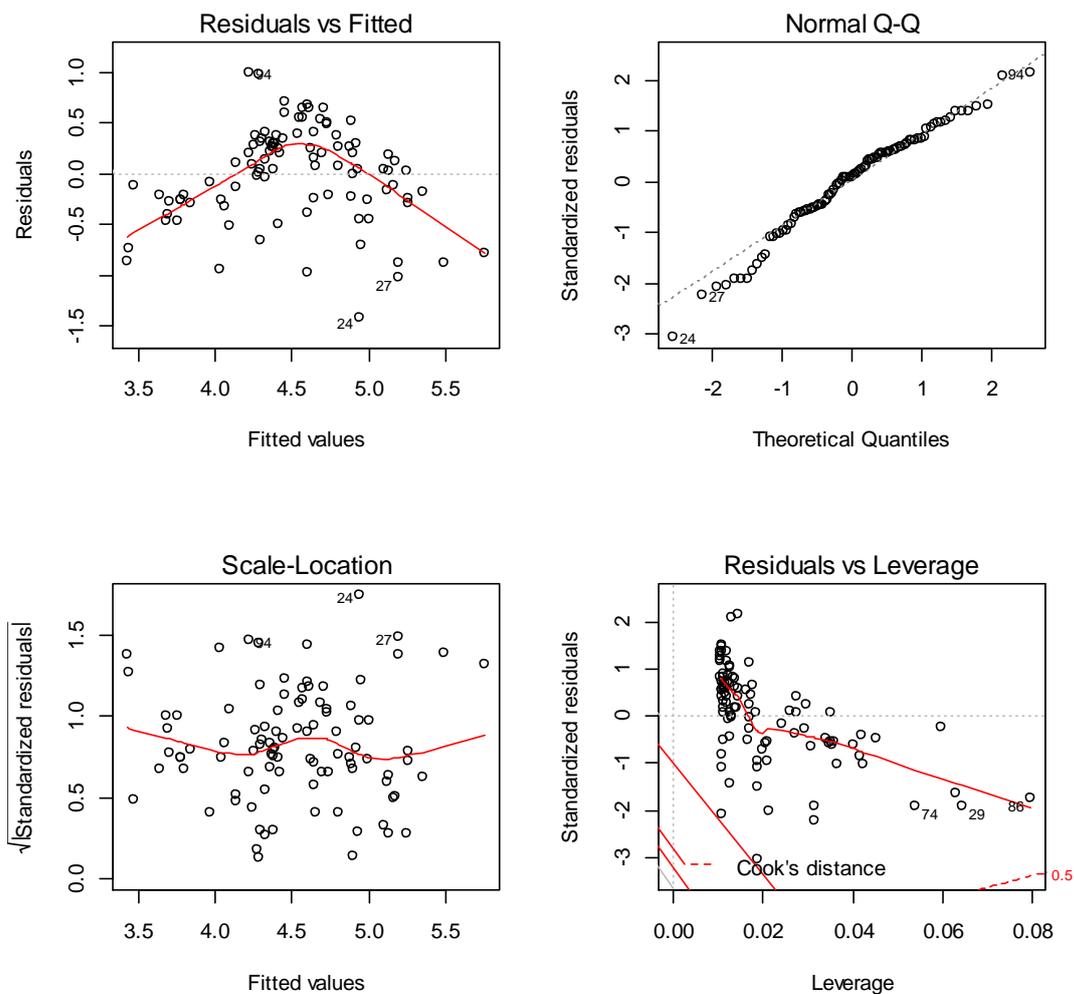
Una vez consolidados y controlados los datos obtenidos a través del análisis de la red social y los factores económicos, se busca una relación entre esos datos. Veamos cómo se comporta un análisis lineal entre BALANCE y DENSIDAD:



**Gráfica 50 análisis lineal entre BALANCE y DENSIDAD**

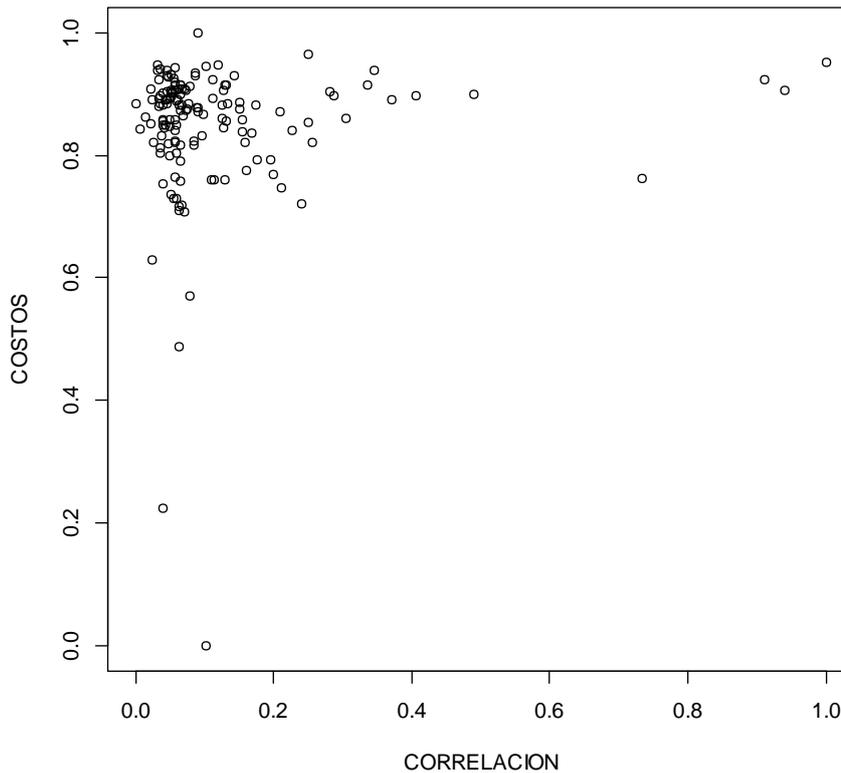
En particular en la observación de comparación entre los residuos y el apalancamiento, con el fin de identificar aquellos valores que no concuerdan con la predicción, se obtienen valores de poco apalancamiento sin importar el residuo. También la distancia de Cook muestra poca interdependencia o correlación de los dos valores.

Veamos ahora el mismo procedimiento tomamos como valor COMPONENTES en vez de DENSIDAD:



Gráfica 51 Análisis lineal entre Balance y Componentes

Una comparación directa entre COSTOS y CORRELACION, da una primera aproximación de que si hay una influencia entre estos valores:



**Gráfica 52 Correlación entre Costos y Correlación de la red social**

A primera vista se establece la teoría de que cuanto menos correlación en el transcurso del desarrollo de la red social los costos permanecen a nivel alto.

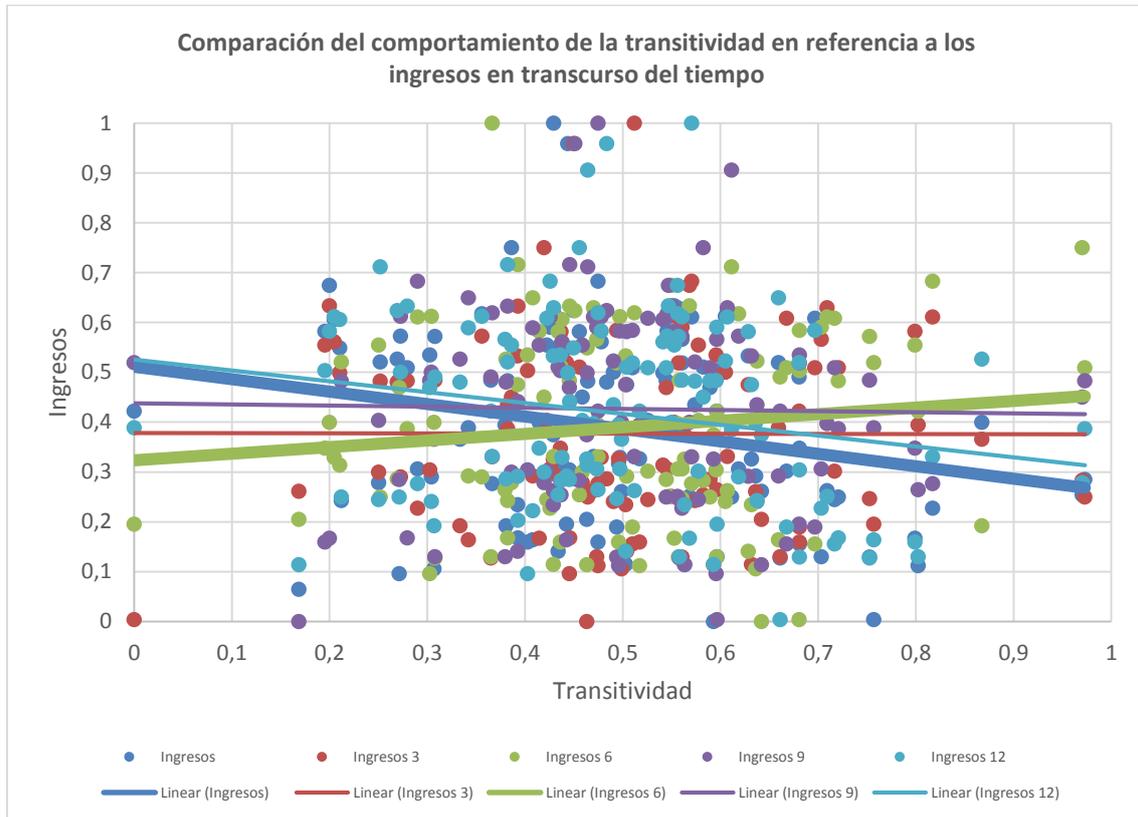
## **6.2. Propuesta para un modelo**

Un efecto económico que distorsiona el análisis son los valores anuales, que por lo general ocurren al final del año financiero, tales como depreciaciones especiales, impuestos anuales o

correcciones de la balanza por reversión de reservas o creación de nuevas reservas. En el caso de la empresa en cuestión, el año financiero acaba en el mes 12. Para evitar esos efectos extremadamente altos y significantes comparativamente con el resto de valores financieros, no tomo en cuenta ese último mes en el análisis detallado. Se podría tomar como efecto del mes 12 el promedio de meses anteriores, pero esto no cambiaría el resultado analítico, más sí el posible efecto posterior con referencia al desarrollo de la red social en el desarrollo económico. Como lo primero que se quiere identificar es si ese efecto realmente existe, en la investigación se propone ignorar el mes en el cual transcurren los efectos financieros anuales. Naturalmente hay valores tales como licencias, concesiones etc. que son anuales, pero estas por lo general ocurren en el transcurso del año y por lo tanto podemos ignorarlas. Por otro lado, si nos concentramos en los valores que normalmente son mensuales y significativos, como lo son los ingresos, los valores anuales en los costos no tendrían un efecto significativo.

Como se puede apreciar, la búsqueda de una influencia de algún factor medible de la red social estimada y el desarrollo económico no es clara y única. Mi propuesta se basa en concentrar la búsqueda en un valor congruente lógico y ese compararlo en su posible efecto a través del tiempo.

He tomado el efecto de transitividad de una red social como he descrito en la página 136 inmediatamente y en los meses siguientes en intervalos de 3 meses.

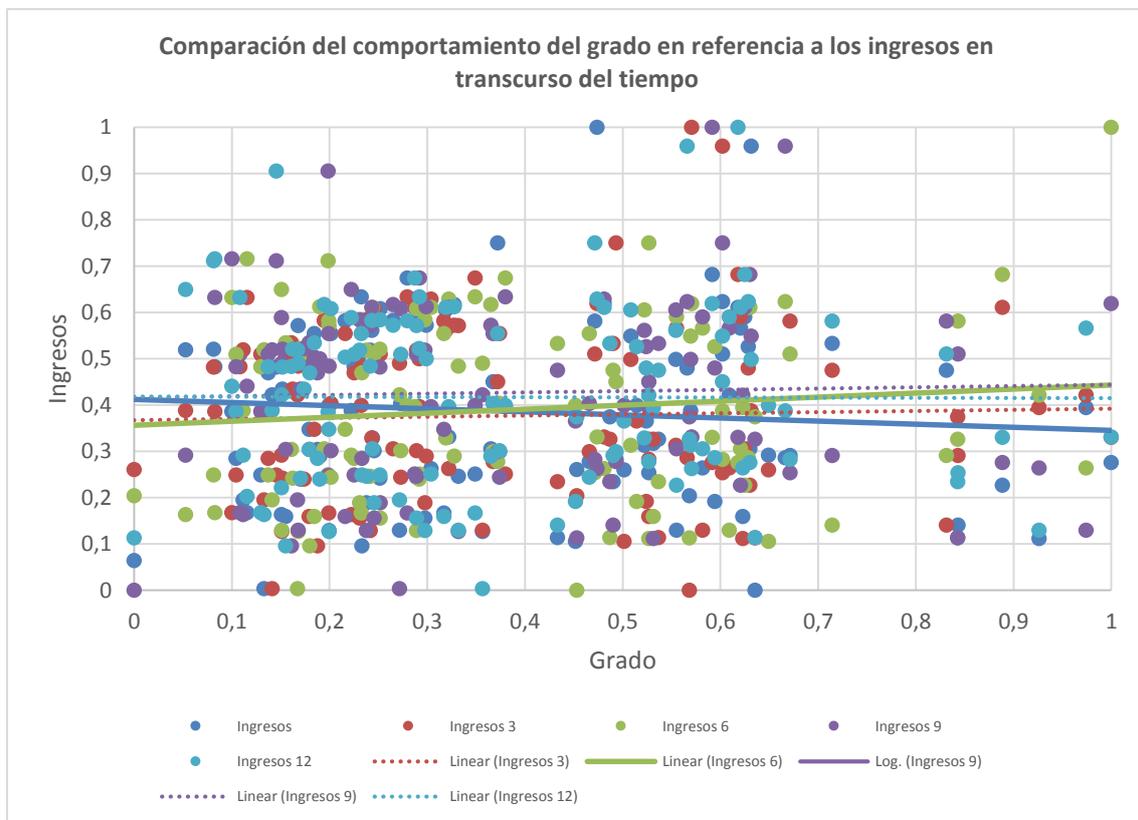


**Gráfica 53 Comparación del comportamiento de la transitividad en referencia a los ingresos en transcurso del tiempo**

Por primera vez se observa claramente un efecto directo del desarrollo de un valor de la red social, en este caso transitividad, en un valor económico. La transitividad crece a medida que el tiempo pasa en función de los ingresos, pero solo hasta 6 meses después. Podemos deducir de esta gráfica, que el efecto de la transitividad y por lo tanto de la actividad social tiene un efecto directo en el desarrollo económico,

pero solo hasta un máximo de 6 meses. Posteriormente el efecto se neutraliza regresando al valor inicial.

Para comprobar este valor tomo otro valor significativo de la red social como lo es el grado de la misma y los mismos valores de ingresos como fue mostrado anteriormente.



**Gráfica 54 Comparación del comportamiento del grado en referencia a los ingresos en transcurso del tiempo**

Podemos observar que a los nueve meses aún puede verse una influencia, sin embargo, el postulado anterior de que el mayor efecto se ve a los seis meses al observar la transitividad, es comprobado objetivamente al observar el desarrollo del grado de la red social.

### **6.3. Descripción del modelo**

El modelo se basa inicialmente en la fijación de valores únicos y generales de la red social y su desarrollo. La estimación de la red social como tal requiere de mucho cuidado y trabajo detallado. Eliminando aquellos valores que no son significativos - como lo son spam y boletines de noticias o propaganda - con el método descrito anteriormente, se consigue una red social de un tamaño analizable y con poder descriptivo suficiente para hacer un análisis. Posteriormente se tomarán los valores de grado y transitividad de la red social para continuar el análisis.

De esta forma obtenemos valores lineales descriptivos de la red social sin tener que estudiar detalladamente el contenido de esta, evitando así conflictos con las leyes de protección de datos.

El siguiente paso es la depuración de los valores económicos, que de por sí son lineales en factor del tiempo, pero por su característica no homogéneos. Hay que suprimir aquellos valores anuales que puedan falsificar un análisis mensual de la evolución de los valores a ser analizados.

Una vez estimados los valores y éstos distribuidos en aquellos que podamos observar detalladamente, podemos entrar a analizarlos.

En la compañía en cuestión fue posible disgregar estos valores en ingresos, costos influenciados y costos no influenciados directamente o inmediatamente como los son los costos de personal, impuestos y depreciaciones.

Una vez coleccionado los datos se observa que la magnitud de los mismos es muy diferente. Mientras que los valores económicos pueden variar en muchos millones según la empresa, los valores de la red social generalmente son mucho más pequeños:

	MINIMO	MÁXIMO
# Componentes	20	273
Grado	387	3858
Densidad	0,033100233	0,2106955
Transitividad	0,076625054	0,540909091
# Conexiones	3422	83810
Diámetro	11	335
# Aristas	130	871
# Vértices	59	290
Ingresos	2.102.075,39	6.618.948,14
Costos	-4.115.379,65	-2.582.464,26
Balanza	-1.131.196,91	3.668.862,33
Costo de materiales	619.522,75	2.219.299,96
Gastos de personal	1.251.419,58	2.310.364,93
Otros gastos	-192.124,32	1.662.841,79

**Tabla 22 Mínimos y máximos de valores a ser analizados**

Para poder usar estos valores recurre al método de normalización de los mismos.

#### **6.4. Factores necesarios para una normalización**

La normalización de valores en este contexto, significa buscar un valor representativo en un marco diseñado para un valor específico.

$$y_i = \frac{x_i - \min(x_1, \dots, x_z)}{\max(x_1, \dots, x_z) - \min(x_1, \dots, x_z)} \quad (12)$$

Con esta fórmula obtenemos una distribución representativa de todos los valores en un rango de 0 a 1. En el caso de que en el marco de operaciones se tenga que dividir por el valor normalizado, como se ha explicado anteriormente (y se ha aplicado en algunas partes de la investigación), esta fórmula se puede ampliar sin cambiar su representación por la siguiente:

$$y_i = \frac{x_i - \min(x_1, \dots, x_z)}{\max(x_1, \dots, x_z) - \min(x_1, \dots, x_z)} + 1 \quad (13)$$

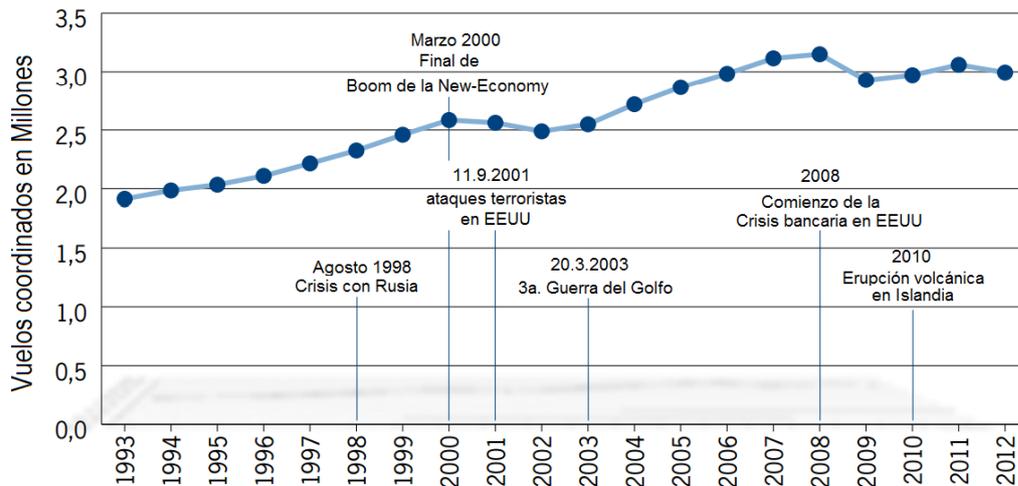
Siendo el mayor valor 2 y el menor 1, evitando de esta forma una división por 0 no definida.

