

ELABORACIÓN DEL PERFIL DE USO DE HERRAMIENTAS FINANCIERAS EN UNA MUESTRA DE EMPRESAS URUGUAYAS USANDO MÉTODO DE CLUSTERING K-MODES

RAMÓN ÁLVAREZ-VAZ ^a, ELENA VERNAZZA MAÑ ^a, MARGARITA ROLDÓS ^b, MAURO DE LA VEGA ^b

^aInstituto de Estadística,
Facultad de Ciencias Económicas y de Administración,
Universidad de la República de Uruguay
e-mail: ramon@iesta.edu.uy, evernazza@iesta.edu.uy

^bUnidad Curricular Finanzas Corporativas,
Facultad de Ciencias Económicas y de Administración,
Universidad de la República de Uruguay
e-mail: mroldos@gmail.com, maurodelavega@hotmail.com

El objetivo de esta investigación es profundizar el conocimiento en relación a la aplicación de herramientas financieras por parte de empresas uruguayas para la administración de las mismas. A partir de un trabajo de campo, se trata de elaborar perfiles de uso de un set de 4 instrumentos financieros y 5 de proyectos de inversión por partes de las empresas. Si bien originalmente se consulta sobre la frecuencia de uso de esos 9 indicadores, se opta por trabajar con variables binarias que dan cuenta del uso de los mismos.

Se propone emplear el método de clustering k-modes de perfil modal para la determinación de tipologías de encuestados en base a los atributos binarios, de forma de obtener perfiles bien diferenciados.

Los datos a utilizar corresponden a un estudio en una muestra de empresas donde participan estudiantes del Posgrado de Especialización en Finanzas de la Udelar y de un grupo de docentes de la UC Finanzas Corporativas y del IESTA.

Keywords: Clustering, Herramientas Financieras, TIR, variables binarias, VPN.

1. Introducción

Las Finanzas de Empresas, ponen a disposición de los empresarios un conjunto de herramientas financieras que procuran contribuir a una mejor asignación de los recursos financieros, y en consecuencia al alcance de una mayor creación de valor. La toma de decisiones financieras en entornos particularmente inciertos como son los que presentan los mercados emergentes, así como el tamaño de las empresas y el menor desarrollo relativo de sus mercados financieros, parecen condicionar el accionar de los agentes tomadores de decisiones financieras. Las finanzas aplicadas a las empresas tienen como propósito la creación de valor para lo cual la utilización más adecuada de los recursos financieros resulta ser principal [21]. El autor plantea que estas decisiones financieras se pueden clasificar en decisiones de inversión, decisiones de financiamiento y decisiones

de dividendos. Por su parte, [24] propone analizar en qué activos de larga duración debería invertir la empresa (presupuesto de capital), cómo obtener los recursos para realizar dichas adquisiciones (proceso que determina la estructura de capital y que muestra las proporciones en que participan las distintas fuentes de financiamiento, incluyendo fondos propios y fondos de terceros de corto y de largo plazo) y finalmente cómo deben administrarse los flujos de corto plazo derivados de la operativa de la empresa, conocido como manejo del capital de trabajo o de la liquidez. El presupuesto de capital implica el uso de diversos métodos, como Valor Presente Neto (VPN), la Tasa Interna de Retorno (TIR) o el Período de Repago (PR), [24]. El proceso de presupuestación de capital incluye a su vez la evaluación de proyectos, a través de técnicas como el análisis de sensibilidad, análisis de punto de equilibrio, análisis de escenarios, simulación Monte

Carlo, las opciones reales y árboles de decisión [5]. Todas las técnicas antes presentadas forman parte de las herramientas financieras (HF) usadas para la toma de decisiones empresariales. Existe variada literatura especializada, que se dedica a la revisión de la aplicación de estas técnicas, en diferentes contextos y países como [12] para empresas de USA y Canadá, [6], que se focalizan en grandes economías europeas que incluyen Reino Unido, Holanda, Alemania y Francia, [2] que analizan las prácticas presupuestarias de inversión en las economías en transición de Europa Central y del Este. [18] que se enfocan en empresas latinoamericanas, destacando las diferencias de tamaño y de desarrollo relativo menor de los mercados en los que éstas empresas desarrollan su actividad. Dada esta evidencia empírica para el uso de este tipo de herramientas, interesa entonces conocer la forma que en Uruguay, las empresas hacen uso de las mismas y como esta práctica se asocia con las características de estas, lo cual obliga a sistematizarla y donde la elaboración de perfiles de uso es un objetivo relevante. A su vez la metodología de análisis que se presenta en la sección 2, no muestra ser un camino utilizado, ya que no aparecen publicaciones en revistas especializadas en temas de clustering para variables de tipo financieras.

