

ANÁLISIS DE LAS VENTAS EMPLEANDO REDES COMPLEJAS: COMUNIDADES Y CENTRALIDAD DE PRODUCTOS

SALVADOR HERNÁNDEZ-GONZÁLEZ, JOSÉ EDUARDO HERNÁNDEZ-TORRES,
MANUEL DARÍO HERNÁNDEZ-RIPALDA*

Tecnológico Nacional de México / Instituto Tecnológico de Celaya, México

Recibido: 7 de febrero del 2020 / Aprobado: 9 de abril del 2020

doi: 10.26439/ing.ind2020.n038.4820

RESUMEN: Se analizan las ventas del área de telefonía de una tienda departamental con enfoque de redes complejas. Con las medidas de centralidad y la formación de comunidades, se muestra el ordenamiento de las compras y se cuantifica la relevancia de cada producto. Las ventas se ordenan formando comunidades alrededor de los chips de las compañías telefónicas, en cada caso los teléfonos celulares son los artículos relevantes de cada comunidad.

Palabras clave: análisis de la cesta de la compra / redes complejas / teléfonos celulares / ventas

SALES ANALYSIS USING COMPLEX NETWORKS: COMMUNITIES AND CENTRALITY OF ITEMS

ABSTRACT: Sales in the cell phone and accessories area of a department store are analyzed with a complex network approach. Using measures of centrality and item grouping (communities), purchase ordering is shown, and the importance of each product is quantified. Sales are ordered by forming 'communities' around the cell phone chips. Based on the results, it can be concluded that cell phones are the most significant items of each community.

Keywords: market basket analysis / complex networks / cell phones / sales

* m1803006@itcelaya.edu.mx, salvador.hernandez@itcelaya.edu.mx, dario.hernandez@itcelaya.edu.mx

1. INTRODUCCIÓN

Las empresas, tiendas y negocios de toda índole generan grandes cantidades de comprobantes de pago y estos son, con frecuencia, la única fuente de información en la que se relacionan y agrupan los productos y a partir de los cuales se pueden identificar relaciones e incluso hábitos de compra y consumo (Videla-Cavieres y Ríos, 2014).

Los hábitos de compra de los clientes pueden representarse a través de asociaciones o relaciones entre los diferentes productos que se integran en el comprobante de compra (también llamado "cesta"), permitiendo conocer y entender el comportamiento de los clientes sobre sus preferencias. Cabe señalar que los analistas necesitan herramientas para seleccionar datos y para analizarlos (Raeder y Chawla, 2011) y así determinar la manera en que se agrupan los artículos al momento de ser vendidos (Videla-Cavieres y Ríos, 2014). Con frecuencia la única fuente de tal información son los registros y los comprobantes de venta.

El paradigma de redes complejas ayuda a visualizar un sistema formado por cientos de individuos, visualizarlos, agruparlos en comunidades y cuantificar la relevancia de cada uno dentro de la red de relaciones (Barabási, 2016). Las ventas forman redes donde los nodos corresponden a un artículo, un arco que une el artículo i con el artículo j indicará que ambos se adquirieron juntos, al aplicar las métricas que cuantifican la relevancia de los nodos en la red, se responderán preguntas sobre ¿cómo cuantificar la relevancia de un artículo dentro de la red de ventas?, ¿cómo se agrupan los artículos vendidos?, ¿cómo se afecta la estructura de las ventas si el artículo X se elimina?

En este documento se analizan las ventas de telefonía celular de una tienda departamental: mediante el enfoque de análisis de la cesta de supermercado se cuantifica la certidumbre de las relaciones entre productos, a continuación, cada cesta de supermercado se representa como una red y, finalmente, se cuantifica la relevancia de los artículos dentro de la red y la forma en que se agrupan.