#### **6.5. Influencias no posibles de cuantificar**

En cuanto a los efectos no posibles de cuantificar y que de manera casi imposible de predecir afectan a este tipo de estudios, se encuentra los hechos de carácter mundial y ambiental.

Si observamos la evolución del número de vuelos desde el año 1993, vemos que los efectos a nivel mundial han afectado el desarrollo de la industria en forma temporal.

### Desarrollo del tráfico aéreo en los últimos 20 años



Fuente: DFS

### Gráfica 55 Desarrollo del tráfico aéreo en Alemania desde el año 1993

Es por ello que la crisis de las empresas vinculadas a internet en el año 2000, el ataque terrorista del 11 de septiembre 2001, la crisis bancaria que comenzó en el 2008 y la erupción del volcán Eyjafjallajökull en 2010 han influido notoriamente en el desarrollo económico de la empresa. Igualmente, en sentido positivo para el aeropuerto fue el incendio catastrófico en el aeropuerto de Düsseldorf en 1996, que obligó a las empresas aéreas a buscar alternativas temporales.

Otro factor que ha influenciado fuertemente el desarrollo económico de la empresa es la increíble fluctuación en el mercado de empresas aéreas. En el año 2002 Lufthansa adquiere la empresa EuroWings que hasta ese momento era una gran competencia, acabando de ese modo un crecimiento debido a la competencia. También el desarrollo económico de la empresa se ha visto afectado por la quiebra, en los años posteriores de empresas como OLT, CityAir, SpanAir, SwissAir, Jetisfaction, CirrusAirlines y HamburgAirlines, que también tenían un servicio permanente en el aeropuerto objeto de investigación. A pesar de estos casos, sin embargo, se observa que la influencia de la red social es de enorme importancia y muy probablemente gracias a la existencia de esa red la empresa ha logrado salir adelante a pesar de todos estos imprevisibles desarrollos.

## **CAPITULO VII. VERIFICACIÓN CON AYUDA DE REDES NEURONALES**

### **7.1. Redes Neuronales**

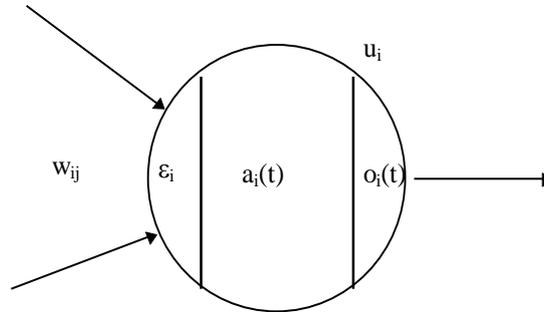
El fundamento para las redes neuronales a finales de los ochenta fue puesto a principios de los años 40 a finales del siglo pasado. Viendo las computadoras Enigma y Purple, con que los aliados en la segunda guerra mundial descifraron el sofisticado código del enemigo, esperaban muchos ingenieros, que esos aparatos se desarrollarían tan rápidamente, que al final pensarían en una forma lógica y resolverían problemas al igual que entenderían el idioma humano y entenderían mensajes visuales (Hassing, 1989), (Johnson & Brown, 1988). Los descubrimientos en neurofisiología han mostrado que las neuronas son mucho más complejas de lo que los pioneros McCulloch y Pitt (1943) creían. Por ejemplo, las neuronas naturales no cambian de un estado pasivo a activo solamente, sino que tienen muchos estados intermedios.

Al concepto de la red neuronal, se llega cuando se resumen los procesos neuronales son descritos en un modelo abstracto altamente simplificado. Este modelo es descrito por las siguientes reglas (Hoffmann, 1993):

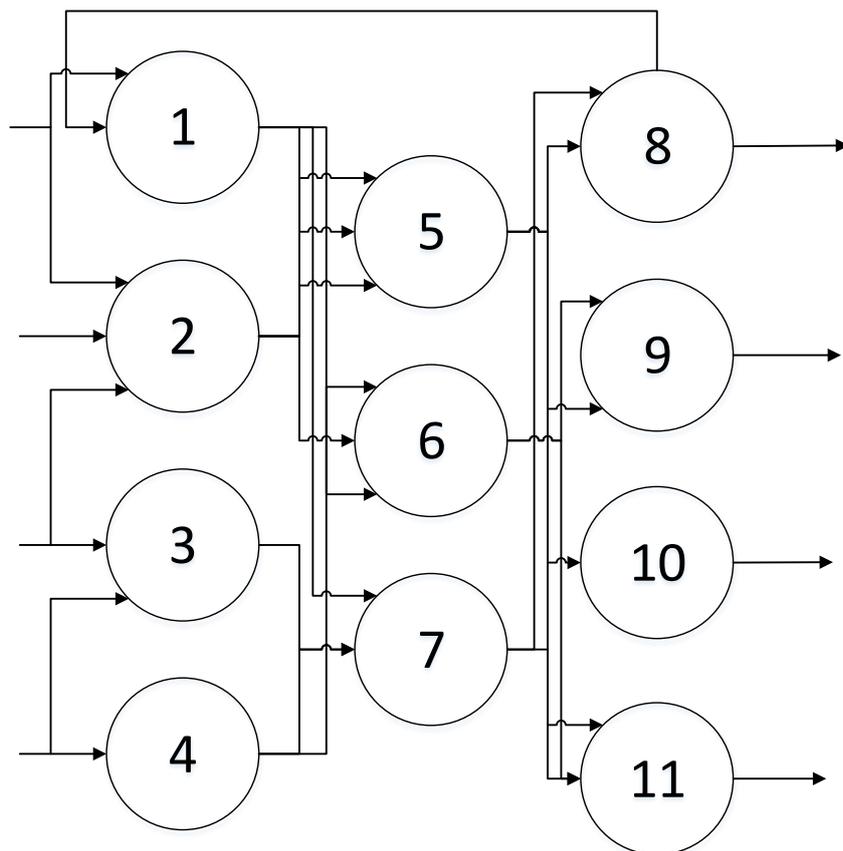
1. Construcción de una neurona: Una neurona tiene muchas entradas, llamadas conexiones sinápticas y una salida llamada axón.
2. Estado de una neurona: Una neurona puede tener dos estados: pasivo y activo
3. Conexiones de las neuronas entre sí: La salida de una neurona es la entrada a otra neurona
4. Entradas al sistema neuronal: Unas neuronas son activadas por el medio ambiente
5. Salidas del sistema neuronal: Algunas salidas de las neuronas accionan sobre el medio ambiente
6. Estado de excitación: Una neurona es activada cuando sus entradas en estado activado son suficientes para activarla
7. Independencia de la neurona: El estado de la neurona es solo dependiente de las entradas. Es decir, cada neurona actúa independiente de las otras neuronas.

La base de una red neuronal es la neurona (Gráfica 56). Una red neuronal es compuesta en principio por una cantidad no determinada de neuronas, que se comunican entre sí. Las conexiones entre las neuronas pueden ser convergente o divergente o una combinación de ambas. Convergente significa en la teoría de redes

neuronales, que la salida de una neurona va a otra, y divergente que la conexión puede ir a varias.



**Gráfica 56 Neurona u:  $w$  ponderación de las conexiones,  $\epsilon$  entrada efectiva,  $a(t)$  función de activación y  $o(t)$  señal de salida**



**Gráfica 57 Una red neuronal con 10 neuronas, 4 entradas y 4 salidas**

## 7.2. Elección de una red adecuada

Las redes neuronales son clasificadas en redes de aprendizaje supervisado y redes de aprendizaje no supervisado (Hoffmann, 1993).

Red	construcción	modo de trabajo	tipo de neurona	regla de aprendizaje	literatura
Rede de aprendizaje supervisado					
Hopfield	1 nivel, realimentado	auto asociativo	Hopfield	Hopfield	Brause 1991
BAM	2 niveles, realimentado	hetero asociativo	Hopfield	Hopfield	Brause 1991
Boltzmann	3 niveles, realimentado	hetero asociativo	Boltzmann	Boltzmann	Ackley 1985
Contra-corriente	2 niveles, acoplada hacia delante	Hetero y auto asociado	Reproducción	Aprendizaje	Hertz 1991
Muestra asociativa	1 nivel, acoplada hacia delante	hetero asociativo	indiferente	Hebb, Delta	Brause 1991
Willshaw	1 nivel, acoplada hacia delante	hetero asociativo	McCulloch-Pitts	especial	Ritter 1991
Perceptron	2 niveles, acoplada hacia delante	hetero asociativo	McCulloch-Pitts	perceptron	Block 1962, Ritter 1991
Adaline	Muestra asociativa con Bias	hetero asociativo	Adaline	Delta	Brause 1991, Kratzer 1990
Madaline	2 niveles expansión de Adaline	hetero asociativo	especial	especial	Kratzer 1990

CAPITULO VII. VERIFICACIÓN CON AYUDA DE REDES  
NEURONALES

Auto asociativa	1 nivel, realimentado	auto asociativo	indiferente	Hebb, Delta	Hoffmann 1992
BSB	como Auto asociativa	auto asociativo	BSB	Delta	Hoffmann 1992
DMA	como Auto asociativa	auto asociativo	DMA	Delta	Hoffmann 1992
Backpropagation	varios niveles, acoplado hacia adelante	hetero asociativo	salida de función diferenciable	Backpropagation	Ruiz 1991, Kratzte 1990
Rede de aprendizaje no supervisado					
memorias asociativas					Storman 1992
Mapa auto organizado					Kohonen 2007
ART (RNA)					Carpenter 1987

**Tabla 23 tipos de redes neuronales acuerdo a tipo de aprendizaje**

Las redes pueden ser diseñadas de tal forma que los niveles interiores no son vistos y no tienen ninguna interactividad con el exterior. Entre las reglas de aprendizaje más comunes está la hipótesis de Hebb que fue formulada 1949: La propiedad sináptica (reforzante o inhibitoria) cambia proporcionalmente al producto del anterior y posterior actividad sináptica (Hebb, 1949).

$$\Delta w_{ij} = \sigma \cdot e_i \cdot a_j \quad (14)$$

Es  $w_{ij}$  la ponderación de la red con entrada  $e_i$  y salida  $a_j$  entonces se cambia  $w_{ij}$  si hay una entrada, siendo  $\sigma$  una constante determinada. Normalmente la regla de Hebb es usada en una forma cambiada y es llamada la regla Delta

$$\Delta w_{ij} = \sigma \cdot e_i \cdot \Delta a_j \quad (15)$$

De esta forma la ponderación de la neurona no se cambia si la salida coincide con la salida esperada. Esta regla es también llamada regla de Widrow-Hoff.

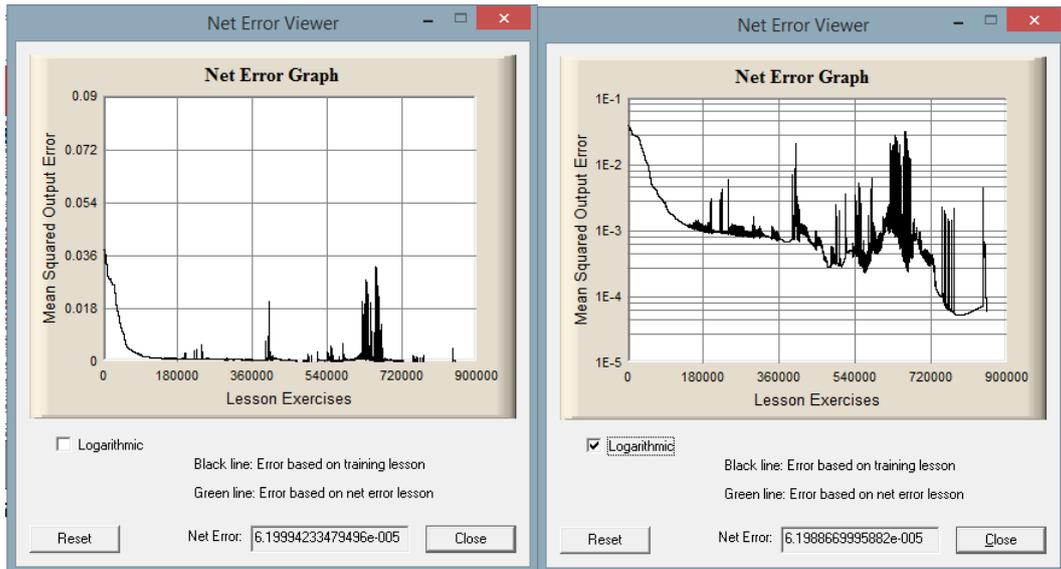
Con ayuda de MemBrain (Jetter, 2004) y después de hacer múltiples pruebas con los datos adquiridos, concebí una red neuronal que es capaz de predecir los datos adquiridos.

La red diseñada tiene las siguientes características:

Característica	
# de neuronas de entrada	2
# de neuronas de salida	1
# de niveles de neuronas ocultos	2
# de neuronas ocultas	34
# de enlaces	340
Función de activación para las neuronas de entrada	IDENT. 0 TO 1
Función de activación para la neurona de salida	LOGISTIC
Función de activación para las neuronas de ocultas	LOGISTIC
Regla de aprendizaje	BP (Full loopback)
Cantidad de lecciones para aprendizaje	138

**Tabla 24 Características de la red neuronal**

La red es una red muestra asociativa (Rey & Wender, 2010) con neuronas ocultas y entrenamiento supervisado. Después de un entrenamiento con el algoritmo Backpropagation de más de 900.000 repeticiones, se alcanzó un error de red menor de  $6,1e-5$



**Gráfica 58 Error de red**

El algoritmo llamado Backpropagation aplica un método derivado de la regla de aprendizaje Delta o Widrow-Hoff (Kohle, 1990), que a su vez se basa en la regla de aprendizaje de Hebb formulada en 1949 (Kohle, 1990): Si la neurona  $a$  y la neurona  $b$  están activadas fuerte y repetidamente, entonces se eleva el nivel de su conexión.

El algoritmo de aprendizaje Backpropagation (error regresivo) se puede formular de la siguiente manera:

1. Determine la ponderación de todas las conexiones en forma aleatoria
2. Entre una muestra de entrada y salida y calcule las asignaciones del nivel no visible
3. Para el valor de entrada y valor esperado corrija de acuerdo a la fórmula 15

4. Repita el paso 2

El siguiente paso trata sobre lo que pasa con las neuronas escondidas: Las señales de error de neuronas ocultas se generan por la diferencia entre lo deseado y el valor real de salida; ese error se reparte proporcionalmente entre las neuronas ocultas de tal manera que la señal de error de una neurona es mayor, cuanto más participan en la salida. Esta participación se mide por la intensidad del peso o ponderación de la conexión.

El cálculo de la activación de una neurona está dado por la siguiente función (Jetter, 2004):

$$o(t) = o(t) \cdot S + f(\text{Sum}(w_{ij}) \cdot T) \quad (16)$$

Donde S es el valor adaptado y T el nivel de sensibilidad de la neurona. Jetter aplica en su programa este método para activar una neurona:

#	Activación	Nivel de activación	# de pasas desde última activación	$o(t)$
1	$a(t) \leq \text{umbral inferior}$	no relevante	no relevante	0
2	no relevante	no relevante	# < Tiempo de recuperación de salida	0
3	$a(t) \geq \text{Umbral superior}$	1	# $\geq$ Tiempo de recuperación de salida	1
4	$a(t) \geq \text{Umbral superior}$	Activación	# $\geq$ Tiempo de recuperación de salida	$a(t)$
5	$\text{Umbral inferior} < a(t) < \text{Umbral superior}$	1	# $\geq$ Tiempo de recuperación de salida	0 o 1
6	$\text{Umbral inferior} < a(t) < \text{Umbral superior}$	Activación	# $\geq$ Tiempo de recuperación de salida	0 o $a(t)$

**Tabla 25 método de activación de una neurona**

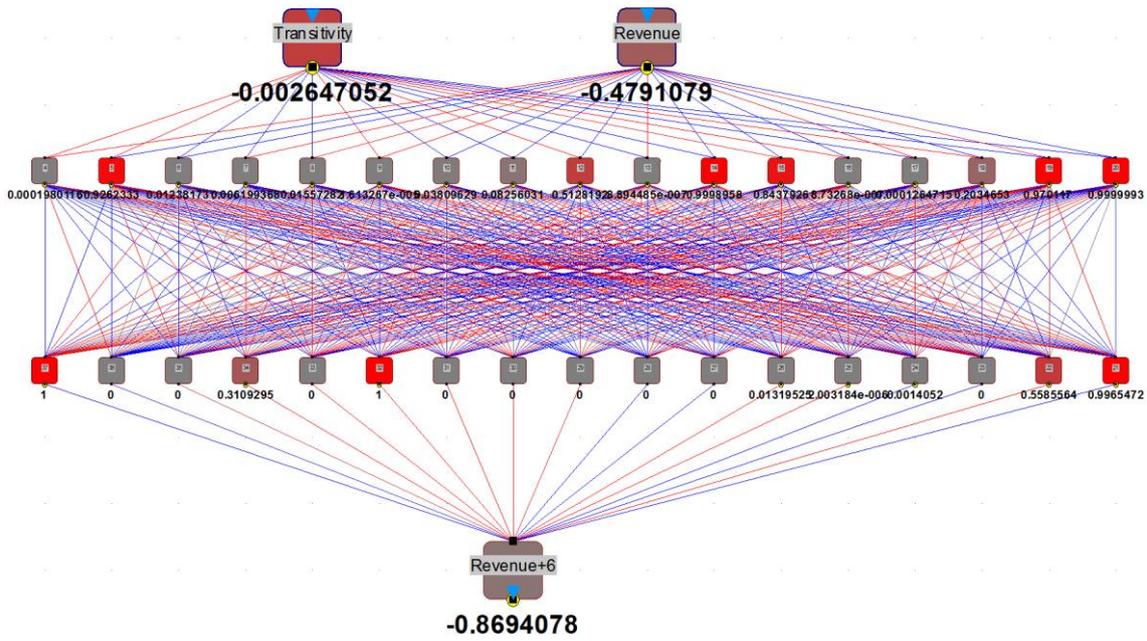
En este caso debido a que escogí para todas las neuronas un valor de umbral superior e inferior idénticos (0) no se aplica el sistema de probabilidad que usa Jetter para diferenciar el valor de salida en los casos 5 y 6; es decir estos casos no son tenidos en cuenta.

### **7.3. La red neuronal**

Las entradas a la red son los valores de transitividad de la red social y los ingresos del mismo mes. Como salida y al efecto de comprobar lo establecido en 5.2. se espera el valor de ingresos 6 meses después. Como se observó con ayuda de Excel la influencia de los valores de la red social y de los valores económicos actuales es máxima a los 6 meses del mismo. Es decir que los efectos son claros.

Con esta red se predicen los valores futuros confirmando lo ya hallado anteriormente. La red se ve en la Gráfica 59.