2. Metodología de Análisis

Se usa parte de la metodología propuesta por [25] para clasificar atributos categóricos, a través del algoritmo mixto *Fuzzy C-modes*, y utilizada por [1] para encontrar perfiles de infección parasitaria en escolares de Montevideo. En ambos trabajos cada individuo es previamente clasificado con algún método de clustering y luego pasa por una etapa de difusión, donde pasa a pertenecer a más de un cluster con diferentes grados de participación o membresía. En el método original antes mencionado, se utilizaba el algoritmo *k-modes*, que es de tipo *modal*, y que no es más que un caso particular de un *k-prototipo* descrito por [13]. En este caso el algoritmo tiene una lógica de funcionamiento similar a la del algoritmo *k-means*, y dada la naturaleza de las variables que son binarias, es necesario el uso de otras medidas de disimilaridad, usando un método basado en frecuencias para actualizar los modos [27].

Por lo tanto del método mixto original antes planteado de [25] se trabaja solamente con el algoritmo *k-modes* que aplica la siguiente disimilaridad, siendo x_i, y_i 2 individuos de los que se mide los atributos

$$d(x_i, y_i) = \sum_{j=1}^m \delta(x_j, y_j); \delta(x_j, y_j) = \begin{cases} 0 & \text{si } x_j = y_j \\ 1 & \text{si } x_j \neq y_j \end{cases} \quad (1)$$

El algoritmo trabaja de la siguiente manera, a través de 4 pasos:

1. Selecciona k modos iniciales, uno para cada *clustreseq:2* con x, y variables categóricas binarias en este caso, actualizando el modo;
2. Luego de que todos los individuos han sido asignado, reestima la disimilaridad de los objetos contra el actual modo y si encuentra que un individuo tiene un modo de otro grupo que está más próximo lo reasigna, actualizando los modos de ambos grupos que se modificaron;
3. Repite el paso 3 hasta que ningún individuo haya cambiado de cluster hasta haber visitado todo el conjunto de datos.

El resultado de este algoritmo es entonces una partición de los individuos en grupos cuyo representante es el perfil modal, es decir la combinación de respuestas que es más frecuente en cada cluster.

2.1. Pregunta(s) e hipótesis de investigación. En función de lo planteado en la sección anterior es que la pregunta que se intenta contestar es si el uso de algunas de las HF se asocia con los tributos de las empresas y sus directivos, lo que equivale a plantearse si existen perfiles de uso que diferencien a las prácticas empresariales, y que se podrían resumir en una tipología de uso de las HF, formada por grupos de empresas.

La hipótesis de trabajo entonces es que esa tipología puede ponerse de manifiesto y caracterizarse mediante diferentes aproximaciones metodológicas.

Existe ya una primera experiencia a través de modelos de respuesta discreta (modelos de regresión logística), donde se intentó modelar la propensión al uso de HF en función de características de las empresas y de sus gerentes, con resultados insuficientes, y dado ese antecedente, se propone una alternativa que implica el uso de una técnica estadística descriptiva complementaria y que a su vez supera algunas limitaciones que pueden explicar la pobre performance alcanzada.

En este trabajo el objetivo es obtener perfiles bien diferenciados de uso de las HF en base a los atributos binarios, proponiendo la creación de grupos mediante la estrategia mencionada antes.

3. Descripción del problema en estudio

Se encuestaron 620 empresas para con una tasa de respuesta del 17 %.

Algunos de los atributos evaluados para las empresas y para las características de los tomadores de decisión de las mismas son:

- El tamaño de las empresas, medido a través de sus ventas y de la cantidad de personal ocupado;

- El tipo de negocio, incluyendo el sector de actividad y el mercado en que trabaja (local o exterior);
- El origen del capital (local o extranjero) y en particular el carácter familiar de las empresas. La forma jurídica adoptada para llevar adelante el negocio;
- El nivel de educación de quien toma las decisiones financieras, así como su área de formación, participación en capacitaciones gerenciales, su edad, el tiempo de permanencia en el cargo y su antigüedad en la empresa.

Se utiliza una matriz de datos D_{ij} de 104×8 , a su vez para las HF de cada empresa donde las 8 variables binarias son el uso de algunas de las HF.