El valor de la ponderación de las conexiones puede variar si el entrenamiento se inicializa nuevamente, debido a que la inicialización de la red es aleatoria inicialmente. Es decir, si el experimento se repite partiendo de cero, los valores de las conexiones pueden variar ligeramente, no obstante, el resultado final es siempre el mismo. Con esta configuración siempre es posible encontrar una red que prediga los posibles valores esperados.



**Gráfica 59 Red neuronal**

### 7.3.1. Neuronas ocultas

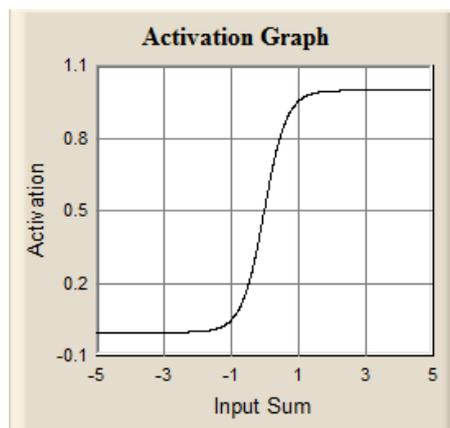
Las neuronas ocultas o intermedias tienen todas las mismas funciones de activación  $a(t)$ . En total se logró el mejor resultado con dos niveles de neuronas ocultas, cada nivel con 17 neuronas y todas interconectadas entre sí mismas.

Nivel	Neurona	$a(t)$	Valor de activación	Umbral	$o(t)$	Norm.	Norm. baja	Norm. alta
H1	4	LOG	0,85482704	-0,30545178	$a(t)$	0	-1	1
H1	5	LOG	0,54164114	-0,21521553	$a(t)$	0	-1	1
H1	6	LOG	0,24472563	0,11138736	$a(t)$	0	-1	1
H1	7	LOG	0,71524616	-0,3245131	$a(t)$	0	-1	1
H1	8	LOG	0,42981064	0,09173508	$a(t)$	0	-1	1
H1	9	LOG	0,73315537	-0,24872509	$a(t)$	0	-1	1
H1	10	LOG	0,60299521	-0,24848777	$a(t)$	0	-1	1
H1	11	LOG	0,6731491	-0,17168324	$a(t)$	0	-1	1
H1	12	LOG	0,35903373	0,105764	$a(t)$	0	-1	1

## CAPITULO VII. VERIFICACIÓN CON AYUDA DE REDES NEURONALES

H1	13	LOG	0,6333993	-0,11702396	a(t)	0	-1	1
H1	14	LOG	0,46450603	0,05535036	a(t)	0	-1	1
H1	15	LOG	0,36148327	0,30647981	a(t)	0	-1	1
H1	16	LOG	0,65709695	-0,47571428	a(t)	0	-1	1
H1	17	LOG	0,82165569	-0,4167213	a(t)	0	-1	1
H1	18	LOG	0,21913692	0,41684404	a(t)	0	-1	1
H1	19	LOG	0,69044687	-0,15139493	a(t)	0	-1	1
H1	20	LOG	0,4221684	0,34481257	a(t)	0	-1	1
H2	21	LOG	0,11407647	0,29987441	a(t)	0	-1	1
H2	22	LOG	0,48860867	0,23620359	a(t)	0	-1	1
H2	23	LOG	0,92705445	-0,4246396	a(t)	0	-1	1
H2	24	LOG	0,3065234	0,12207291	a(t)	0	-1	1
H2	25	LOG	0,87501551	0,07122884	a(t)	0	-1	1
H2	26	LOG	0,47784184	-0,26737692	a(t)	0	-1	1
H2	27	LOG	0,33185559	0,25713675	a(t)	0	-1	1
H2	28	LOG	0,10918774	0,32902924	a(t)	0	-1	1
H2	29	LOG	0,29544162	-0,09077841	a(t)	0	-1	1
H2	30	LOG	0,13193027	0,31532105	a(t)	0	-1	1
H2	31	LOG	0,07488982	0,45516816	a(t)	0	-1	1
H2	32	LOG	0,28671764	-0,16112541	a(t)	0	-1	1
H2	33	LOG	0,99267727	-0,0547235	a(t)	0	-1	1
H2	34	LOG	0,0398926	0,31008556	a(t)	0	-1	1
H2	35	LOG	0,41727186	-0,04357595	a(t)	0	-1	1
H2	36	LOG	0,23288582	0,24017026	a(t)	0	-1	1

**Tabla 26 Neuronas ocultas**



**Gráfica 60 Grafica de activación función logística**

### 7.3.2. Neuronas de entrada

Las dos neuronas de entrada, "Transitividad" y "Entradas", tienen las siguientes características:

Nivel	Entrada	
Neurona	Transitividad	Entradas
a(t)	IDENT01	IDENT01
Valor de activación	0,51327813	0,28198645
Umbral	0	0
Ponderación	1	1
o(t)	a(t)	a(t)
Normalización	1	1
Normalización baja	-1,06	-1,06
Normalización alta	1	1

**Tabla 27 Neuronas de entrada**

### 7.3.3. Neurona de salida

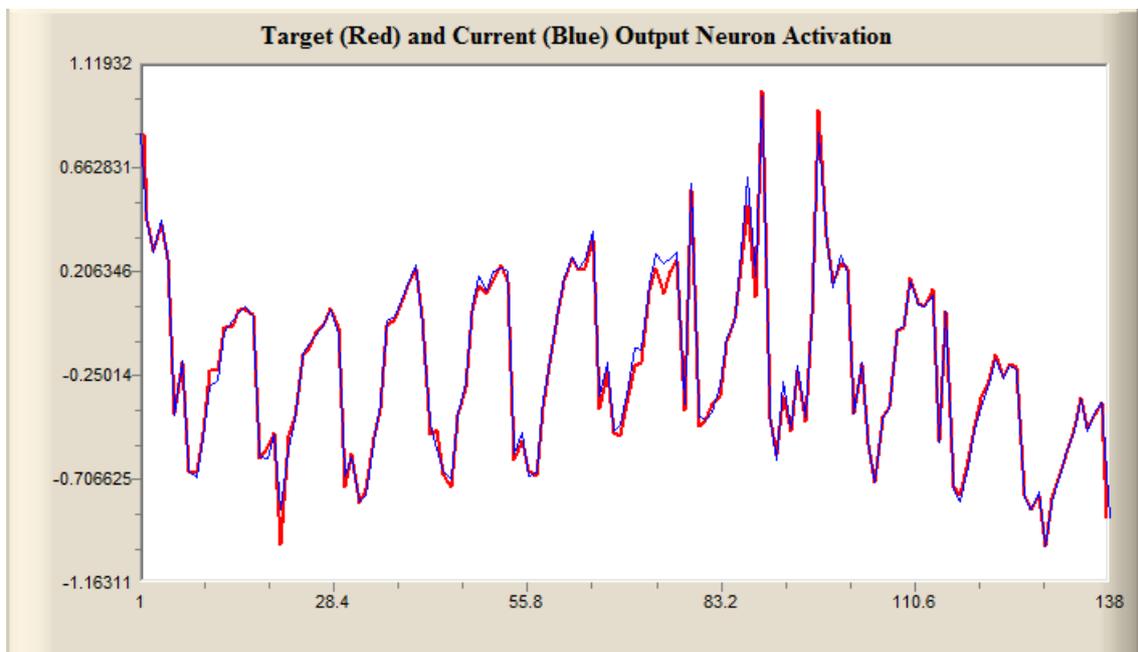
La neurona de salida, "Entradas 6 meses después", tiene la siguiente característica:

Nivel	Salida
Neurona	Entradas 6 meses después
a(t)	Logística
Valor de activación	0,244083679
Umbral	-0,28837272
Ponderación	1
o(t)	a(t)
Normalización	1
Normalización baja	-1,06
Normalización alta	1,06

**Tabla 28 Neurona de salida**

### 7.3.4. Resultado de la red

La red encontrada muestra un error aceptable tanto en la observación visual como tanto en el resultado matemático de computación del error de la muestra. Como se ve en la Gráfica 58 Error de red, el error calculado es  $6,1e-5$ . Y si observamos visualmente el resultado de la misma, vemos una coordinación muy grande entre el resultado esperado y el resultado obtenido:



**Gráfica 61 Pattern Error**

Es decir, con esta red es posible teniendo los datos actuales de Transividad e Ingresos predecir los posibles Ingresos en los próximos 6 meses. Confirmando de esta manera en forma clara y evidente la influencia de la red social a los resultados económicos de una empresa PYME.

## **CAPITULO VIII. CONCLUSIONES Y RECOMENDACIONES**

### **8.1. Conclusiones generales**

La meta de este trabajo, como formulado en capítulo uno, es mostrar instrumentos en el análisis de redes sociales con el propósito de utilizar su desarrollo como un indicador económico en empresas PMYE. En ningún caso digo que los métodos mostrados son aplicables para todo tipo de análisis científico y que este sea la estrategia correcta. El trabajo instaura preguntas importantes sobre la estrategia, los déficits estructurales y la estructura social y describe alternativas que se pueden escoger y evaluar. Con estos conceptos y perspectivas del desarrollo de la red social a través del tiempo y valores macroeconómicos, se consigue exitosamente descubrir una relación puramente matemática entre dos conceptos, que en su origen investigativo buscan por separado metas, desde el punto de vista procedural y estratégico absolutamente diferentes.

Por último, quiero hacer énfasis en unos aspectos en relación a objetivos futuros basados en redes sociales y valores económicos especialmente en lo concerniente a la falta de instrumentos y conceptos que ayuden a integrar conceptos uní dimensionales en bidimensionales.

### **8.1.1. Dinámica de la red social y su evolución**

En los estudios hechos por Hansen (1999) y Gulati y Gargulio (1999) se muestran métodos para analizar los cambios de redes en su evolución, Stockman y Doreian (1997) diferencia explícitamente entre la dinámica de una red y su evolución. Debido a que estos contextos son abarcados en el campo del análisis de redes sociales, los objetivos seguidos, en especial en lo referente a anticipaciones, tienen en general como meta predecir la probabilidad de por ejemplo que un grupo se una a otro o que actores cambien su influencia en la red.

Un método comprobado, en el cual la evolución de la red desde un punto matemático lineal, no es en esos estudios bien especificado.

### **8.1.2. Déficit interpretativo**

Un problema encontrado y también descrito por Giddens (1984) , Mohr (1998) y Jansen (2006), consiste en la interpretación de la relación de los actores o estructuras y sus interacciones dentro de las instituciones y los valores morales y culturales de las mismas. Giddens hace una aproximación en la investigación de redes de compañías sobre como cognición y normas de una perspectiva científica se pueden conectar. Estos métodos son específicos y a pesar que existe metodologías, en general se refieren a redes sociales

históricas, donde desde mi punto de vista la meta es explicativa y no predictiva.

Aun así, el método mostrado en este trabajo no deja sacar conclusiones en su significado, para la red en sí. Es decir, los efectos a la red social por el hecho de que sus parámetros descriptivos cambien en el tiempo y muestren un determinado comportamiento, no son efecto de esta investigación, a pesar de que el resultado analítico muestra la influencia de ellos sobre el desarrollo económico.

### **8.1.3. Foco de las redes como variables aclaratorias**

En estudios clásicos de redes sociales, se han usado estos generalmente para responder preguntas específicas, tales como si la transferencia de conocimiento se produce por conexiones amplias (Weak ties) o conexiones fuertes (Strong ties) (Hansen M. T., 1999). Los resultados son importantes y aplicables en estudios empíricos cuantitativos y cualitativos en diferentes sectores, tales como el bancario (Baker, 1984), (Uzzi, 1999), por lo que la referencia económica sería en este caso esperada. Pero no es el caso. Las metas son en un sentido amplio entre el desarrollo económico y el desarrollo de la red, más sin embargo estos estudios recalcan la influencia de

valores puntuales descriptivos (cantidad de conexiones Fuertes) y no generales de la red.

## **8.2. Conclusiones – red social**

El análisis de las redes sociales, se está convirtiendo en una potente herramienta para determinar los fenómenos de relación entre agentes o actores. Más concretamente, y desde el punto de vista económico, las empresas necesitan analizar sus datos con el objetivo de recabar información que pueda ser útil para la toma de decisiones. Una fuente muy importante de información para la empresa pueden ser los correos electrónicos. La investigación ha demostrado que el estudio del correo electrónico proporciona una visión de la estructura de las redes de contacto de una empresa. Y nos proporciona un contexto a través del cual se puede analizar las relaciones internas y externas de la empresa. La obtención de datos del correo electrónico puede ser un proceso costoso y difícil puede requerir mucho tiempo para su procesamiento (consolidación, depuración, control de calidad etc.). El estudio, sin establecer premisas absolutas o únicas, muestra una forma de adquisición de la información a través de una metodología sencilla y factible que permite la captación de datos del sistema de correo electrónico de las empresas. Dicha información se ha tomado para analizar la estructura de la red y describir el comportamiento de la red social a lo largo del tiempo. La clasificación

hecha a nivel de los agentes sirve para en una forma macro analítica se saquen algunas, tales como influencia subjetiva de determinados grupos, cliques u organizaciones en el desarrollo de la red.

Desde el punto de vista de las implicaciones para la gestión, podemos señalar que esta metodología es factible de aplicación para las empresas, respetando los criterios legales de privacidad. El uso de esta técnica puede permitir a las empresas el acceso a una potencial fuente de datos que, modelizados a través del análisis de redes sociales, proporcione información sobre la evolución de la gestión de la empresa. En este sentido, a través de la red de relaciones de la empresa es factible analizar, las estrategias corporativas, de negocio y funcionales de la compañía. Por ejemplo, una tendencia a diversificar, puede suponer un mayor número de nuevos contactos, o la mayor densidad de la red, una intensificación en las relaciones con nuestros suministradores.

El trabajo no está exento de limitaciones. La principal es la necesidad de filtrar los datos de email, con el objetivo de buscar tendencias a largo plazo, y eliminar posibles espurios y ruido a corto plazo. Esta labor de filtrado, puede permitir, establecer patrones de comportamiento con las diferentes variables económicas de las empresas.

Esta contribución abre una línea de investigación que permite comparar la gestión interna de las empresas, en función de las características y evolución de la red social derivada de sus contactos a través del correo electrónico.

### **8.3. Conclusiones – efectos al desarrollo económico**

Los efectos de la red social al desarrollo económico de la empresa han sido posibles de comprobar, aunque no en una forma absoluta como en los inicios de la investigación se esperaban. El problema más grande que se encontró es el hecho de que los estudios de las redes sociales son en su mayoría puntuales en el tiempo y bidimensionales en su observación, ya que el objeto principal de la investigación típica en este campo, es el de estudiar el comportamiento de los miembros de la red, su fuerza y su cambio en la forma. Mientras que el desarrollo económico es unidimensional y en el tiempo no estable e influenciado por medidas de contabilidad estratégicas.

Qué valores genéricos de la red social en su desarrollo - o en su cualidad como tal - son aptos para tomar como valor puntual y qué significan en su correlación con los valores económicos, ha sido un reto y un desafío grande que ha tomado mucho tiempo de paciencia, persistencia y constancia, hasta encontrar un común valor que en su

desarrollo muestre un contexto claro en la interoperabilidad entre los dos mundos en investigación.

El hecho de que tanto la transitividad como el grado de la red social tengan una influencia en las entradas económicas o que la correlación de los costos y la correlación de la red social den una clara muestra de que la red social pueda ser estudiada a partir del uso de correos electrónicos y sin ayuda de cuestionarios tradicionales, no era de esperarse. En especial, debido al enfoque de la herramienta de software UCINET, en el desarrollo del estudio, se llega límites de diseño. Es por eso que una herramienta más universal como es R, con un marco de referencia más universal y versátil se constituye en la herramienta principal del estudio. Los últimos valores encontrados fueron calculados y comprobados con ayuda de Excel, basada en los datos adquiridos por medio de R.

La influencia de la red social en el desarrollo económico de una empresa mediana ha sido comprobada. Sin embargo, hechos tales como el mal manejo de gestión, la posible dependencia grande de otras empresas o catástrofes ambientales no implican que el desarrollo económico de la empresa entre en dificultades, a pesar de que el pronóstico de la influencia de la red social sea positivo.

Estas herramientas presentan instrumentos y metodologías que puede ser usadas para efectos preventivos y darle a la gestión de

la empresa la oportunidad de ver si su acción es positiva o no. Igualmente observando los resultados históricos se pueden usar desde el punto de vista económico para analizar el porqué del crecimiento o decrecimiento.

La influencia de la red social en el desarrollo de la empresa se ha comprobado, pero la aplicación metodológica a otras empresas puede variar significativamente de acuerdo a la posibilidad de acceso a los datos necesarios para el análisis. Este estudio no es representativo, sin embargo, muestra métodos detallados y claros que pueden ser aplicados a diferentes empresas con el objeto de analizar si sus redes sociales son determinantes o no en su desarrollo económico. Las dos herramientas usadas, R y MemBrain son accesibles y fáciles de usar. El apoyo de Excel en la transformación de los valores a formatos usables ha sido de un gran valor en esta investigación, pero dependiendo de las fuentes disponibles se podría prescindir de él.

#### **8.4. Puntos abiertos**

Como en toda investigación quedan muchas preguntas abiertas.

¿Qué otros valores de una red social podrían ser considerados como tales que influyeran en el desarrollo económico?

Si tuviéramos la posibilidad de usar todos los correos de todos los empleados, ¿cuál sería el desarrollo de la investigación?

Si no se tiene acceso al correo electrónico, ¿Qué tipo de información podría ser representativa? ¿Tráfico de internet? ¿Análisis del comportamiento telefónico?

¿Qué efectos tiene el hecho de que métodos de comunicación instantáneos tales como iMessage, WhatsApp, SMS u otros estén avanzando en el ámbito de negocios en un posible análisis futuro?

Otro campo que no se tocó en esta investigación pero que podría ser objeto de investigación complementaria, ya que los datos se encuentran disponibles, es el análisis de la red social interna de la empresa. Para esta investigación se tuvo solo en cuenta la interacción de la empresa con no miembros de la empresa excepto en los casos en que estos eran involucrados en la conversación electrónica. Pero como se ve en la Tabla 7 Características del contenido de la base de datos (p. 70) la cantidad de correos internos es mucho mayor.

Por último cabe resaltar, que la motivación para hacer este tipo de análisis puede ser en el futuro más grande debido a las nuevas leyes de conservación de datos<sup>23</sup>. Aunque la directiva aún no está implementada en su totalidad por los miembros de la CE (a

---

<sup>23</sup> DIRECTIVA 2006/24/CE DEL PARLAMENTO EUROPEO Y DEL CONSEJO

31.12.2014), puede ser una ayuda colateral el disponer de esos datos en forma legal. A pesar de ello si estos son anonimizados como se muestra aquí, pueden ser de mucha ayuda para investigaciones posteriores.

## Bibliografía

Adamic, L. A., & Adar, E. (2003). Friends and neighbors on the Web.

*Social Networks*, 25, págs. 211–230. Recuperado el 11 de 09 de 2013, de doi:10.1016/S0378-8733(03)00009-1:

<http://social.cs.uiuc.edu/class/cs591kgk/friendsadamic.pdf>

Ahmed, N., & Gokhale, D. (Mayo de 1989). Entropy expressions and their estimators for multivariate distributions. *IEEE Transactions on Information Theory*(35 Issue:3 ), págs. 688 - 692.