Las variables a usar que se dividen en 2 bloques son

Variable	Descripción	Bloque	Tipo
V1	Presupuesto de caja	1	Indicadores financieros
V2	Margen operativo	1	Indicadores financieros
V3	Razón corriente insuficiente	1	Indicadores financieros
V4	Ciclo de conversión de Caja.	1	Indicadores financieros
V5	VPN (Valor presente neto)	2	Proyectos de inversión
V6	TIR (Tasa interna de retorno)	2	Proyectos de inversión
V7	Período de repago	2	Proyectos de inversión
V8	Análisis de sensibilidad	2	Proyectos de inversión

Table 1. Bloques de Variables HF utilizadas

Variable	Descripción	Bloque	Proporción
V1 (PC)	Presupuesto de caja	1	71.2
V2 (MargOp)	Margen operativo	1	58.4
V3 (RazCorr)	Razón corriente insuficiente	1	43.5
V4 (CCC)	Ciclo de conversión de Caja.	1	31.7
V5 (VPN)	VPN	2	32.7
V6 (TIR)	TIR	2	41.5
V7 (PR)	Período de repago	2	59.4
V8 (AS)	Análisis de sensibilidad	2	47.5

Table 2. Proporciones de las variables estudiadas

4. Resultados

Para el análisis global se trabaja con el software [22], para la determinación de los clusters con el algoritmo presentado en la metodología de la sección 2 se usa la librería *kmodes* [27], donde a través de la función *kmodes* de determinan los clusters.

En la Tabla 3 se puede ver como es el comportamiento cambiando de cantidad de clusters, para lo cual se definen 4 escenarios, en los que se presentan varias métricas. Se considera el tamaño en cada grupo, y la métrica *withindiff* que expresa la distancia de *simple matching* intracluster, que expresa la cantidad de desacuerdos para todas las variables dentro de cada cluster. También se construye una métrica que combina las 2 primeras y que se expresa como densidad, relativizando el total de desacuerdos con respecto al total

de observaciones dentro de cada cluster. Esta última es un indicador de homogeneidad interna que puede verse como varía y decrece en algún cluster al cambiar la cantidad de grupos que se establece como restricción.

Analizando la densidad puede verse que esta decrece a medida que hay más cantidad de grupos, con lo cual los escenarios 2 y 3 que consideran 3 y 4 grupos parecen ser una buena solución.

En la Figura 2 y Figura 3 se ve como es la proporción de las variables a la interna de cada bloque para el escenario 2 y 3 respectivamente, siendo que la opción de 4 grupos la mejor, mostrando grupos que en promedio la cantidad de desacuerdos intracluster es más baja (densidad más baja) que para el caso del escenario 2 con 3 grupos.

Métricas para Descripción de los grupos

Métricas para la división de los grafos					
Escenario 1 (2 clusters)	1	2			
tamaño	42	59			
withindiff	58	105			
densidad	1.38	1.77			
Escenario 2 (3 clusters)	1	2	3		
tamaño	40	50	11		
withindiff	50	78	15		
densidad	1.25	1.58	1.36		
Escenario 3 (4 clusters)	1	2	3	4	
tamaño	45	14	8	34	
withindiff	63	23	7	34	
densidad	1.40	1.64	0.87	1.00	
Escenario 4 (5 clusters)	1	2	3	4	5
tamaño	36	14	8	33	10
withindiff	40	23	7	31	9
densidad	1.11	1.64	0.87	0.93	0.9

Table 3. Caracterización de los clusters mediante algoritmo k-modes

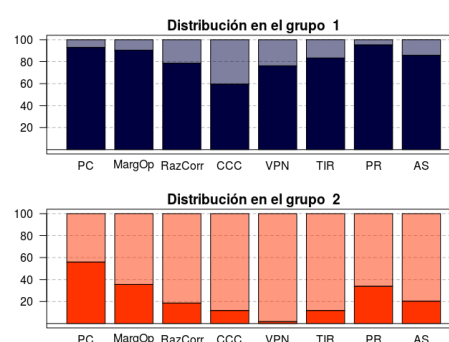


Fig. 1. Proporción de las variables en cada grupo (2 grupos)