Albert, R., & Barabási, A. (2002). *Statistical mechanics of complex Networks. Reviews of Modern Physics*. Obtenido de <http://arxiv.org/pdf/cond-mat/0106096.pdf>

Aldrich, J. H., Nelson, F. D., & Sullivan, J. L. (1984). *Linear Probability Logit & Probit Models (Quantitative Applications in the Social Sciences)*. Sage Pubn Inc.

Alstyne, M., & Zhang, J. (2003). *EmailNet: A system for automatically mining social networks from organizational email communication*. Obtenido de <http://www.researchgate.net/publication/215439725>

Amschlinger, M. C. (2011). *Internationalisierung im Mittelstand: Eine empirische Analyse kulturabhängiger Erfolgsfaktoren von*

*Direktinvestitionen deutscher kleiner und mittlerer Unternehmen in Rumänien* (Vol. 21). (J.-A. Meyer, Ed.) Lohmar - Colonia: Josef Eul.

Arbib, M., Kfoury, A. J., & Moll, R. N. (1984). *A basis for theoretical computer science*. (D. Gries, Ed.) New York, Heidelberg, Berlin: Springer-Verlag.

Arellano, M. (25 de 07 de 2001). *Introducción al Análisis Clásico de Series de Tiempo*. Recuperado el 12 de 09 de 2013, de 5campus.com: <http://www.5campus.com/leccion/seriest>

Audioguiaroma. (2011). *Historia sencilla de la Antigua Roma*. Recuperado el 11 de 2011, de <http://www.historia-roma.com>: <http://www.historia-roma.com>

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2011). *Multivariate Analysemethoden: eine anwendungsorientierte Einführung*. Berlin, Heidelberg: Springer.

Baker, W. E. (1984). The social structure of a national security market. *Journal of Mathematical Sociology*, 16, págs. 775-881.

Barabási, A.-L. (2002). *Linked: How Everything is Connected to Everything Else and What It means for Business, Science and Everyday Life*. Cambridge: Perseus.

Barnard, A. (1992). Through Radcliffe-Brown's spectacles: reflections on the history of anthropology. *History of the Human Sciences*, 5(4), págs. 1-20. doi:10.1177/095269519200500401

Barnes, J. A. (1954). Class and Committee in a Norwegian Island Parish. *Human Relations*, 7.

Barnett, A. G., & Dobson, A. J. (2010). *Analysing Seasonal Health Data*. Springer: Berlin, Heidelberg.

Barnett, A. G., Baker, P. J., & Dobson, A. J. (2012). Manual "season: Analysing Seasonal Data R Functions". *R package version 0.3-2*.

Bartoň, S., Dohnal, V., Sedmidubský, J., & Zezula, P. (2 de 10 de 2007). *Gauging the Evolution of Metric Social Network*. Recuperado el 1 de 11 de 2013, de Faculty of Informatics, Masaryk University, Brno, Czech Republic:  
[http://is.muni.cz/th/60474/fi\\_r/report.pdf](http://is.muni.cz/th/60474/fi_r/report.pdf)

Basso, M. (27 de Diciembre de 2011). *Magic Quadrant for Enterprise Wireless Email Market*. Recuperado el 3 de Octubre de 2012, de Gartner(r):  
<http://www.gartner.com/technology/reprints.do?id=1-18KGC8M&ct=120105&st=sb>

Baym, N., Zhang, Y., & Lin, M. (2004). Social interactions across media: Interpersonal communication on the internet, telephone and face-to-face. *New Media & Society*, 6(3), págs. 299–318.

Bea, F. X., & J., H. (2012). *Strategisches Management* (6 ed.). Stuttgart: UTB.

Beaudreau, B. C. (2004). *World Trade: A Network Approach*. iUniverse.

Bekkerman, R., McCallum, A., & Huang, G. (2004). *Automatic Categorization of Email into Folders: Benchmark Experiments on Enron and SRI Corpora*. (Technical Report IR-418 ed.). (CIIR, Ed.) University of Massachusetts.

Berry, M., & Browne, M. (2005). Email Surveillance Using Nonnegative Matrix Factorization. En *Computational & Mathematical Organization Theory*, 11 (págs. 249–264).

Bird, C., Gourley, A., Devanbu, P., Swaminathan, P., & Gertz, M. (2006). Mining email social networks in postgres. En *MSR '06: Proceedings of the International Workshop on Mining Software Repositories* (págs. 225-237).

Bohn, A., Feinerer, I., Hornik, K., & Mair, P. (06 de 2011). Content-Based Social Network Analysis of mailing Lists. *The R Journal*, 3 (1), págs. 11-18.

- Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, 27, págs. 55-71.
- Borgatti, S. P. (2009). Network analysis in the social sciences. En *Science* 323 (pág. 892).
- Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. En *Social networks* (Vol. 19, págs. 243-269).
- Borgatti, S. P., & Foster, P. C. (2003). The Network Paradigm in Organizational Research: A Review and Typology. En *Journal Management* 29 (págs. 991-1013).
- Borgatti, S. P., & Halgin, D. S. (2011). On Network Theory. En *Organization Science* 22(5). 1168-1181.
- Borgatti, S. P., & Lopez-Kidwell. (2011). Network Theorizing. En P. Carrington, & J. Scott (Edits.), *The Sage Handbook of Social Network Analysis*. Sage Publications.
- Borgatti, S. P., Jones, C., & Everett, M. G. (1988). Network measures of social capital. En *Connections* 21 (2) (págs. 27-36).
- Borgatti, S. P., Mehra, A., Brass, D., & Labianca, G. (2009). Network Analysis in the Social Sciences. En *Science* 323 (5916) (págs. 892-895).

Borgatti, S., Everett, M., & Freeman, L. (2002). Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Borgatti, S., Everett, M., & Freeman, L. (2012). UCINET 5 for Windows, Software for Social Network Analysis - User's guide. (I. Analytic Technologies, Ed.)

Borkar, V. R., Deshmukh, K., & Sarawagi, S. (2000). Automatically extracting structure from free text addresses. *Boletín del IEEE Computer Society Technical committee in Data Engineering*.

Bortz, J. (1999). *Statistik für Sozialwissenschaftler*. Berlin, Heidelberg, New York, Barcelona, Hongkong, London, Mailand, Paris, Singapur, Tokio: Springer.

Bott, E. (1955). Urban Families: Conjugal Roles and Social networks. *Human relations*, 9.

Bott, E. (1964). *Family and social network: Roles, norms, and external relationships in ordinary urban families*. London: Tavistock.

Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. (R. Kreckel, Ed.) *Soziale Welt, Edición especial* 2(Soziale Ungleichheiten), págs. 183-198.

- Brass, D. (1995). A Social Network Perspective on Human Resources Management. En *Research in Personnel and Human Resources Management, 13* (págs. 39-79).
- Brosius, F. (2007). *SPSS für Dummies: [den Daten sei Dank! - durchdachte Entscheidungen durch Datenanalyse ; statistische Analyse statt Datenchaos ; Daten aufbereiten und analysieren ; grafische Darstellung der Ergebnisse ; Export in Excel, Word & Co.]*. Weinheim: Wiley-VCH-Verl.
- Burbules, N. C., & Smith, R. (2005). What it makes sense to say: Wittgenstein, rule-following and the nature of education. En *Educational Philosophy and Theory, 37(3)* (págs. 425-430).
- Burguière, A., Klapisch-Zuber, C., & Segalen, M. (1996). *Geschichte der Familie* (Vol. 1). (F. Zonabend, Ed.) Franckfort: Alianza Editorial.
- Burguière, A., Klapisch-Zuber, C., Segalen, M., & Zonabend, F. (1988). *Historia de la Familia - Tomo I con prólogos de Claude Lèvi-Strauss y Georges Duby* (Vol. 1). Madrid: Alianza Editorial.
- Burkart, M., Panunzi, F., & Shleifer, A. (Septiembre de 2002). Family Firms. *NOTA DI LAVORO 74.2002*. The Fondazione Eni Enrico Mattei/Università di Bologna.

Burt, R. (1992). *Structural Holes*. Cambridge: Harvard University Press.

Butts, C. T. (2013). sna: Tools for Social Network Analysis. *R package version 2.3-1*. Obtenido de <http://CRAN.R-project.org/package=sna>

Butts, C. T., & Carley, K. M. (2001). *Multivariate Methods for Inter-Structural Analysis*. Mellon University, Department of Social and Decision Sciences, Carnegie. Recuperado el 05 de 01 de 2014, de <http://www.casos.cs.cmu.edu/publications/papers/multiv001a.pdf>

Cain, M. W. (9 de Agosto de 2012). *MarketScope for Email Systems*. Recuperado el 3 de Octubre de 2012, de Gartner(r): <http://www.gartner.com/technology/reprints.do?id=1-1BOC85F&ct=120809&st=sb>

Camiller, P., & Monsutti, A. (2005). *War and Migration: Social Networks and Economic Strategies of the Hazaras of Afghanistan (Middle East Studies: History, Politics & Law)*. Routledge Chapman & Hall.

- Cartwright, D., & Harary, F. (1956). Structural Balance: A Generalization of Heider's Theory. *Psychological Review*, 63, pág. 277.
- Casson, M., Buckley, P. J., Dark, K., & Della Giusta, M. (2010). *Entrepreneurship: Theory, Networks, History*. Edward Elgar Publishing Ltd.
- Castells, M. (2001). La era de la información. En *La Sociedad en Red* (1 ed., Vol. 1). Madrid: Alianza Ed.
- Chapanond, A., Krishnamoorthy, M., & Yener, B. (2005). Graph Theoretic and Spectral Analysis of Enron Email Data. En *Computer Mathematical Organization Theory* 11(3) (págs. 265-281).
- Colaboradores de Wikipedia. (5 de Sep. de 2013). *Hojas de Cálculo*. (Wikipedia, La enciclopedia libre, Ed.) Recuperado el 12 de 09 de 2013, de [http://es.wikipedia.org/wiki/Hoja\\_de\\_c%C3%A1lculo](http://es.wikipedia.org/wiki/Hoja_de_c%C3%A1lculo)
- Colaboradores de Wikipedia. (16 de 07 de 2013). *Lenguaje R*. (Wikipedia, La enciclopedia libre, Editor) Recuperado el 12 de 09 de 2013, de [http://es.wikipedia.org/wiki/Lenguaje\\_R](http://es.wikipedia.org/wiki/Lenguaje_R)

Colaboradores de Wikipedia. (13 de 07 de 2013). *SPSS*. (Wikipedia, La enciclopedia libre, Editor) Recuperado el 12 de 09 de 2013, de <http://es.wikipedia.org/wiki/SPSS>

Coleman, J. (1986). Social theory, social research, and a theory of action. *American Journal of Sociology, 91 (supplement)*, págs. 1309-1335.

Coleman, J. (1998). Social capital in the creation of human capital. *American Journal of Sociology, 94 (supplement)*, págs. 95-120.

Comisión Europea de la UE, . (20 de 5 de 2003). Recomendación 2003/361/CE de la Comisión, de 6 de mayo de 2003, sobre la definición de microempresas, pequeñas y medianas empresas. *Diario Oficial L 124*.

Comisión Europea de la UE, . (2005). *The new SME definition - A guide and model declaration*. Publications Office of EC.

Contractor N., S., & Monge, P. (2003). *Theories of communication networks*. New York: Oxford University Press.

Crocker, D. H. (13 de Agosto de 1982). Standard for the format of Arpa Internet Text Messages. Dept. of Electrical Engineering, University of Delaware.

- Crosona, R., & Gächterb, S. (2009). The science of experimental economics. *Journal of Economic Behavior & Organization*, 73, págs. 122–131.
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, pág. 1695. Obtenido de <http://igraph.sf.net>
- Cuadras, C. M. (2011). Nuevos métodos de análisis multivariante. Barcelona.
- Culotta, A., Bekkermann, R., & McCallum, A. (2004). Extracting social networks and contact information from email and the Web. *Resumen de Conference on Email and Anti-Spam CEAS*.
- de Klepper, M., Sleebos, E., van de Bunt, G., & Agneessens, F. (2010). Similarity in friendship networks: Selection or influence? The effect of constraining contexts and non-visible individual attributes. *Social Networks*, 32, págs. 82-90.
- Diesner, J., & Carley, K. (2005). Revealing and Comparing the Organizational Structure of Covert Networks with Network Text Analysis. *XXV Sunbelt Social Network Conference. February 16-20 2005*. Redondo Beach, CA.
- Diesner, J., Frantz, T. L., & Carley, K. M. (2005). Communication networks from the Enron e-mail corpus it's always about the

- people Enron is no different. En *Computational and Mathematical Organization Theory*, 11 (3) (págs. 201–228).
- Doreian, P. (1997). *Evolution of Social Networks*. (F. N. Stokman, Ed.) New York: Gordon and Breach.
- Dubossarsky, E., & Norrie, M. (2012). snort: Social Network-Analysis On Relational Tables. *R package version 1.0*. Obtenido de <http://CRAN.R-project.org/package=snort>
- Dürr, M. (2013). *Sicherheit und Privatsphäre in Online Sozialen Netzwerken*. München: Ludwig–Maximilians–Universität München.
- Ebel, H., Mielsch, L., & Bornholdt, S. (2002). Scale-free topology of e-mail networks. Obtenido de arXiv:cond-mat/0201476v2 [cond-mat.dis-nn].
- Esser, H. (1987). Zum Verhältnis von qualitative und quantitative Methoden in der Sozialforschung. *Methoden der Biographie- und Lebenslaufforschung*, págs. 87-101.
- Fahrmeir, L., Hamerle, A., & Tutz, G. (1996). *Multivariate statistische Verfahren*. Gruyter.
- Faust, S., & Wasserman, K. (1998). *Social network analysis : methods and applications*. Cambridge: Cambridge Univ. Press.

- Frank, O. (2005). Network Sampling and Model Fitting. En J. S. Carrington, & S. S. Wasserman, *Models and Methods in Social Network Analysis* (Vols. 31–56). Cambridge: Cambridge Univ. Press.
- Freeman, L. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*(1.3), págs. 215-239.
- Freemann, L. C., White, D. R., & Romney, A. K. (1989). *Research Methods in Social Network Analysis*. New Brunswick, NJ: Transaction Books.
- Fremann, L. (1977). A set of measures of centrality based upon betweenness. *Sociometry*(40), págs. 35-41.
- Fridson, M., & Alvarez, F. (2002). The Adversarial Nature of Financial Reporting. En M. Fridson, & F. Alvarez, *Financial Statement Analysis: A Practitioner's Guide: A Practitioner's Guide (Wiley Finance)* (págs. 3-6;74-83). Crystal Dreams Pub.
- García Ruiz, J. L. (1995). *Historia económica de la empresa moderna*. Madrid: Ediciones Istmo.
- Giddens, A. (1984). *The constitution of society. Outline of the theory of structuration*. Cambridge: Polity Press.

Gong, N. Z., Xu, W., Huang, L., Mittal, P., Stefanov, E., Sekar, V., & Song, D. (14-16 de 11 de 2012). *Evolution of Social-Attribute Networks: Measurements, Modeling, and Implications using Google+*. Recuperado el 1 de 11 de 2013, de EECS, UC Berkeley:

<http://www.cs.berkeley.edu/~stevgong/papers/imc12.pdf>

González Casimiro, M. P. (2009). *Técnicas de predicción económica*. Universidad del País Vasco (UPV-EHU), Facultad de Ciencias Económicas y Empresariales. Sarriko-on.

Granovetter, M. (1973). The strenght of weak ties. *American Journal of Sociology*, 78, págs. 1360-1364.

Greene, W. H. (2002). *Economics Analysis*. Pearson.

Gross, R., & Acquisti, A. (2005). Information Revelation and Privacy in Online Social Networks (The Facebook case). *Procceding ACM Workshop on Privacy in the Electronic Society (WPES)*. AC Workshop.

Gruzd, A. (2012). *Non-Academic and Academic Social Networking Sites for Online Scholarly Communities Social Media for Academics: A practical guide*. (D. Neal, Ed.) Chandos Publishing.

Gruzd, A., & Haythornthwaite, C. A. (2013). Enabling Community Through Social Media. *Journal Mediated Internet Research*, 15(10), pág. 248.

Guinalíu, M., & Flavián, C. (2003). La Comunidad Virtual. Apuntes para asignatura Economía del Comercio Electrónico. (U. d. Zaragoza, Ed.) Obtenido de <http://www.5campus.org/leccion/comunidadvirtual/comunidadvirtual.doc>

Gulati, R., & Gargiulo, M. (1999). Where do interorganizational networks come from? *American Journal of Sociology*, 29, págs. 1439-1493.

Günterberg, B., & Pahnke, A. (11 de 09 de 2012). *Kennzahlen zum Mittelstand 2010/2012 in Deutschland*. Recuperado el 03 de 10 de 2012, de Institut für Mittelstandsforschung Bonn: <http://www.ifm-bonn.org/index.php?id=99>

Haagedorn, J., Link, A. N., & Vonortas, N. S. (2000). .Research partnerships. En *Research Policy* 29 (págs. 567-586).

Halgin, D. (2008). *An Introduction to Ucinet and NetDraw*. Recuperado el 21 de 01 de 2012, de Boston College: [http://www.hks.harvard.edu/netgov/files/NIPS/Halgin\\_NIPS\\_2008.pdf](http://www.hks.harvard.edu/netgov/files/NIPS/Halgin_NIPS_2008.pdf)

- Hannan, M. T., & Carroll, G. R. (1992). *Dynamics of Organizational Populations*. New York, Oxford: Oxford University Press.
- Hanneman, R., & Riddle, M. (2005). Introduction to social network methods. Riverside, University of California. Obtenido de <http://faculty.ucr.edu/~hanneman/>
- Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly*, 44, págs. 82-111.
- Hansen, N. (2011). The CMA Evolution Strategy: A Tutorial. Recuperado el 01 de 07 de 2015, de <https://www.lri.fr/~hansen/cmaesintro.html>
- Hassing, L. (Ed.). (1989). *Computer verstehen* (Vol. Computer der Zukunft). Turnhout, Belgica: Time-Life Books B. V.
- Hatzinger, R., Hornik, K., & Nagel, H. (2011). *R - Einführung durch angewandte Statistik*. München: Pearson Education.
- Haythornthwaite, C. A., & Wellman, B. (1988). Work, friendship and media use for information exchange in a networked organization. En *Journal of the American Society for Information Science*, 49(12) (págs. 1101-1114).
- Hebb, D. (1949). *Organisation of Behavior*. New York: Wiley.

Heidler, R. (2 de 10 de 2007). *Zur Evolution sozialer Netzwerke - theoretische Implikationen einer aktorsbasierten Methode*. Recuperado el 1 de 11 de 2013, de Universidad Frankfurt am Main: [http://www.soz.uni-frankfurt.de/Netzwerktagung/Heidler\\_evolution\\_sozialer\\_Netzwerke.pdf](http://www.soz.uni-frankfurt.de/Netzwerktagung/Heidler_evolution_sozialer_Netzwerke.pdf)

Henningsen, A. (2012). micEcon: Microeconomic Analysis and Modelling. *R package version 0.6-10*. Obtenido de <http://CRAN.R-project.org/package=micEcon>

Hibbert, C. (1999). *The House of Medici: Its Rise and Fall*. William Morrow Paperbacks .

Hipp, J. R., Faris, R. W., & Boessen, A. (2012). Measuring 'neighborhood': Constructing network neighborhoods. *Social Networks, Volume 34(1)*, págs. 128-140.

Ho, R. (2006). *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS*. Crc Pr Inc.

Hoffmann, N. (1993). *Neuronale Netze*. Braunschweig, Wiesbaden: Vieweg.

Hoffmeister, W. (2008). Investitionsrechnung und Nutzwertanalyse (Evaluación de inversiones y análisis de costo-beneficio). págs. 278 - 307.

- Holmes, E. E., Ward, E. J., & Wills, K. (2012). MARSS: Multivariate autoregressive state-space models for analyzing time-series data. *The R Journal*, 4(1), págs. 11-19.
- Holmes, E., & Wills, E. W. (2012). *MARSS: Multivariate Autoregressive State-Space Modeling*. Recuperado el 08 de 10 de 2012, de R package version 3.2: <http://cran.r-project.org/web/packages/MARSS/>
- Homans, G. (1972). *Grundfragen soziologischer Theorie*. Opladen: Westdeutscher Verlag.
- Hückstädt, J. (1993). *Schnell-Übersicht Visual Basic für Windows*. (J. Steiner, & R. Valentin, Edits.) Haar bei München/Alemania: Markt&Technik.
- Huinink, J., & Konietzka, D. (2007). Familiensoziologie: Eine Einführung. En *Begriffliche undtheoretische Grundlagen* (págs. 23-55). Campus Verlag GmbH.
- Huisman, M., & van Duijn, M. A. (2011). A reader's guide to SNA software. En J. Scott, & P. Carrington (Edits.), *The SAGE Handbook of Social Network Analysis* (págs. 578-600). London: SAGE.
- Hummell, H.-J., & Opp, K.-D. (1972). *Die Reduzierbarkeit der Soziologie auf Psychologie*. Braunschweig: Vieweg.

Hurlbert, J. S., Beggs, J. J., & Haines, V. A. (2008). Social Networks and Social Capital in Extreme Environments. En K. S. Nan Lin (Ed.), *Social Capital: Theory and Research*. New Brunswick, London: Aldine Transaction.

Ibaraki, T., Kato, N., & Ono, H. (2003). *Algorithms and Computation: 14th International Symposium, ISAAC 2003*. Springer.

Jansen, D. (2006). *Einführung in die Netzwerkanalyse* (3. ed.). Wiesbaden: VS Verlag für Sozialwissenschaften.

Jansen, D. (2007). *Theoriekonzepte in der Analyse sozialer Netzwerke*. Speyer: Deutsches Forschungsinstitut für öffentliche Verwaltung .

Jansen, W. (1999). Mobile Agents And Security. *Computer Communications*, 23.

Jetter, T. (2004). *MemBrain Neural Network Simulator*. Mainz-Kostheim: Dipl. Ing. Thomas Jetter.

Johnson, C., & Brown, C. (1988). *Cognizers: neural Networks and machines that think*. New York: John Wiley & Son.

- Jones, Q. (1977). Virtual-communities, Virtual settlements & cyber-archaeology: A theoretical outline. *Journal of Computer Mediated Communication*, 3(3).
- Kadushin, C. (2012). *Understanding Social Networks: Theories, Concepts, and Findings*. Oxford : Oxford University Press.
- Kali, R., & Reyes, J. (2007). The architecture of globalization: a network approach to international economic integration. *Journal of International Business Studies*, 38, págs. 595–620.
- Keila, P., & Skillicorn, D. (2005). *Detecting Unusual and Deceptive Communication in Email*. Ontario, Canada: Queen's University. Obtenido de <http://research.cs.queensu.ca/TechReports/Reports/2005-498.pdf>
- Kennedy, P. (1989). *Rise and Fall of the Great Powers: Economic Change and Military Conflict from 1500-2000*. Fontana.
- Kennedy, P. (2011). *Aufstieg und Fall der großen Mächte*. Fischer Taschenbuchverlag.
- Kim, J., Kim, G., Park, H., & Rice, R. (2007). Configurations of Relationships in Different Media: FtF, Email, Instant Messenger, Mobile Phone, and SMS. En *Journal of Computer-Mediated Communication* 12 (págs. 1183–1207).

- Kleiber, C., & Zeileis, A. (2008). Applied Econometrics with R. *R package*. Obtenido de <http://CRAN.R-project.org/package=AER>
- Klensin (Ed.), J. (Abril de 2001). Simple Mail Transfer Protocol (RFC2821). The Internet Society/ Network Working Group.
- Klensin, J., Freed, N., Rose, M., Stefferud, E., & Crocker, D. (Julio de 1994). SMTP Service Extensions (RFC 1651). Network Working Group.
- Klimt, B., & Yang, Y. (2004). The Enron Corpus: A New Dataset for Email Classification Research. En *European Conference on Machine Learning*.
- Kmenta, J. (1997). *Elements of Econometrics: Solution Manual*. Univ of Michigan Press.
- Knoke, D., & Yang, S. (2008). Social Network Analysis. En *Series: Quantitative Applications in the Social Sciences* (2 ed.). Los Angeles: Sage publications Inc.
- Kohler, W. (1924 (Org.)). *The Mentality of Apes*. Fisher Press (10-01-2011).
- Kokali, L., Paffenholz, G., & Moog, P. (2003). Neue Tendenzen in der Mittelstandsfinanzierung. En *Documentos sobre investigación*

*de PYME* (Vol. 99). Institut für Mittelstandsforschung Bonn  
(Ed.).

Korte, B., & Vygen, J. (2008). *Combinatorial Optimization. Theory and Algorithms*. (4a. ed.). Berlin: Springer.

Kossinets, G., & Watts, D. J. (06 de 01 de 2006). Empirical Analysis of an Evolving Social Network. *Science*(311), págs. 88-92.

Krumke, S. O., & Noltemeier, H. (2012). Graphentheoretische Konzepte und Algorithmen. Vieweg+Teubner Verlag.

Kruse, P., Dittler, A., & Schomburg, F. (2007). nextexpertizer und nextcoach: Kompetenzmessung aus der Sicht der Theorie kognitiver Selbstorganisation. (J. Erpenbeck, & L. v. Rosenstiel, Edits.) *Manual de medición de competencias*, págs. 515-543.

Kuper, A. (1977). *The Social Anthropology of Radcliffe-Brown*. Londres, Henley y Boston: Routledge & Kegan Paul.

Leinhardt, S. (1977). *Social Networks: A Developing Paradigm*. New York, San Francisco, London: Academic Press.

Levitt, M. (Julio de 2006). Worldwide Integrated Collaborative Environments - 2006-2010 Forecast and 2005 Vendor Shares. *Market Analysis, 1(Markets)*. IDC.

Löffelmann, K., & Vüllers, D. (1994). *Visual Basic 3.0*. München: Te-wi-Verlag.

Mackay, H. (1997). *Networking : das Buch über die Kunst, Beziehungen aufzubauen und zu nutzen / Harvey Mackay*. Düsseldorf, München: ECON.

Malik, F. (16 de Dic. de 2003). *Die malik-Kolumne: Gefährliche Worte*. Recuperado el 07 de 08 de 2012, de manager magazin online: <http://www.manager-magazin.de/unternehmen/karriere/0,2828,278499,00.html>

Malkin, I., Constantakopoulou, C., & Panagopoulou, K. (2012). *Greek and Roman Networks in the Mediterranean*. Routledge.

McCallum, A., Wang, X., & Corrada-Emmanuel, A. (2007). Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email. En *Journal of Artificial Intelligence Research* 30 (págs. 249-272).

McCulloch, W., & Pitts, W. (Diciembre de 1943). A logical calculus of the ideas immanent in nervous activity. *The bolletin of mathematical biophysics*.

McDonnell, S. (08 de 01 de 2001). *EBITDA*. Recuperado el 08 de 07 de 2012, de Computerworld: <http://www.computerworld.com/s/article/55895/EBITDA>

- McMillan, D., & Chavis, D. (1986). Sense of community: A definition and theory. En *Journal of Community Psychology*, 14 (págs. 6-23).
- Melanchthon, D. (2010). Microsoft Exchange. (Microsoft, Ed.) *Presentación en el CEBIT 2010*, pág. 2.
- Microsoft Corporation. (1993). *Visualbasic Programming System for Windows Professional Features Book 1*. Microsoft Corporation.
- Microsoft Corporation. (1993). *Visualbasic Programming System for Windows Professional Features Book 2*. Microsoft Corporation.
- Miller, D., & Le Breton-Miller, I. (2005). Management Insights from Great and Struggling Family Businesses. En *Long Range Planning* (Vol. 38/6, págs. 517–530). Elsevier.
- Mitchell, J. C. (1969). The Concept and Use of Social networks. En J. C. Mitchell (Ed.), *Social Networks in Urban Situations*. Manchester: Manchester University Press.
- Mohr, J. (1998). Measuring meaning structures. *Annual Review of Sociology*, 24, págs. 345-370.
- Molina, J. L. (2004). La ciencia de la redes. En 1. Apuntes de Ciencia y Tecnología.

- Morris, C. (30 de Marzo de 2011). Man And His Ancestor A Study In Evolution.
- Mosler, K., & Schmid, F. (2011). *Wahrscheinlichkeitsrechnung und schließende Statistik*. Berlin, Heidelberg: Springer.
- Mugler, J. (2008). *Grundlagen der BWL der Klein- und Mittelbetriebe* (2a. ed.). Viena: facultas.vuw.
- Nagel, H., & Hatzinger, R. (2009). *SPSS Statistics: Statistische Methoden und Fallbeispiele*. München: Pearson Studium .
- Nelson, J. (04 de 03 de 2013). *A Wolf or a Ram? What is Wolfram Alpha?* Recuperado el 12 de 09 de 2013, de Public Libraries Online: <http://publiclibrariesonline.org/2013/03/a-wolf-or-a-ram-what-is-wolfram-alpha/>
- Newman, M. (2001). Who is the best connected scientist? a study of scientific coauthorship networks. *Working Papers 00-12-064*. Santa Fe Institute.
- Newman, M. E. (2010). *Networks: An Introduction by M. E. J. Newman, a college-level textbook about the science of networks*. Oxford: Oxford University Press.
- Núñez, J. F., & Cárdenas, E. (2013). Identificación de asociaciones y complicidades, vía-email, de un grupo de analistas de redes

sociales: ¿Qué intercambian los rederos? *Revista Hispana para el Análisis de Redes Sociales*, 24(1), págs. 27-52.

Ostoic, J. A. (2013). multiplex: Analysis of Multiple Social Networks with Algebra. *R package version 1.0*. Obtenido de <http://CRAN.R-project.org/package=multiplex>

Parkhe, A., Wasserman, S., & Ralston, D. A. (2006). New Frontiers in Network Theory Development. En *Academy of Management Review*, 31 (págs. 560-569).

Parsons, T. (1976). *Zur Theorie sozialer Systeme*. Opladen: Westdeutscher Verlag.

Parthasarathi, P., & Riello, G. (2011). *The Spinning World: A Global History of Cotton Textiles, 1200-1850*. Oxford Univ Pr.

Popp, A., & Wilson, J. (2003). *Industrial Clusters and Regional Business Networks in England, 1750-1970 (Modern Economic and Social History)*. Ashgate Publishing Limited.

Postel, J. B. (Agosto de 1982). Simple Mail Transfer Protocol. Information Sciences Institute, University of Southern California.

Potter, B., Maxwell, T., & Scott, B. (1993). *Visual Basic Superbible* (2 ed.). (M. Waite, Ed.) Waite Group Press.

- Priebe, E. C., Conroy, M. J., & Marchette, J. D. (2005). Scan statistics on Enron graphs. En *Computational and Mathematical Organization Theory*, 11(3) (págs. 229–247).
- Reips, U.-D. (2002). Standards for Internet-based experimenting. En *Experimental Psychology*, 49(4) (págs. 243-256).
- Resnic (Ed.), P. (Abril de 2001). Internet Message Format (RFC 2822). The Internet Society/Network Working Group.
- Rey, G. D., & Wender, K. F. (2010). *Neuronale Netze: Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung*. Verlag Hans Huber.
- Rodriguez, F. (1994). *Steuerung eines Segelbootes mit Hilfe neuronaler Netze*. Bremen: Universität Bremen.
- Rodriguez, F. (2009). *El cálculo del capital humano en empresas medianas tomando como ejemplo al aeropuerto de Münster y Osnabrück en Alemania*. Curso de Economía de la tecnología, Dpto. Economía Aplicada e Historia Económica. UNED.
- Rodriguez, F. (2010). *El desarrollo de las redes sociales en función del tiempo de empresas PYME a nivel europeo*. Dpto. Economía Aplicada e Historia Económica. UNED.

Rodriguez, F., Arranz, N., & Fernandez, J. C. (2016). Proceso de adquisición de datos del sistema de correo electrónico: Una aplicación a la modelización de una red social. *Redes*, 27(1). doi:<http://dx.doi.org/10.5565/rev/redes.589>

Rowe, R., Creamer, G., Hershkop, S., & Stolfo, S. J. (2007). Automated social hierarchy detection through email network analysis. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007. En *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007. Workshop on Web mining and social network analysis* (págs. 109–117).

Rürup, M., Rübken, H., Emmerich, M., & Dunkake, I. (2015). *Netzwerke im Bildungswesen: Eine Einführung in ihre Analyse und Gestaltung*. Wiesbaden: Springer Fachmedien. doi: 10.1007/978-3-658-06737-3\_2

Saisana, M. (2012). *Monitoring SMEs'performance in Europe Indicators fit for purpose*. European Commission. JRC - Institute for the Protection and Security of the Citizen.

Schlittgen, R. (2012). *Angewandte Zeitreihenanalyse mit R* (2. ed.). München: Oldenburg wissenschaftsverlag GmbH.

Schmidtke, M. (9 de Julio de 2006). *Teoría de grafos*. Recuperado el 23 de Octubre de 2012, de Enciclopedia Libre Universal en

Español:

[http://enciclopedia.us.es/index.php/Teor%C3%ADa\\_de\\_grafos](http://enciclopedia.us.es/index.php/Teor%C3%ADa_de_grafos)

Scholtyssek, J. (2011). *Der Aufstieg der Quandts: Eine deutsche Unternehmerdynastie*. C.H.Beck.

Schubert, H. (Ed.). (2008). *Netzwerk-Management, Koordination von professionellen Vernetzungen – Grundlagen und Beispielen*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Scott, J. (1991). *Social Network Analysis: A Handbook*. Londres: Sage.

Scott, J. (2000). *Social Network Analysis* (2 ed.). Los Angeles: Sage publications Inc.

Shakes, J., Langheinrich, M., & Etzioni, O. (1997). Dynamic reference sifting: A case study in the homepage domain. *Resumen del 6º World Wide Web Conference* .

Shetty, J., & Adibi, J. (2004). The Enron email dataset database schema and brief statistical report. En *Information Sciences Institute. Technical Report*. University of Southern California.

Simmel, G. (1890). *Über sociale Differenzierung : soziologische und psychologische Untersuchungen*. Leipzig: Duncker & Humblot.

- Smith, M. (1999). Invisible crowds in cyberspace: Measuring and mapping the social structure of usenet. En *Communities in Cyberspace: Perspectives on New Forms of Social Organization*. Londres: Routledge Press.
- Smith, S., & Jain, A. (Septiembre de 1988). A test to determine the multivariate normality of a data set. *IEEE Transactions on Pattern Analysis and Machine Intelligence*(10 Issue:5 ), págs. 757 - 761.
- Smith, V. (2009). Theory and experiment: What are the questions? *Journal of Economic Behavior & Organization*, 73, págs. 3–15.
- Snijders, T. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology.*, págs. 361–395.
- Snijders, T., van de Bunt, G., & Steglich, C. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32, págs. 44–60.
- SPSS Inc. (2007). *SPSS Base 16.0 User's Guide*. Prentice Hall.
- Standage, T. (2013). *Writing on the Wall: Social Media - The First 2,000 Years* (Kindle Edition ed.). USA: Bloomsbury USA.

- Stegbauer, C. (2005). *Mediensoziologie: Grundfragen und Forschungsfelder*. En M. Jäckel (Ed.). Verlag für Sozialwissenschaften.
- Stokman, F. N., & Doreian, P. (1997). *Evolution of Social networks: Processes and principles*.
- Sun, C. (2013). erer: Empirical Research in Economics with R. *R package version 1.4*. Obtenido de <http://CRAN.R-project.org/package=erer>
- Trier, M. (2008). Research Note: Towards Dynamic Visualization for Understanding Evolution of Digital Communication Networks. En *Information Systems Research*, 19(3) (pág. 350).
- Upstill, T., Craswell, N., & Hawking, D. (2003). Query independent evidence in home page finding. *ACM Transactions On Information Systems*.
- Uzzi, B. (1999). Embeddedness in the Making of Financial Capital: How Social Relations and Networks Benefit Firms Seeking Finance. *American Sociological Review*, 64, págs. 482-505.
- Van Alstyne, M., & Zhang, J. (2003). Emailnet: A system for automatically mining social networks from organizational email communication. *Resumen del NAACSOS2003 (North American*

*Association for Computational Social and Organizational Science*).

Vetter, S., & Köhler, J. (2014). *Unternehmensdemographie und – dynamik in Europa. Aktuelle Trends in der Unternehmenslandschaft. Research Briefing*. Deutsche Bank AG. Frankfurt am Main: Deutsche Bank Research.

W. N. Venables, D. M. (2013). *An Introduction to R - Manual parte de la distribución de R*.

Wassermann, S., & Faust, K. (1999). *Social Network Analysis – Methods and Applications*. Cambridge : Cambridge University Press.

Wassermann, S., & Faust, K. (2007). *Social network analysis: Methods*. Cambridge: Cambridge University Press.

Weber, J., & Schäffer, U. (2008). *Einführung in das Controlling* (12 ed.). Schäffer-Poeschel.

Weisstein, E. W. (6 de 09 de 2012). *Determinant*. Recuperado el 10 de 09 de 2012, de MathWorld:  
<http://mathworld.wolfram.com/Determinant.html>

Weisstein, E. W. (6 de 09 de 2012). *Oriented Graph*. Recuperado el 10 de 09 de 2012, de MathWorld:

<http://mathworld.wolfram.com/OrientedGraph.html>

Welfens, P. J., Zoche, P., Jungmittag, A., Beckert, B., & Joisten, M. (2005). *Internetwirtschaft 2010: Perspektiven und Auswirkungen. Studie für das Bundesministerium für Wirtschaft und Arbeit*. Heidelberg: Physica-Verlag .

Wellman, B. (2002). Little Boxes, Glocalization, and Networked Individualism. *Proceeding: Revised Papers from the Second Kyoto Workshop on Digital Cities II, Computational and Sociological Approaches*.

Wellman, B., & Leighton, B. (1979). Networks, Neighborhoods and Communities. Approach to the Study of the Community Question. En *Urban Affairs Quartely*, 14(3) (págs. 363-390).

Wellmann, K. (2002). *Das Intranet als Kommunikationsinstrument im internen Marketing*. Hamburg: Diplomatica GmbH.

Weyer, J. (2000). *Soziale Netzwerke*. Munchen: Oldenbourg Wissensch.

Weyer, J. (2000). *Soziale Netzwerke : Konzepte und Methoden der sozialwissenschaftlichen Netzwerkforschung*. München, Wien: Oldenbourg.