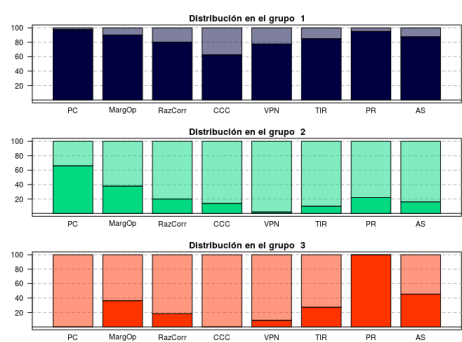


Fig. 2. Proporción de las variables en cada grupo (3 grupos)

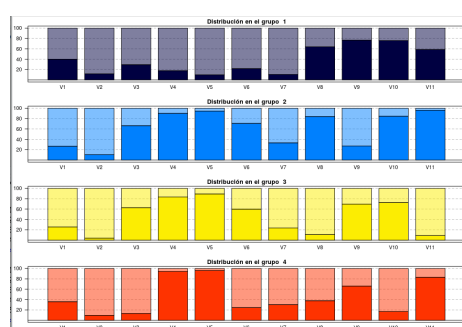


Fig. 3. Proporción de las variables en cada grupo (4 grupos)

Se puede apreciar cual es el perfil modal de cada grupo para las 4 soluciones planteadas en la Tabla 4

Perfil modal								
Escenario k=3 clusters								
Grupo	V1	V2	V3	V4	V5	V6	V7	V8
1	si	si	si	si	si	si	si	si
2	si	no	no	no	no	no	no	no
3	no	no	no	no	no	no	si	no
Escenario k=4 clusters								
Grupo	V1	V2	V3	V4	V5	V6	V7	V8
1	si	no	no	no	no	no	no	no
2	si	si	no	no	no	si	si	si
3	no	no	no	no	no	no	si	no
4	si	si	si	si	si	si	no	si

Table 4. Perfil modal de las variables en cada grupo

En la Tabla 5 se presenta como es la proporción en cada grupo para los 2 bloques de variables y que se puede comparar con la prevalencia global de cada variable, para la opción de 4 grupos, donde cada grupo es (G(1),G(2),G(3),G(4)).

En la Tabla 6 se hace una caracterización de los grupos en función de atributos personales y por otra parte la carga de variables intra-grupo (Cantidad de respuestas SI en las variables de los bloques 1, 2 y 3 de la Tabla 2)

Proporción de cada variable

	G(1)	G(2)	G(3)	G(4)	Total
PC	62.2	78.6	0.0	97.1	71.3
MargOp	33.3	85.7	12.5	91.2	58.4
RazCorr	22.2	21.4	12.5	88.2	43.6
CCC	15.6	7.1	0.0	70.6	31.7
VPN	2.2	28.6	0.0	82.4	32.7
TIR	2.2	71.4	12.5	88.2	41.6
PR	13.3	100.0	100.0	94.1	59.4
AS	13.3	57.1	50.0	88.2	47.5
Total	45	14	8	34	101

Table 5. Perfiles de los grupos creados mediante *k-modes*

Perfiles de los grupos

Total de respuestas(SI)	G(1)	G(2)	G(3)	G(4)	
0	0	13	0	0	13
1	0	11	0	0	11
2	0	14	0	0	14
3	6	3	2	0	11
4	7	0	4	1	12
5	3	0	2	1	6
6	2	0	3	6	11
7	0	0	0	10	10
8	0	0	0	13	13
Total	18	41	11	31	101

Sector

Comercial	50.00	5.90	8.80	35.30	36.7
Industria	47.10	11.80	11.80	29.40	16.8
Otros	53.80	15.40	0.00	30.80	12.9
Servicios	35.10	21.60	8.10	35.10	36.6

Tamaño

0 a 10 personas	56.20	6.20	18.80	18.80	31.7
10 a 50	57.10	17.10	2.90	22.90	34.6
50 a 100	33.30	8.30	8.30	50.0	11.9
más de 100	13.60	22.70	0.00	63.60	31.7

Volumen de ventas

Menos de 5 *10 ⁶ U\$S	53.00	12.10	9.10	25.80	65.3
Mas de 5 *10 ⁶ U\$S	28.60	17.10	5.70	48.60	34.7

Table 6. Asociación de los clusters creados algoritmo *k-modes*, con características de las empresas

Por último se presenta los patrones de respuestas más frecuentes de un total de 50 observados de 256 posibles (2^8)

V1	V2	V3	V4	V5	V6	V7	V8	Observados
No	No	No	No	No	No	No	No	13
Si	Si	Si	Si	Si	Si	Si	Si	13
Si	No	No	No	No	No	No	No	7
Si	Si	Si	No	Si	Si	Si	Si	6
No	No	No	No	No	No	Si	Si	4
Si	Si	No	No	No	Si	Si	No	3

Table 7. Patrones de respuestas más frecuentes

5. Discusión

Los resultados del relevamiento realizado ponen de manifiesto aspectos característicos de la administración financiera de empresas de menor tamaño relativo y que desarrollan sus negocios en el marco de economías emergentes, que son señalados por diversos autores y sobre los cuales se aprecia coincidencia.