- White, D. R., & Jorion, P. (1992). Representing and Computing Kinship: A New [Network] Approach. *Current Anthropology*, 33(4), págs. 454-463.
- White, H. C. (1963). *Anatomy of Kinship*. Englewood Cliffs: Prentice Hall.
- Wilhelm, B. E. (2000). *Systemversagen im Innovationsprozess: Zur Reorganisation des Wissens- und Technologietransfers*. Wiesbaden: Springer Fachmedien.
- Wilson, T. (1982). Qualitative 'oder' quantitative Methoden in der Sozialforschung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 34.
- Wissen.de. (1 de Enero de 2011). Und täglich grüßt die Evolution: Wie der Mensch wurde, was er ist. Wissenmedia.
- Wymenga, P., Spanikova, D. V., Barker, A., Konings, D., & Canton, D. (2012). *EU SMEs in 2012: at the crossroads - Annual report on small and medium - sized enterprises in the EU*. OECD. Rotterdam: Ecorys.
- Xi, W., Fox, E. A., Shu, J., & Tan, R. (2002). Machine learning approach for homepage finding task. *Resumen del 9º International Symposium on String Processing and Information Retrieval*, págs. 145-159.

Yardi, S. (2009). Social Learning and technical capital on the social web. *The ACM Student Journal Crossroads*, 16.2, págs. 9-11.

Yelupula, K., & Ramaswamy, S. (2008). Social network analysis for email classification. En *Proceedings ACM, 2008. 46th Annual Southeast Regional Conference* (págs. 469-474).

## A. Apéndice – Programas en R

### A.1. Programa para computar las principales características de cada red social por año y mes

```
#####
# Program to extract properties of a Network from a group of files
# Every file has data for a specific month in matrix form
# Every node has a weight with the amount of exchanged emails
#
# For the investigation following properties become a sense:
# Number of components, degree, density, transitivity,
# number of ties, diameter, number of edges (ecount) and
# number of vertices (vcount)
#
# The result is extracted to a csv file to be read on eg Excel
#
# Autor   : Francisco Rodriguez
# Date    : 05.01.2013
# Version: 1.0
#####

# Libraries for Social Network Analysing
library(sna)
library(igraph)
# Working directory
setwd('W:\\doctorado\\DATA-R')
# Definition of vectors with year and month data
Year<-c(2002:2013)
Month<-c(1:12)
# Initialisation of the matrix to be fulfil with the calculated data
MatrixF<-c('year', 'month', 'Components', 'Degree', 'Density', _
'Transitivity', 'Ties', 'Diameter', 'ECount', 'VCount')
# Start of the loop
for (y in Year) {
  for (m in Month){
# Data from a file is reading
    g<-read.table(sprintf('%s-%s.csv', y, m), row.names=1, sep=';')
    # the table is converted into a matrix
    gM<-as.matrix(g)
# For some calculation is necessary to have a graph instead of a matrix
```

```

    gG<-graph.adjacency(gM, weighted=T, mode='directed')
# Start of the calculation of the properties
    gComponents<-components(gM) # Number of components
    gDegree<-sum(gM)           # Degree of the network
    gDensity<-gden(gM)         # Density of the network
    gTransitivity<-gtrans(gM)  # Transitivity of the network
    gTies<-nties(gM)           # Number of ties
    gDiameter<-diameter(gG)    # Diameter of the graph
    gECount<-ecount(gG)        # Number of edges on the graph
    gVCount<-vcount(gG)        # Number of vertices on the graph
# The values are put together to be exported
    VectorT<-c(y, m, gComponents, gDegree, gDensity, _ gTransitivity, gTies,
gDiameter, gECount, gVCount)
    MatrixF<-rbind(MatrixF, VectorT)
}}
# Export of the calculated data
write.csv2(MatrixF, 'DataFinal.csv')
#####
# END
#####

```

## A.2. Programa para calcular los valores de evolución de la red social en transcurso del tiempo (Covariancia y correlación)

```

#####
# Program to extract properties of an analysis of the evolution of
# a Network step by step
# Graph theory is basis of the calculation
# Every file has data for a specific month in matrix form
# Every node has a weight with the amount of exchanged emails
#
# For the investigation following properties become a sense:
# Covariance (two ways) and correlation (two ways)
# Because the amount of data growth every year, the calculation take
# more time as the go to the last years!!!
# The result is extracted to a csv file to be read on eg Excel
#
# Autor   : Francisco Rodriguez
# Date    : 05.01.2013
# Version: 1.0
#####

# Libraries for Social Network Analysing

```

```

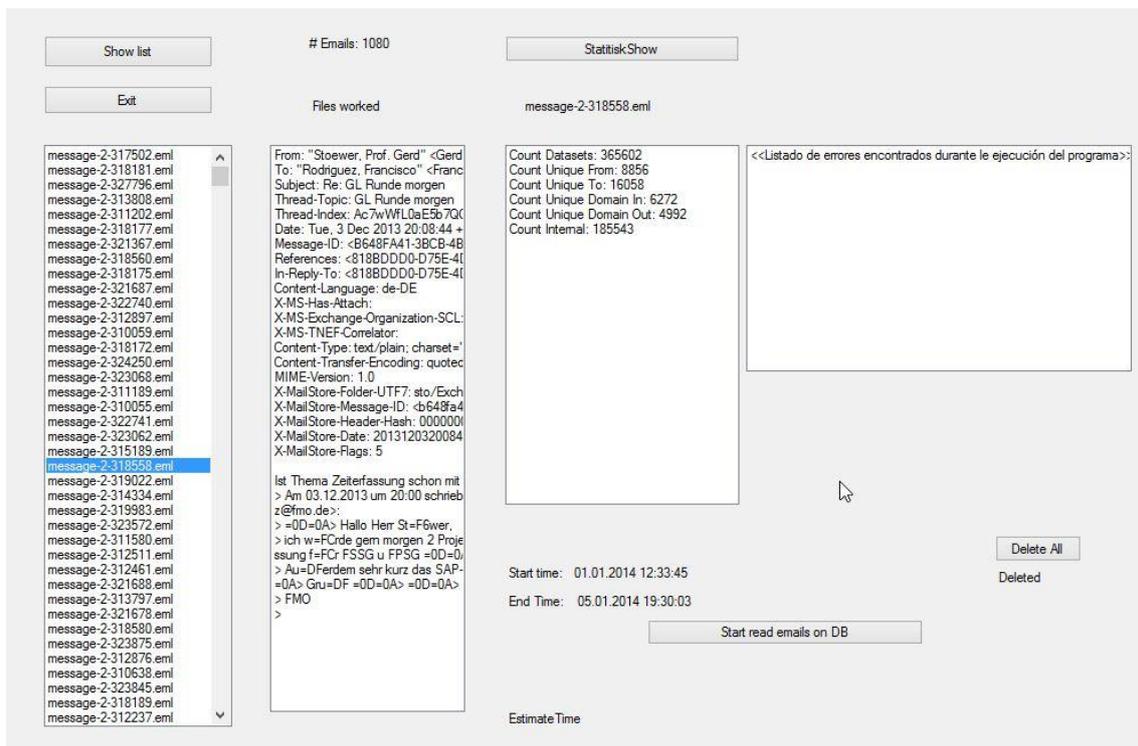
library(sna)
library(igraph)
# Working directory
setwd('W:\\doctorado\\DATA-R')
# Definition of vectors with year and month data
Year<-c(2002:2013)
Month<-c(1:12)
# Before the loop is started I load the first network as reference
# The first value is 0!
g<-read.table(sprintf('%s-%s.csv', 2002, 1), row.names=1, sep=';')
gM<-as.matrix(g)
MatrixF<-c("year", "month", "Covariance", "SCovariance", "Correlation",
"SCorrelation")
# Startting the loop
for (y in Year) {
  for (m in Month){
    g<-read.table(sprintf('%s-%s.csv', y, m), row.names=1, sep=';')
    gMnew<-as.matrix(g)
# Graph covaraince
    gMCv<-gcov(gM, gMnew)
    gMsCv<-gscov(gM, gMnew)
# Graph correlation
    gMCr<-gcor(gM, gMnew)
    gMsCr<-gscor(gM, gMnew)
# new matrix get the old values to be compare with the new data
    gM<-gMnew
# The values are put together to be exported
    VectorT<-c(y, m, gMCv, gMsCv, gMCr, gMsCr)
    MatrixF<-rbind(MatrixF, VectorT)
  }}
# Export of the calculated data

write.csv2(MatrixF, 'DataFinal-Evolution.csv')
#####
# END
#####

```

## B. Apéndice – Programa en Visual Basic

### B.1. Programa para leer los archivos con correos electrónicos en una base de datos



**Gráfica 62 GUI del Programa para extraer información de correos y guardarlos en una base de datos**

```

|
| *****
| VERSION: 3.0
| FECHA : 05.01.2014
| AUTOR: FRANCISCO RODRIGUEZ
| DESCRIPCIÓN: Este programa lee archivos en formato EML que contienen correos
|               electrónicos
|               los analiza y extrae la información necesaria para el análisis
|               y los guarda en una base de Datos de tipo Microsoft SQL 2012
|
| *****
Imports System.Data.SqlClient
Imports System.IO
    
```

```
Public Class Form1
Dim con As New SqlClient.SqlConnection
Dim cmd As New SqlClient.SqlCommand
Dim reader As SqlClient.SqlDataReader
' FORMA DE ESCOGER LA BASE DE DATOS - EN ESTE CASO MICROSOFT SQL 2012
Const CONNECT_STRING As String = "Data
Source=.\SQLEXPRESS;AttachDbFilename=C:\Users\rod\Documents\DATA\
EMAIL-DB.mdf;Integrated Security=True;Connect Timeout=30;User Instance=True"
Dim SN As Integer
Dim Dirs() As String
Dim SourcePfad As String = "W:\Gesendete Objekte" & "\"
Dim WorkedPfad As String = "W:\Gesendete Objekte\read" & "\"

Private Sub Form1_Load(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles MyBase.Load
' COMIENZO DEL PROGRAMA
Count_DB_Datasets()
Me.Refresh()
Try
' MANEJO DE ERRORES
Catch ex As Exception
Dim MsgTxt, MsgTxt2
MsgTxt = ex.Message
MsgTxt2 = ex.InnerException
MsgBox(MsgTxt & MsgTxt2)
End Try
End Sub

Private Sub ExitApp_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles ExitApp.Click
' PARA SALIR DEL PROGRAMA
Application.Exit()
End Sub

Private Sub ShowList_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles ShowList.Click
' LLAMA EL SUB DEL LISTADO DEL CONTENIDO DEL DIRECTORIO A SER LEIDO
Next_Dir()
End Sub

Sub Next_Dir()
' LISTADO DEL CONTENIDO DEL DIRECTORIO A SER LEIDO
Dim sPath As String
sPath = SourcePfad
Dim oDir As New System.IO.DirectoryInfo(sPath)
Dim oFiles As System.IO.FileInfo() = oDir.GetFiles()
Dim oFile As System.IO.FileInfo
Dim DirsNo = 0
If sPath.EndsWith("\") And sPath.Length > 3 Then
sPath = sPath.Substring(0, sPath.Length - 1)

```

```

End If
ListBox3.Items.Clear()
For Each oFile In oFiles
    ListBox3.Items.Add(oFile.Name)
    DirsNo = DirsNo + 1
    ReDim Preserve Dirs(DirsNo)
    Dirs(DirsNo - 1) = oFile.Name
Next
Label5.Text = "# Emails: " & ListBox3.Items.Count
SN = ListBox3.Items.Count
End Sub

Private Sub ListBox3_SelectedIndexChanged(ByVal sender As System.Object, ByVal e As System.EventArgs) Handles ListBox3.SelectedIndexChanged
' PARA EFECTOS DE CONTROL ES POSIBLE SELECCIONAR UN ARCHIVO Y MOSTRAR SU
CONTENIDO
ListBox4.Items.Clear()
Label8.Text = ListBox3.Text
For Each oLine In IO.File.ReadAllLines(SourcePfad & ListBox3.Text)
    ListBox4.Items.Add(oLine)
Next
End Sub

Private Sub StartReadEmails_Click(ByVal sender As System.Object, ByVal e As System.EventArgs) Handles StartReadEmails.Click
' EN ESTE SUB SE LEEN LOS ARCHIVOS, SE EXTRACTA LA INFORMACIÓN Y ES GUARDADA
EN LA ' BASE DE DATOS
Dim EMAILID, EMAILFROM, EMAILTO, EMAILTOTEMP, EMAILCCTEMP, EMAILBCCTEMP As String
Dim EMAILDATE, DOMAINTO, DOMAINFROM, INTERN, FILENAME As String
Dim EMAILCAT As String
Dim YEAR, MONTH, QUARTER, SIZE As Integer
Dim HEADEND As Boolean
Dim InFromField, InToField, InCCField, InBCCField, InMIDField As Boolean
Dim FileFocus As String
Dim A$, EMAILCC, EMAILBCC
Dim ErrorStr, NumberOfTo, NumberOfCC, NumberOfBCC, EndOfEmail
Dim EmailsCounter As Integer = 0
Dim elementOnDB As Boolean
Dim dtStart As Date = Date.Now
Dim tsDuration As TimeSpan, dtEstFinishTime As Date
Label9.Text = DateTime.Now
'Inside the initialising procedure for the process
dtStart = Now()
For j = 1 To 2
    If j = 1 Then
        SourcePfad = "G:\Gesendete Objekte" & "\"
        WorkedPfad = "G:\Gesendete Objekte\read" & "\"
    End If
    If j = 2 Then
        SourcePfad = "G:\Posteingang Archiv" & "\"
        WorkedPfad = "G:\Posteingang Archiv\read" & "\"
    End If

```

```

End If
Next_Dir()
For i = 0 To Dirs.Length - 2
    ' CALCULO ESTIMADO DE DURACIÓN DEL PROCESO - Nota: este valor no es
    confiable!
    tsDuration = Now() - dtStart
    dtEstFinishTime = dtStart.AddSeconds(tsDuration.TotalSeconds *
(Dirs.Length) / (i + 1))
    EstimateTime.Text = dtEstFinishTime
    If (i + 1) Mod 1000 = 0 Then
        Count_DB_Datasets()
    End If
    FileFocus = Dirs(i)
    Label8.Text = FileFocus
    Label7.Text = i
    FILENAME = Label8.Text
    EMAILID = ""
    EMAILFROM = ""
    EMAILDATE = ""
    EMAILTO = ""
    EMAILCC = ""
    EMAILB CC = ""
    EMAILCAT = ""
    HEADEND = False
    InFromField = False
    InToField = False
    InCCField = False
    InBCCField = False
    InMIDField = False
    Me.Refresh()
    elementOnDB = vbFalse
    If FileFocus = SourcePfad & FileFocus Then
        Continue For
    End If
    ' PRUEBA SI EL ELEMENTO SE ENCUENTRA EN LA BASE DE DATOS
    If Test_if_DS_on_DB(FileFocus) Then
        elementOnDB = vbTrue
        File.Move(SourcePfad & FileFocus, WorkedPfad & FileFocus)
        Continue For
    End If
    ' COMIENZO DEL ANALISIS DEL ARCHIVO LEIDO
    For Each oLine In IO.File.ReadLines(SourcePfad & FileFocus)
        SIZE = FileFocus.Length
        A$ = Trim(oLine)
        If A$.Length = 0 Then
            A$ = " "
        End If
    ' PARA NO TENER QUE LEER TODAS LAS LINEAS DEL ARCHIVO
    ' SE TERMINA CUANDO SE LLEGA A ESTE PUNTO
    ' ERRORSTR SOLO PARA EFECTOS DE BUSQUEDA DE ERRORES

```

```
    If Mid(A$, 1, 17) = "X-MailStore-Flags" Then 'Or Mid(A$, 1, 17) =
"Content Language:" Then
        HEADEND = True
        ErrorStr = "1"
    End If
' CORREOS CON ESTE SUBJETO SON IGNORADOS - ESTOS MENSAJES NO SON CORREOS
    If A$ = "Subject: Konfliktnachricht" Then
        ErrorStr = "2"
    End If
' PROBANDO SI To, CC o BCC TIENEN MAS DE UNA LINEA
    If InFromField And (Asc(Mid(A$, 1, 1)) = 9 Or Asc(Mid(A$, 1, 1)) = 32)
Then
        InFromField = True
    Else
        InFromField = False
    End If
    If InToField And (Asc(Mid(A$, 1, 1)) = 9 Or Asc(Mid(A$, 1, 1)) = 32)
Then
        InToField = True
    Else
        InToField = False
    End If
    If InCCField And (Asc(Mid(A$, 1, 1)) = 9 Or Asc(Mid(A$, 1, 1)) = 32)
Then
        InCCField = True
    Else
        InCCField = False
    End If
    If InBCCField And (Asc(Mid(A$, 1, 1)) = 9 Or Asc(Mid(A$, 1, 1)) = 32)
Then
        InBCCField = True
    Else
        InBCCField = False
    End If
    If InMIDField And Asc(Mid(A$, 1, 1)) = 32 Then
        InMIDField = True
    Else
        InMIDField = False
    End If
' EVALUACION DE LA LINEA LEIDA
    If Mid(A$, 1, 5) = "From:" Or InFromField Then
        If InFromField Then
            EMAILFROM = Trim(EMAILFROM & Mid(A$, 2, 100))
        Else
            EMAILFROM = Trim(Mid(A$, InStr(A$, "From:") + 6, 100))
        End If
        EMAILFROM = EMAILFROM.Replace("'", "")
        InFromField = True
    End If
    If Mid(A$, 1, 3) = "To:" Or InToField Then
        If InToField Then
            EMAILTO = Trim(EMAILTO & Mid(A$, 2, 100))
```

```

        Else
            EMAILTO = Trim(Mid(A$, InStr(A$, "To:") + 4, 100))
        End If
        EMAILTO = EMAILTO.Replace("'", "")
        InToField = True
    End If
    If Mid(A$, 1, 3) = "CC:" Or InCCField Then
        If InCCField Then
            EMAILCC = Trim(EMAILCC & Mid(A$, 2, 100))
        Else
            EMAILCC = Trim(Mid(A$, InStr(A$, "CC:") + 4, 100))
        End If
        EMAILCC = EMAILCC.Replace("'", "")
        InCCField = True
    End If
    If Mid(A$, 1, 4) = "BCC:" Or InBCCField Then
        If InBCCField Then
            EMAILBCC = Trim(EMAILBCC & Mid(A$, 2, 100))
        Else
            EMAILBCC = Trim(Mid(A$, InStr(A$, "BCC:") + 5, 100))
        End If
        EMAILBCC = EMAILBCC.Replace("'", "")
        InBCCField = True
    End If
    If Mid(A$, 1, 11) = "Message-ID:" Or InMIDField Then
        If InMIDField Then
            EMAILID = Trim(EMAILID & Mid(A$, 2, 100))
        Else
            EMAILID = Trim(Mid(A$, InStr(A$, "Message-ID:") + 12, 100))
        End If
        EMAILID = EMAILID.Replace("'", "")
        InMIDField = True
    End If
    If Mid(A$, 1, 5) = "Date:" Then
        EMAILDATE = DateOnField(Mid(A$, InStr(A$, "Date:") + 6, 100))
        YEAR = CInt(Mid(EMAILDATE, 7, 4))
        MONTH = CInt(Mid(EMAILDATE, 4, 2))
    End If
    If InStr(A$, "X-MailStore-Date") > 0 Then
        EMAILDATE = Trim(Mid(A$, InStr(A$, "X-MailStore-Date") + 17, 30))
        YEAR = CInt(Mid(EMAILDATE, 1, 4))
        MONTH = CInt(Mid(EMAILDATE, 5, 2))
        QUARTER = CInt(Mid(EMAILDATE, 7, 2))
        EMAILDATE = Mid(EMAILDATE, 7, 2) & "-" & Mid(EMAILDATE, 5, 2) & "-" &
Mid(EMAILDATE,
1, 4)
        HEADEND = True
    End If
    If HEADEND Then
        Exit For
    End If

```

```

    End If
Next
If elementOnDB = False Then
    If EMAILFROM = "" Then
        EMAILFROM = "<unknown>"
    End If
    If EMAILDATE = "" Then
        ErrorStr = "Email fecha vacia?"
        'Next ' Normalmente no es posible!!! (pero hubo un caso)
    Else
' INFORMACIÓN ES GUARDADA EN LA BASE DE DATOS
' BUCLE/LASO POR MAS DE UN CORREO EN UN ARCHIVO
        NumberOfTo = CountOfSymbols(EMAILTO, "<")
        NumberOfCC = CountOfSymbols(EMAILCC, "<")
        NumberOfBCC = CountOfSymbols(EMAILBCC, "<")
        EMAILFROM = ExtractEmail(EMAILFROM)
        INTERN = "NO"
        ' PARA LEER MEJOR
        EMAILID = EMAILID.Replace(">", "")
        EMAILID = EMAILID.Replace("<", "")
        EMAILID = EMAILID.Replace("@", "")
        If NumberOfTo > 0 Then
            For Loops = 1 To NumberOfTo
                SN = SN + 1
                EmailsCounter = EmailsCounter + 1
                EMAILTOTEMP = ExtractEmail(EMAILTO)
                EndOfEmail = InStr(EMAILTO, ">") + 1
                EMAILTO = Mid(EMAILTO, EndOfEmail)
                DOMAINTO = ExtractDomain(EMAILTOTEMP)
                DOMAINFROM = ExtractDomain(EMAILFROM)
                If UCase(DOMAINTO) = UCase(DOMAINFROM) Then
' FMO.DE ES EL DOMINIO DE LA COMPAÑIA EN INVESTIGACIÓN
                    If LCase(DOMAINTO) = "fmo.de" Then
                        INTERN = "YES"
                    Else
                        INTERN = "NO"
                    End If
                End If
                If TestDatasetOnDB(EMAILID, EMAILFROM, EMAILTOTEMP, EMAILDATE,
FILENAME) = 0
Then
                    DataInput(SN, EMAILID, EMAILFROM, EMAILTOTEMP, EMAILDATE, "TO",
FILENAME
, DOMAINTO, DOMAINFROM, INTERN, YEAR, MONTH, QUARTER, SIZE)
                End If
            Next
        End If
        If NumberOfCC > 0 Then
            For Loops = 1 To NumberOfCC
                SN = SN + 1
                EmailsCounter = EmailsCounter + 1

```

```

EMAILCCTEMP = ExtractEmail(EMAILCC)
EndOfEmail = InStr(EMAILCC, ">") + 1
EMAILCC = Mid(EMAILCC, EndOfEmail)
DOMAINTO = ExtractDomain(EMAILCCTEMP)
DOMAINFROM = ExtractDomain(EMAILFROM)
If UCase(DOMAINTO) = UCase(DOMAINFROM) Then
    If LCase(DOMAINTO) = "fmo.de" Then
        INTERN = "YES"
    Else
        INTERN = "NO"
    End If
End If
If TestDatasetOnDB(EMAILID, EMAILFROM, EMAILCCTEMP, EMAILDATE,
FILENAME) = 0
Then
    DataInput(SN, EMAILID, EMAILFROM, EMAILCCTEMP, EMAILDATE, "CC",
FILENAME
, DOMAINTO, DOMAINFROM, INTERN, YEAR, MONTH, QUARTER, SIZE)
    End If
Next
End If
If NumberOfBCC > 0 Then
For Loops = 1 To NumberOfBCC
    SN = SN + 1
    EmailsCounter = EmailsCounter + 1
    EMAILBCCTEMP = ExtractEmail(EMAILBCC)
    EndOfEmail = InStr(EMAILBCC, ">") + 1
    EMAILBCC = Mid(EMAILBCC, EndOfEmail)
    DOMAINTO = ExtractDomain(EMAILBCCTEMP)
    DOMAINFROM = ExtractDomain(EMAILFROM)
    If UCase(DOMAINTO) = UCase(DOMAINFROM) Then
        If LCase(DOMAINTO) = "fmo.de" Then
            INTERN = "YES"
        Else
            INTERN = "NO"
        End If
    End If
End If
If TestDatasetOnDB(EMAILID, EMAILFROM, EMAILBCCTEMP, EMAILDATE,
FILENAME) =
0 Then
    DataInput(SN, EMAILID, EMAILFROM, EMAILBCCTEMP, EMAILDATE,
"BCC",
FILENAME, DOMAINTO, DOMAINFROM, INTERN, YEAR, MONTH, QUARTER, SIZE)
    End If
Next
End If
End If
File.Move(SourcePfad & FileFocus, WorkedPfad & FileFocus)
Next
Next
Label12.Text = DateTime.Now

```

```

Count_DB_Datasets()
Me.Refresh()
End Sub

Sub DataInput(ByRef SNfield As String, ByRef idfield As String, ByRef
fromfield As String, ByRef tofield
As String, ByRef datumfield As String, ByRef categoryfield As String, ByRef
filenamefield As String,
ByRef dinfield As String, ByRef doutfield As String, ByVal internfield As
String, ByVal yearfield As
Integer, ByVal monthfield As Integer, ByVal quarterfield As Integer, ByVal
sizefield As Integer)
' ESTE SUB GUARDA LOS ELEMNETOS ENCONTRADOS EN LA BASE DE DATOS
Dim Count As Integer = 0
Dim inputString As String
Dim con As SqlConnection = New SqlConnection(CONNECT_STRING)
Dim cmd As SqlCommand = New SqlCommand("INSERT INTO EMAILS (SN, ID, [FROM],
[TO], DATETIME, CATEGORY
, FILENAME, DOMAINTO, DOMAINFROM, INTERNAL, YEAROFEMAIL, MONTHOFEMAIL,
QUARTEROFEMAIL, SIZEOFEMAIL) " &
-
"VALUES(" & SNfield & ",'" & idfield.ToString & "', '" _
& fromfield.ToString & "', '" _
& tofield.ToString & "', '" _
& datumfield.ToString & "', '" _
& categoryfield.ToString & "', '" _
& filenamefield.ToString & "', '" _
& dinfield.ToString & "', '" _
& doutfield.ToString & "', '" _
& internfield.ToString & "', " _
& yearfield & ", " _
& monthfield & ", " _
& quarterfield & ", " _
& sizefield & ")")
cmd.CommandType = CommandType.Text
cmd.Connection = con
Try
    con.Open()
    Count = cmd.ExecuteNonQuery()
' MANEJO DE ERRORES
Catch ex As Exception
    ActuallyData.Items.Add(ex.Message & "-" & filenamefield)
    inputString = ex.Message & "-" & filenamefield & " - " & WorkedPfad &
Chr(10)
    inputString = inputString & "VALUES: " & SNfield & ",'" & idfield.ToString &
"', '" _
& fromfield.ToString & "', '" _
& tofield.ToString & "', '" _
& datumfield.ToString & "', '" _
& categoryfield.ToString & "', '" _
& filenamefield.ToString & "', '" _
& dinfield.ToString & "', '" _

```

```

    & doutfield.ToString & ", " _
    & internfield.ToString & ", " _
    & yearfield & ", " _
    & monthfield & ", " _
    & quarterfield & ", " _
    & sizefield & Chr(10)
' ERRORES SON GUARDADOS EN UN ARCHIVO
    My.Computer.FileSystem.WriteAllText("G:\log.txt", inputString, True)
End Try
If con.State = ConnectionState.Open Then
    con.Close()
End If
End Sub

Private Sub DELETEDB_Click(sender As System.Object, e As System.EventArgs) Handles DELETEDB.Click
' SUB PARA BORRAR EL CONTENIDO DE LA BASE DE DATOS
' NECESARIO AL PRINCIPIO CUANDO NO TODOS LOS ERRORES ESTABAN DEFINIDOS Y ACLARADOS
Dim Count As Integer = 0
Dim con As SqlConnection = New SqlConnection(CONNECT_STRING)
Dim cmd As SqlCommand = New SqlCommand("DELETE FROM EMAILS")
cmd.CommandType = CommandType.Text
cmd.Connection = con
Try
    con.Open()
    Count = cmd.ExecuteNonQuery()
    Catch ex As Exception
        MessageBox.Show(ex.Message)
End Try
Deleted.Text = Count
Me.Refresh()
If con.State = ConnectionState.Open Then
    con.Close()
End If
Count_DB_Datasets()
End Sub

Private Sub StatitiskShow_Click(sender As System.Object, e As System.EventArgs) Handles StatitiskShow.Click
' MOSTRAR LAS ESTADISTICAS EN CUALQUIER MOMENTO
Count_DB_Datasets()
End Sub

Function Test_if_DS_on_DB(ByVal filename As String) As Boolean
' PRUEBA DE SI UN ELEMENTO SE ENCUENTRA EN LA BASE DE DATOS BASADO SOLO EN EL NOMBRE DEL ARCHIVO
Dim CountDatasets As Integer = 0
Dim con As SqlConnection = New SqlConnection(CONNECT_STRING)
Dim cmd As SqlCommand = New SqlCommand("SELECT COUNT(*) FROM EMAILS WHERE FILENAME='" & filename &

```

```
""")
cmd.CommandType = CommandType.Text
cmd.Connection = con
Try
    con.Open()
    CountDatasets = cmd.ExecuteScalar()
    Catch ex As Exception
        MessageBox.Show(ex.Message)
End Try
If CountDatasets > 0 Then
    Return vbTrue
Else
    Return vbFalse
End If
End Function

Function DateOnField(ByVal TextCell As String) As String
' CONVERSIÓN DE UNA FECHA DE TEXTO A NUMERICO
Dim Y, M, D As String
Dim tempcell, MTemp As String
If TextCell.Length < 10 Then
    Return TextCell
End If
tempcell = Mid$(TextCell, InStr(TextCell, " ") + 1)
D = Mid(tempcell, 1, InStr(tempcell, " ") - 1)
MTemp = Mid(tempcell, InStr(tempcell, " ") + 1, 3)
M = ""
Select Case MTemp
    Case "Jan"
        M = "01"
    Case "Feb"
        M = "02"
    Case "Mar"
        M = "03"
    Case "Apr"
        M = "04"
    Case "May"
        M = "05"
    Case "Jun"
        M = "06"
    Case "Jul"
        M = "07"
    Case "Aug"
        M = "08"
    Case "Sep"
        M = "09"
    Case "Oct"
        M = "10"
    Case "Nov"
        M = "11"
    Case "Dec"

```

```

        M = "12"
    End Select
    Y = Mid(tempcell, InStr(tempcell, " ") + 5, 4)
    DateOnField = D & "-" & M & "-" & Y
End Function

```

```

Function CountOfSymbols(ByVal StringToControl As String, ByVal CharToLook As String) As Integer

```

```

    ' FUNCIÓN PARA CONTAR ELEMENTOS EN UN ELEMENTO PARA ELIMINARLOS ("LEER MEJOR LOS CORREOS")

```

```

    Dim Counter1
    Dim Counter2
    Counter2 = 0
    For Counter1 = 1 To Len(StringToControl)
        If Mid(StringToControl, Counter1, 1) = CharToLook Then
            Counter2 = Counter2 + 1
        End If
    Next
    CountOfSymbols = Counter2
End Function

```

```

Function ExtractEmail(ByVal stringWithEmail As String) As String

```

```

    ' FUNCIÓN PARA EXTRAER LA DIRECCIÓN DE CORREO QUE ESTÉ EN UN ELEMENTO
    Dim StartOfEmail, EndOfEmail, LenghtOfEmail
    ' ELIMINAR UN CARACTER SIN SIGNIFICADO
    stringWithEmail = stringWithEmail.Replace("'", "")
    StartOfEmail = InStr(stringWithEmail, "<") + 1
    EndOfEmail = InStr(stringWithEmail, ">")
    LenghtOfEmail = EndOfEmail - StartOfEmail
    If LenghtOfEmail > 1 Then
        ExtractEmail = LCase(Mid(stringWithEmail, StartOfEmail, LenghtOfEmail))
    Else
        ExtractEmail = LCase(stringWithEmail)
    End If
End Function

```

```

Function TestDatasetOnDB(ByRef Idtemp As String, ByVal Fromtemp As String, ByVal Totemp As String, ByVal

```

```

Datetemp As String, ByVal Filenametemp As String) As Integer

```

```

    ' FUNCIÓN PARA COMPROBAR SI UN ELEMENTO YA SE ENCUENTRA EN LA BASE DE DATOS -
    ' UTIL PARA EL CASO DE REPETIR LA LLAMADA DEL PROGRAM VARIAS VECES SIN
    ' DUPLICAR LOS VALORES

```

```

    ' LEIDOS - BASADA EN LOS ELEMENTOS LEIDOS
    Dim Count As Integer = 0
    Try
        con.ConnectionString = CONNECT_STRING
        cmd.Connection = con
        cmd.CommandText = "select COUNT(*) from EMAILS WHERE ID='" & Idtemp & _
            "' AND [FROM]='" & Fromtemp & _
            "' AND [TO]='" & Totemp & _
            "' AND DATETIME='" & Datetemp & _

```

```

    "' AND FILENAME='" & Filenametemp & "'"
    con.Open()
    Count = cmd.ExecuteScalar()
    con.Close()
    Catch ex As Exception
    Dim MsgTxt, MsgTxt2
    MsgTxt = ex.Message
    MsgTxt2 = ex.InnerException
    MsgBox(MsgTxt & MsgTxt2)
End Try
Return Count
End Function

Function ExtractDomain(ByVal OriginalTxt As String) As String
' FUNCIÓN QUE EXTRAER EL DOMINIO DEL CORREO ANALISADO
Dim StartOfTxt, EndOfTxt, LenghtOfTxt
StartOfTxt = InStr(OriginalTxt, "@") + 1
EndOfTxt = Len(OriginalTxt)
LenghtOfTxt = EndOfTxt - StartOfTxt
If LenghtOfTxt > 1 Then
    ExtractDomain = UCase(Mid(OriginalTxt, StartOfTxt, LenghtOfTxt + 1))
Else
    ExtractDomain = UCase(OriginalTxt)
End If
End Function

Function Count_DB_Datasets()
' FUNCIÓN PARA MOSTRAR LA ESTADISTICA DE LOS DATOS EN LA BASE DE DATOS
Dim CountDatasets As Integer = 0
Dim CountUniqueFrom As Integer = 0
Dim CountUniqueTo As Integer = 0
Dim CountUniqueDOMAINTO As Integer = 0
Dim CountUniqueDOMAINFROM As Integer = 0
Dim CountInternal As Integer = 0
ListTab.Items.Clear()
Dim con As SqlConnection = New SqlConnection(CONNECT_STRING)
Dim cmd As SqlCommand = New SqlCommand("SELECT COUNT(ID) FROM EMAILS")
cmd.CommandType = CommandType.Text
cmd.Connection = con
Try
    con.Open()
    CountDatasets = cmd.ExecuteScalar()
    Catch ex As Exception
    MessageBox.Show(ex.Message)
End Try
ListTab.Items.Add("Count Datasets: " & CountDatasets)
cmd.CommandText = "SELECT COUNT(DISTINCT [FROM]) FROM EMAILS"
cmd.CommandType = CommandType.Text
cmd.Connection = con
Try
    CountUniqueFrom = cmd.ExecuteScalar()

```

```
    Catch ex As Exception
        MessageBox.Show(ex.Message)
End Try
ListTab.Items.Add("Count Unique From: " & CountUniqueFrom)
cmd.CommandText = "SELECT COUNT(DISTINCT [TO]) FROM EMAILS"
cmd.CommandType = CommandType.Text
cmd.Connection = con
Try
    CountUniqueTo = cmd.ExecuteScalar()
    Catch ex As Exception
        MessageBox.Show(ex.Message)
End Try
ListTab.Items.Add("Count Unique To: " & CountUniqueTo)
cmd.CommandText = "SELECT COUNT(DISTINCT DOMAINTO) FROM EMAILS"
cmd.CommandType = CommandType.Text
cmd.Connection = con
Try
    CountUniqueDOMAINTO = cmd.ExecuteScalar()
    Catch ex As Exception
        MessageBox.Show(ex.Message)
End Try
ListTab.Items.Add("Count Unique Domain In: " & CountUniqueDOMAINTO)
cmd.CommandText = "SELECT COUNT(DISTINCT DOMAINFROM) FROM EMAILS"
cmd.CommandType = CommandType.Text
cmd.Connection = con
Try
    CountUniqueDOMAINFROM = cmd.ExecuteScalar()
    Catch ex As Exception
        MessageBox.Show(ex.Message)
End Try
ListTab.Items.Add("Count Unique Domain Out: " & CountUniqueDOMAINFROM)
cmd.CommandText = "SELECT COUNT(*) FROM EMAILS WHERE INTERNAL='YES'"
cmd.CommandType = CommandType.Text
cmd.Connection = con
Try
    CountInternal = cmd.ExecuteScalar()
    Catch ex As Exception
        MessageBox.Show(ex.Message)
End Try
ListTab.Items.Add("Count Internal: " & CountInternal)
Me.Refresh()
If con.State = ConnectionState.Open Then
    con.Close()
End If
Return vbTrue
End Function
End Class
```

## C. Apéndice – Programa en VBA (Excel)

### C.1. Macro para exportar datos calculados

```

Sub Export()
'
' Macro para exportar los datos calculados con ayuda de una
' Tabla dinámica en archivos csv que puedan ser leídos en R
' Autor: Francisco Rodríguez
' Version: 3.0
' Fecha: 5.1.2014
'
    Dim s As String
    Dim i As Integer
    Dim e As Integer
    Dim y As Integer ' para escojer el año
    Dim m As Integer ' para escojer el mes

    Sheets("DATA").Select

    ' Importante para no exportar sumas generales!!!
    ActiveSheet.PivotTables("PivotTable1").PivotFields("TONEW").Subtotals = _
    Array(False, False, False, False, False, False, False, False, False, False, _
    False, False)

    With ActiveSheet.PivotTables("PivotTable1")
        .ColumnGrand = False
        .RowGrand = False
    End With

    ' Análisis y extracción de los datos por año y mes
    For y = 2002 To 2013
        For m = 1 To 12
            ActiveSheet.PivotTables("PivotTable1").PivotFields("YEAROFEMAIL"). _
            ClearAllFilters
            ActiveSheet.PivotTables("PivotTable1").PivotFields("YEAROFEMAIL"). _
            CurrentPage = y
            ActiveSheet.PivotTables("PivotTable1").PivotFields("MONTHOFEMAIL"). _
            ClearAllFilters
            ActiveSheet.PivotTables("PivotTable1").PivotFields("MONTHOFEMAIL"). _
            CurrentPage = m

            s = "" ' Iniliciación de la variable que contiene los datos

            Open "W:\Doctorado\" & y & "-" & m & ".csv" For Output As #1

            For i = 5 To ActiveSheet.UsedRange.Rows.Count
                For e = 1 To ActiveSheet.UsedRange.Columns.Count
                    If e = cc Then ' la última columna no termina en ;
                        If Cells(i, e).Value > 0 Then
                            s = s & Cells(i, e).Value
                        Else
                            s = s & "0"
                        End If
                    Else
                        If i=5 and e=1 Then
                            ' el primer valor es ignorado para poder usar función de
                            ' importación en R y conversión en matriz