Con respecto a la caracterización de los grupos

- Grupo 1 con un total de (n=45) empresas que se caracterizan por los siguiente: por un lado son las empresas que menos usan cualquiera de la HF y a la vez su comportamiento es que si usan algunas de éstas son mayoritariamente las de Indicadores financieros
- Grupo 2 con menos empresas (n=14), que se caracteriza por tener empresas que hacen mas uso de las HF, donde los proyectos de inversión son las HF mas usadas.
- Grupo 3 minoritario, con apenas 8 empresas, las que se distinguen por usar el PR toda ellas y el AS, mientras que las restantes HF son dejadas de lado.
- Grupo 4 que casi representa la tercera parte del total de empresas, las que usan intensamente la HF de ambos tipos, con proporciones cercanas al doble del valor total

Como todo problema de clustering, este no escapa a la situación donde no se sabe exactamente el número de clusters, sino mas bien aproximaciones a un número óptimo. Teniendo en cuenta a su vez que para el caso de la metodología que se usa el algoritmo *k-modes* (homónimo del *k-means*), no es de extrañar que los individuos una vez clasificados en los clusters muestren perfiles que se alejan de los centroides de los grupos (que son los perfiles modales) a pesar de estar más cerca de estos que si estuviesen en otros clusters. Esta situación puede mejorarse extendiendo el método a un método mixto mediante un proceso de difusión donde cada individuo puede extenderse a otros clusters con diferentes grados de

membresía, donde cada individuo pertenece a más de un cluster con diferentes grados de participación, partiendo previamente de una clasificación previa, que en este caso es el algoritmo *k-modes*. Es importante advertir que el fenómeno de difusión que podría darse en los clusters creados con el algoritmo *k-means* puede verse acrecentada por la propia naturaleza de las variable binarias donde la variabilidad intracluster que surge al aplicar el algoritmo *k-means* está mucho más discretizada.

6. Conclusiones y futuros pasos

A su vez tal como se decía en la introducción el objetivo era plantear una alternativa de elaboración de perfiles mediante diferentes metodologías pero que parte de una estrategia metodológica particular y que consiste en no tener en cuenta para nada la jerarquía que existe en las variables al ponerlas a todas en los 2 bloques en igualdad de condiciones para segmentar la población bajo estudio. La literatura muestra que en general la aproximación es de tipo modelizante donde hay claramente un bloque de variables explicadas, que serían las que integran el bloque 1 y 2, las que puede ser explicadas por los atributos de cada empresa. Por eso se proponen otro camino que puedan ayudar a entender mejor las tipologías resultantes mediante la metodología de de detección de comunidades mediante Análisis de Redes Sociales (ARS), [8], Brandes2001, [26], [4] [14], [16]. que se plantean a continuación.

- Encontrar las comunidades dentro de cada bloque de variables y cruzar entre si las particiones elaboradas, comparando con los resultados que surgen al considerar las 8 variables sin agruparlas jerárquicamente.
- Estudiar el problema desde la perspectiva del “Social Network Analysis” (SNA) o el Análisis de redes sociales trabajando a partir de las mismas variables binarias construyendo la matriz de adyacencias sobre la que se pueden aplicar una batería de métricas (closeness, betweenness, modularidad, clustering) sobre los nodos y enlaces, homofilia de los nodos [15], [19], [20], [7] que permitiría detectar comunidades y medidas de centralidad de proximidad
- Para el aspecto macro de la topología de la red, estudiar el diámetro de la misma y la cohesión , tratando de identificar, componentes gigantes, [3], así como también *cliques* de diferente tamaño, [9], [14]
- Comparar los resultados de las comunidades detectadas con el AR, con lo hecho en este trabajo
- Proponer un análisis de redes validando modelos estadísticos (recordar que esto es solo descripción),

donde algunos de los atributos evaluados en la caracterización se pueden usar como variables explicativas, usando la teoría de los modelos exponenciales aleatorios en grafos (ERGM), [14].