```

```
        Else
            If Cells(i, e).Value > 0 Then
                s = s & Cells(i, e).Value & ";"
            Else
                s = s & "0;"
            End If
        End If
    End If
Next e

    Close #1
Next m
Next y MsgBox "Exporte finalizado!"

End Sub
```

# D. Apéndice – Extracto de la Base de Datos (normalizados) 2002-01 a 2005-09

FECHA	COMPONENTE GRADO	DENSIDAD	TRANSMISIVIDAD/CONEXIONE	DIAMETRO	ARISTAS	VERTICES	INGRESOS	COSTOS	BALANZA	COVARIANZA	CORRELACION	P.A.X.	MOV
2002-01	0.05069124	0.00475436	0.19444444	0	0.05882333	0	0.23495657	0.84947966	0.41980976	0.02318397	0.04301848	0.16757975	0.3821594
2002-02	0.02958267	1	0.24747475	0.00625109	0.0310559	0.03603604	0.25281758	0.84487117	0.42824468	0.11106471	0.12665458	0.23880035	0.468194
2002-03	0.13824885	0.03982182	0.14583333	0.01863678	0.04813665	0.09009009	0.40941362	0.81135939	0.55176488	0	0.04457568	0.46674584	0.76456152
2002-04	0.06451613	0.09190743	0.19444444	0.09803523	0.04532795	0.025252	0.65348769	0.85198979	0.55285585	0.01264481	0.02512909	0.50883713	0.67170363
2002-05	0.0875576	0.03592182	0.23445222	0.13241782	0.1372549	0.063006306	0.60546666	0.83167288	0.66218379	0.06434171	0.09594566	0.72347585	0.86765781
2002-06	0.09577419	0.02060222	0.48632882	0.01130474	0.1372549	0.05416149	0.9094555	0.80430628	0.84975646	0.01858925	0.03599146	0.77365081	0.75578845
2002-07	0.10599078	0.14262607	0.28282828	0.01022604	0.02828282	0.05405405	0.72757459	0.79857801	0.71805009	0.02621772	0.0484059	0.93344127	0.7280039
2002-08	0.0875576	0.14262607	0.50412428	0.0081742	0.01960784	0.04504505	0.66360684	0.82035172	0.6926759	0.02903977	0.05513371	0.88339974	0.70972459
2002-09	0.10599078	0.01601743	0.54453998	0.26756757	0.01022604	0.05405405	0.72342498	0.84313766	0.7545061	0.00819514	0.00543861	0.91326974	0.74677066
2002-10	0.03225806	0.00369783	0.71423486	0.31531532	0.00208703	0.05882333	0.64754758	0.81245918	0.67434748	0.02035963	0.03514781	0.85189513	0.58273702
2002-11	0.0590783	0	0.52483358	0.13207547	0.00285227	0.1960794	0.00465839	0.71957385	0.36438617	0.11574897	0.23959349	0.33602151	0.48622959
2002-12	0.07373272	0.01320655	0.56441424	0.16969597	0.0081742	0.09803922	0.41973777	0.62855963	0.36009826	0.01653066	0.02394126	0.28493909	0.3009927
2003-01	0.12903226	0.02430002	0.3300992	0.1851852	0.01598577	0.03921569	0.70392768	0.88383605	0.42647797	0.00546364	0	0.31980947	0.3821594
2003-02	0.12903226	0.03593556	0.4375787	0.48373984	0.01721799	0.05882333	0.5490962	0.8108108	0.19709554	0.91389529	0.45035591	0.07099885	0.12946977
2003-03	0.13824885	0.0701532	0.37022133	0.15900546	0.03165327	0.17847099	0.08540373	0.1621613	0.91389529	0.91389529	0.14649012	0.33583783	0.42266876
2003-04	0.12903226	0.03328051	0.40833333	0.01598577	0.09803922	0.05434783	0.07657658	0.41118643	0.90851024	0.59479711	0.01478335	0.02235072	0.58364621
2003-05	0.29953917	0.11466329	0.1226684	0.39381153	0.0396883	0.25490196	0.18444059	0.27477477	0.41281331	0.9476908	0.6297065	0.03897416	0.11942768
2003-06	0.21658966	0.07290016	0.3404428	0.2251462	0.05154921	0.15868275	0.09782609	0.17567568	0.89853133	0.64968397	0.1590585	0.48960414	0.79358849
2003-07	0.16589862	0.03388273	0.37451986	0.07909605	0.03498774	0.09803922	0.08229814	0.13153154	0.50355728	0.88986181	0.64247781	0.16779397	0.37150016
2003-08	0.12903226	0.2482831	0.3397743	0.15909091	0.01982678	0.01960794	0.0621118	0.09009009	0.53886473	0.93852615	0.7102963	0.01964481	0.03084433
2003-09	0.12442396	0.03222398	0.49275593	0.09909891	0.01721799	0.07843137	0.07142857	0.08108108	0.54009876	0.90153559	0.67843768	0.02661067	0.05325318
2003-10	0.07834101	0.0306592	0.73629172	0.36458333	0.01022604	0.17547059	0.05405405	0.42907915	0.91230742	0.68028835	0.04097577	0.07913402	0.90271499
2003-11	0.0539954	0.00645219	0.60891337	0.11382114	0.00448711	0.01960794	0.02018634	0.02702703	0.81711706	0.38590173	0.03427421	0.0656316	0.35556662
2003-12	0.11961567	0.02271527	0.477448	0.38888889	0.01355569	0.03215569	0.04347826	0.06756757	0.25122198	0.90416948	0.47951486	0.112453	0.2807891
2004-01	0.1843318	0.0426096	0.92663377	0.23123123	0.03666214	0.01960794	0.07919255	0.13963964	0.27941703	0.85198031	0.45299598	0.01715054	0.02208948
2004-02	0.22580645	0.05124142	0.33561071	0.26376812	0.04956984	0.09803922	0.09782609	0.17117117	0.04338626	0.90178984	0.33261661	0.02320208	0.05072865
2004-03	0.62678281	0.1251981	0.30742368	0.36897436	0.08048976	0.19807843	0.18187702	0.23423423	0.27281207	0.89052745	0.48248963	0.01719931	0.0244008
2004-04	0.19354839	0.0552404	0.27585481	0.52415459	0.04761905	0.47058824	0.11024845	0.16666667	0.31335909	0.84127678	0.55562567	0.01993288	0.03529812
2004-05	0.02304147	0.03856313	0.68633943	0	0.01474834	0.50980392	0.03881988	0.07207207	0.4451084	0.94593289	0.65165239	0.01767641	0.04539854
2004-06	0.29953917	0.07151537	0.3524234	0.22653722	0.09562942	0.47058824	0.1242236	0.26126126	0.45732108	0.92393555	0.6453733	0.02385799	0.04534281
2004-07	0.15688203	0.0687406	0.48412735	0.5480226	0.03165327	0.3915686	0.09161491	0.12612613	0.49088242	0.88190029	0.62878248	0.02184797	0.03881349
2004-08	0.11981567	0.4807184	0.67763903	0.52238806	0.01721799	0.37254902	0.06055901	0.08108108	0.51091663	0.95091758	0.90981758	0.0306311	0.8813884
2004-09	0.16894009	0.07078711	0.43657794	0.58947368	0.03840134	0.35584118	0.0947205	0.14444444	0.54151069	0.92338332	0.69865731	0.8357709	0.91162676
2004-10	0.31336406	0.12466984	0.2273573	0.25490196	0.1124213	0.52841176	0.18012422	0.28828829	0.50426067	0.89774948	0.65012143	0.02037781	0.0366311
2004-11	0.29953917	0.07199043	0.33587251	0.11330836	0.11330836	0.56862745	0.1870497	0.29279279	0.1647548	0.80330713	0.32988538	0.03159105	0.05812522
2004-12	0.2718894	0.08716323	0.4788759	0.26866687	0.10675154	0.37254902	0.15683323	0.27927928	0.23552228	0.82337717	0.39697676	0.02939177	0.05736364
2005-01	0.37788018	0.13734812	0.48182747	0.33243848	0.13336116	0.11764706	0.20341615	0.14981982	0.13193933	0.8198306	0.31991554	0.04767477	0.18592445
2005-02	0.3737189	0.1557427	0.17885377	0.5008726	0.16971025	0.43137255	0.23360248	0.36936937	0.1621714	0.92216452	0.43339647	0.01966644	0.03357963
2005-03	0.28571429	0.08283714	0.21939037	0.7805489	0.49019608	0.11490683	0.23972973	0.270767375	0.91517637	0.50286472	0.05567232	0.13178378	0.43065111
2005-04	0.37788018	0.09984152	0.12669905	0.19592875	0.13333333	0.15983979	0.31305418	0.31981982	0.32005418	0.81672613	0.51672939	0.0623761	0.51070161
2005-05	0.2903258	0.1198722	0.88953714	0.67884478	0.0890306	0.33333333	0.18372671	0.24774775	0.50463721	0.88827784	0.64200604	0.02912651	0.05863067
2005-06	0.31336406	0.10776545	0.20420492	0.09591607	0.10396883	0.50860392	0.13975155	0.27477477	0.11996771	0.90853065	0.07061339	0.03643398	0.06808613
2005-07	0.11981567	0.1056524	0.1326108	0.69292929	0.16971025	0.35894118	0.36992929	0.55327814	0.86416629	0.65456629	0.0313331	0.06806743	0.93486451
2005-08	0.11981567	0.1056524	0.3204796	0.04653371	0.08049976	0.56862745	0.13664596	0.23423423	0.58889146	0.8919137	0.7039196	0.02004187	0.03298377
2005-09	0.4239631	0.24775489	0.22000166	0.49371981	0.21809454	0.29411785	0.30900621	0.4272793	0.62804919	0.90031354	0.73862942	0.02221958	0.04018372

## E. Apéndice - Ponderación de las conexiones de la red neuronal

Fuente	Objetivo	Ponderación	Fuente	Objetivo	Ponderación	Fuente	Objetivo	Ponderación
1	4	0,480417574	4	37	-0,162228801	6	37	0,168896327
1	5	-0,047736205	4	36	0,10400799	6	36	0,43451066
1	6	-0,282564953	4	35	-0,02966235	6	35	-0,136812997
1	7	-0,226375264	4	34	-0,082323407	6	34	-0,206704393
1	8	0,040587055	4	33	0,176906202	6	33	0,474525733
1	9	0,104556508	4	32	-0,258755105	6	32	-0,005572866
1	10	-0,071011376	4	31	0,411896912	6	31	-0,141309431
1	11	0,190144423	4	30	0,078306525	6	30	-0,362792044
1	12	-0,179760713	4	29	-0,141822874	6	29	-0,246183574
1	13	0,331096597	4	28	-0,175702063	6	28	-0,062193141
1	14	-0,10339561	4	27	0,267588483	6	27	-0,09590396
1	15	0,180522656	4	26	-0,247070068	6	26	0,093220796
1	16	-0,331476351	4	25	-0,065946362	6	25	-0,411278634
1	17	0,220939435	4	24	0,021442741	6	24	0,377747602
1	18	0,162887905	4	23	0,38989794	6	23	-0,09365635
1	19	0,257671173	4	22	0,379651997	6	22	-0,023447988
1	20	0,454131118	4	21	-0,399905817	6	21	-0,337319753
2	4	0,138132454	5	37	-0,426819792	7	37	-0,121621749
2	5	-0,478970361	5	36	-0,073046587	7	36	0,197281329
2	6	-0,422807614	5	35	0,166750343	7	35	0,073923163
2	7	0,34994923	5	34	-0,290211711	7	34	-0,216375413
2	8	-0,082646834	5	33	0,199422337	7	33	0,155125481
2	9	0,122365353	5	32	-0,066327462	7	32	-0,143611652
2	10	-0,257880761	5	31	0,022603331	7	31	-0,411014576
2	11	-0,100924349	5	30	-0,096718651	7	30	-0,428876122
2	12	0,017179602	5	29	-0,188617474	7	29	0,372301434
2	13	-0,371268446	5	28	0,257604482	7	28	0,271860479
2	14	0,216379321	5	27	0,114895122	7	27	-0,447800126
2	15	0,085739714	5	26	0,446856048	7	26	-0,435858012
2	16	-0,314838229	5	25	0,469270195	7	25	-0,030318153
2	17	-0,074198174	5	24	0,124595112	7	24	0,453867764
2	18	-0,320336179	5	23	0,052204281	7	23	0,301344163
2	19	-0,057622672	5	22	-0,217964615	7	22	-0,397763493
2	20	0,025146383	5	21	0,375497799	7	21	0,164449069

## Apéndice - Ponderación de las conexiones de la red neuronal

Fuente	Objetivo	Ponderación	Fuente	Objetivo	Ponderación	Fuente	Objetivo	Ponderación
8	37	0,471305745	11	37	-0,196241559	14	37	0,405135171
8	36	0,267134318	11	36	-0,034671938	14	36	0,42962704
8	35	-0,254448953	11	35	-0,463817344	14	35	-0,418235746
8	34	0,378265065	11	34	0,138185985	14	34	0,257984216
8	33	0,266192653	11	33	0,25638722	14	33	0,424958712
8	32	0,212484664	11	32	-0,185572805	14	32	0,231141527
8	31	0,321955481	11	31	0,387246882	14	31	-0,401317599
8	30	0,072531521	11	30	0,156340368	14	30	0,190980628
8	29	-0,049413772	11	29	-0,014700537	14	29	-0,186399604
8	28	-0,494852673	11	28	-0,110242441	14	28	-0,501652377
8	27	-0,445649996	11	27	0,063600455	14	27	0,020212124
8	26	0,003797564	11	26	-0,00372334	14	26	-0,375769344
8	25	0,487082054	11	25	0,299143909	14	25	0,426100678
8	24	0,251826392	11	24	-0,366005603	14	24	0,203764497
8	23	-0,323719981	11	23	-0,141549932	14	23	0,158419926
8	22	0,322925228	11	22	0,299136976	14	22	0,296215237
8	21	0,346524858	11	21	0,433233871	14	21	-0,257261114
9	37	-0,480735436	12	37	0,461120122	15	37	-0,134062383
9	36	0,071262523	12	36	-0,076699753	15	36	0,308095614
9	35	0,344502212	12	35	0,397119028	15	35	0,405378997
9	34	0,067360915	12	34	-0,287746369	15	34	0,101840274
9	33	0,132345023	12	33	-0,136908167	15	33	-0,037702434
9	32	-0,030822265	12	32	0,088549649	15	32	0,314059461
9	31	-0,496764206	12	31	-0,408212701	15	31	0,276183693
9	30	-0,12075966	12	30	-0,423496088	15	30	0,192345204
9	29	0,002934845	12	29	-0,140435409	15	29	0,302831943
9	28	0,053468728	12	28	0,483070803	15	28	-0,473292513
9	27	0,372230914	12	27	-0,218991869	15	27	-0,15146364
9	26	0,280880659	12	26	-0,423995454	15	26	0,407231369
9	25	0,221396763	12	25	-0,210472951	15	25	-0,106038004
9	24	0,022017038	12	24	-0,348610921	15	24	-0,245546753
9	23	-0,26210221	12	23	0,010568962	15	23	0,254797058
9	22	-0,334918626	12	22	-0,078409776	15	22	0,021941117
9	21	0,163863103	12	21	0,215848697	15	21	-0,375451667
10	37	-0,19348535	13	37	0,122959725	16	37	-0,293247897
10	36	-0,152551115	13	36	-0,276694234	16	36	-0,290195932
10	35	-0,357772676	13	35	0,484860318	16	35	-0,4730131
10	34	-0,040565713	13	34	-0,164248576	16	34	-0,387816883
10	33	0,327431448	13	33	0,266232739	16	33	0,023117927
10	32	-0,078641852	13	32	-0,420907976	16	32	-0,092760192
10	31	-0,270871535	13	31	0,365190282	16	31	-0,193154761
10	30	0,457373358	13	30	0,275739008	16	30	0,409546128
10	29	-0,389478841	13	29	0,228162314	16	29	0,030332272
10	28	-0,308041873	13	28	0,3775377	16	28	0,288864273
10	27	-0,368188382	13	27	0,491647202	16	27	0,038293535
10	26	0,104019162	13	26	0,291919141	16	26	-0,187502574
10	25	-0,273921281	13	25	-0,110766184	16	25	0,452238005
10	24	-0,199152297	13	24	-0,106637393	16	24	-0,399295152
10	23	-0,249929179	13	23	0,311172937	16	23	0,300511504
10	22	-0,230622191	13	22	0,294521637	16	22	0,146397439
10	21	0,143275402	13	21	0,098459148	16	21	-0,470528413

## Apéndice - Ponderación de las conexiones de la red neuronal

Fuente	Objetivo	Ponderación	Fuente	Objetivo	Ponderación	Fuente	Objetivo	Ponderación
17	37	0,18563574	19	37	-0,151883968	21	3	0,419428977
17	36	-0,480597681	19	36	0,136929461	22	3	0,123959191
17	35	0,121797096	19	35	-0,30953793	23	3	-0,595585905
17	34	-0,164904384	19	34	-0,241583574	24	3	0,013406585
17	33	0,471174578	19	33	-0,347287492	25	3	0,034703709
17	32	-0,292046727	19	32	0,371182155	26	3	0,116991738
17	31	0,253036311	19	31	-0,320274369	27	3	0,20896136
17	30	-0,19925081	19	30	-0,373155604	28	3	0,307929383
17	29	-0,395802297	19	29	-0,012984684	29	3	0,045971818
17	28	-0,464487608	19	28	0,039373022	30	3	0,149081985
17	27	-0,452064126	19	27	0,415633749	31	3	0,29723705
17	26	-0,296122267	19	26	-0,0401328	32	3	0,310996088
17	25	-0,191391007	19	25	0,10291573	33	3	-0,520924926
17	24	0,249421207	19	24	-0,293258186	34	3	0,424588762
17	23	-0,332283888	19	23	0,020651654	35	3	-0,29635737
17	22	-0,126639967	19	22	0,324031269	36	3	0,133986618
17	21	-0,367408101	19	21	-0,319618489	37	3	0,176332095
18	37	0,062344633	20	37	-0,1344099			
18	36	-0,360028628	20	36	-0,104052331			
18	35	0,380780976	20	35	0,200273924			
18	34	0,363919207	20	34	-0,165191783			
18	33	-0,18382482	20	33	0,209512197			
18	32	-0,074259599	20	32	0,176967369			
18	31	0,169310894	20	31	-0,438745488			
18	30	-0,468984292	20	30	-0,429193741			
18	29	-0,154617374	20	29	0,312347199			
18	28	0,053253489	20	28	0,091784385			
18	27	0,372410147	20	27	-0,08344158			
18	26	0,215959733	20	26	0,082786961			
18	25	0,262139523	20	25	-0,108250345			
18	24	-0,227770801	20	24	0,203181608			
18	23	0,288283528	20	23	0,183432296			
18	22	0,137255399	20	22	-0,475278732			
18	21	0,467334314	20	21	-0,19516975			