Referencias

- [1] Álvarez-Vaz, R. y Massa, F. (2012). Determinación de tipologías de infecciones parasitarias intestinales, en escolares mediante, técnicas de clustering sobre datos binarios. Documento de Trabajo Serie DT (12 / 05) - ISSN : 1688-6453, IESTA.
- [2] Andor, G., Mohanty, S., y Toth, T. (2015). Capital budgeting practices: a survey of central and eastern european firms. *Emerging Markets Review*, (23):148–172.
- [3] Bearman, P. S., Moody, J., y Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American journal of sociology*, 1(110):44–91.
- [4] Borgatti, S. P., Everett, M. G., y Johnson, J. (2013). *Analyzing Social Networks*. SAGE Publications Ltd.
- [5] Brealey, R., Myers, S., y Allen, F. (2011). *Principles of Corporate Finance*. McGraw-Hill / Irwin., New York.
- [6] Brounen, D., de Jong, A., y Koedijk, K. (2004). Corporate finance in europe confronting theory with practice. En *EFA 2004 Maastricht Meetings Paper*, número 2769.
- [7] Currarini, S., Matheson, J., y Vega-Redondo, F. (2016). A simple model of homophily in social networks. *European Economic Review*, 90:18 – 39. Social identity and discrimination.
- [8] Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215.
- [9] Galaso, P., Miranda, A., Goinheix, S., Martínez, C., y Picasso, S. (2018). Redes de cooperación entre empresas: un estudio aplicado a cinco clusters en uruguay. Reporte Técnico 04.
- [10] Giannatale, S. D., Elbittar, A., y Maya, L. (2012). Redes sociales, confianza y hábitos financieros: un análisis empírico. Technical Report 2, CEMLA.
- [11] Ginalski, S. (2011). Las empresas familiares en la red inter-firmas: el caso de la industria suiza de máquinas y metales en el siglo XX. *Redes Revista Hispana para redes sociales*, 2(9):42.
- [12] Graham, J. R. y Harvey, C. R. (2001). The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics*, 60(2):187–243.
- [13] Huang, Z. (1997). A fast clustering algorithm to cluster very large categorical data sets in data mining. in *kdd: Techniques and applications*. Technical report.
- [14] Kolaczyk, E. y Csárdi, G. (2014). *Statistical analysis of network data with R*. Springer, New York.
- [15] Krackhardt, D. y Stern, R. N. (1988). Informal networks and organizational crises: An experimental simulation. *Social psychology quarterly*, pp. 123–140.
- [16] Luke, D. (2015). *A user's guide to network analysis in R*. Springer, Cham.
- [17] Macías-García, A. (2002). Redes sociales y 'clusters' empresariales. *Redes Revista Hispana para redes sociales*, p. 20.
- [18] Maquieira, C., Preve, L. A., y Sarria-Allende, V. (2012). Theory and practice of corporate finance: Evidence and distinctive features in latin america. *Emerging Markets Review*, 2(13):118–148.
- [19] Newman, M. E. J. (2002). Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701.
- [20] Newman, M. E. J. (2003). Mixing patterns in networks. *Phys. Rev. E*, 67:026126.
- [21] Pascale, R. (2009). *Decisiones Financieras*. Prentice Hall-Pearson Education.
- [22] R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [23] Rocha, C. C. (2003). Relaciones estratégicas entre pymes: contraste de hipótesis empresariales mediante ARS. *Redes Revista Hispana para redes sociales*, 4(4):27.
- [24] Ross, S., Westerfield, R., y Jaffe, J. (2012). *Finanzas Corporativas*. McGraw-Hill/Interamericana Editores, México.
- [25] Tsekouras, G., Papageorgiou, D., Kotsiantis, S., Kalloniatis, C., y Pintelas, P. (2005). Fuzzy clustering of categorical attributes and its use in analyzing cultural data. *World Academy of Science, Engineering and Technology*, 1:87–91.

- [26] Wasserman, S. y Faust, K. (1994). *Social network analysis: methods and applications*. Número 8 en *Structural analysis in the social sciences*. Cambridge University Press, Cambridge ; New York.
- [27] Weihs, C., Ligges, U., Luebke, K., y Raabe, N. (2005). Klar analyzing german business cycles. En Baier, D., Decker, R., y Schmidt-Thieme, L., editores, *Data Analysis and Decision Support*, pp. 335–343, Berlin. Springer-Verlag.