2. TRABAJOS PREVIOS

Existen varios trabajos previos sobre el análisis de las ventas con un enfoque de redes: en Huang, Zeng y Chen, (2007) se aplica el enfoque de redes complejas a una cadena de ventas *on-line* de libros. Raeder y Chawla (2011) proponen mejorar la claridad del análisis de la cesta de mercado mediante el modelado de datos como una red. Kim, Kim y Chen (2012) aplican el enfoque de redes complejas para determinar las relaciones entre productos en una cadena de tiendas. Pennacchioli, Coscia, Rinzivillo, Gianotti y Pedreschi (2014) llevan a cabo el análisis de las transacciones de una cadena de tiendas departamentales en Italia. Videla y Ríos (2014) analizan las transacciones de una cadena departamental en Chile. Wan, Chen, Fu y Yang (2015) determinan las preferencias del mercado de automóviles en China. Zhao, McAuley, Li y King (2017) desarrolla un sistema

para proponer sugerencias de compra a los clientes basadas en la construcción de una red de productos sustitutos y productos complementarios. Sha *et al.* (2018) aplican el enfoque de redes complejas para modelar las decisiones de compra en supermercados de China.

Valle, Ruz y Morrás (2018) diseñan un sistema experto de análisis de las compras basado en el modelo de árbol de expansión mínima, para identificar las relaciones de mayor peso y, posteriormente, utilizan métricas de redes para cuantificar la relevancia de los artículos.

Regularmente, el paradigma empleado es construir la red a partir de los comprobantes y a continuación determinar las reglas de asociación empleando, por ejemplo, el índice de agrupamiento (*clustering*); cabe señalar que, los comprobantes de compra contienen numerosas relaciones espurias, es decir, relaciones entre artículos que no indican un patrón o hábito de consumo y que deben descartarse del análisis.

En esta aplicación, primero se construyen las reglas de asociación, a continuación, se construye la red y con las métricas de centralidad se determinan las comunidades existentes y se cuantifica la relevancia de cada artículo.

3. ANÁLISIS DE LA CESTA DE MERCADO

La minería de datos es un conjunto de técnicas que se emplean para encontrar patrones y relaciones entre las variables contenidas en los datos guardados en sistemas de almacenamiento de información; para el caso de las ventas de telefonía celular de una tienda departamental se ha empleado una técnica conocida como análisis de cesta de supermercado (Raeder y Chawla, 2011). El objetivo del análisis de la canasta de mercado es el descubrimiento e interpretación de las reglas de asociación que se forman entre artículos de un catálogo (Agrawal y Srikant, 1994; Kaur y Kang, 2016).

Dado un umbral s , denominado soporte mínimo, y un umbral c , conocido como confianza mínima, se encontrarán todas las reglas de la forma A-B, donde A y B son conjuntos de productos, de manera que:

- A y B aparecen juntos en al menos $s\%$ de las transacciones.
- B ocurre en al menos $c\%$ de las transacciones en las que A ocurre.

Una regla de asociación se admite en una base de datos de transacciones si cumple con los criterios de soporte mínimo y de confianza mínimo (Ayse Nur y Ayvazb, 2018). Formalmente, el nivel de confianza (C) de que A y B se adquieran juntos se plantea como sigue:

$$C(A \rightarrow B) = \frac{P(AB)}{P(A)P(B)} \quad (1)$$

donde $P(X)$ es la proporción de transacciones o compras en las que se incorpora el producto X (Raeder y Chawla, 2011), a mayor valor de C , mayor confianza de que la relación entre pares de productos sea significativa, en consecuencia, es mayor la probabilidad de que se adquieran juntos.

4. REDES COMPLEJAS

Existe un consenso en el sentido de que el inicio de la teoría de grafos se encuentra en los trabajos de Euler, cuando propuso el problema de los puentes de Königsberg. Con el paso del tiempo, la teoría de grafos, el análisis de redes y las redes complejas han sido aplicados a diversas disciplinas, el problema original de la teoría de grafos surgió de un problema de transporte. El auge del estudio de las redes complejas inició a finales de los años noventa dando lugar a un amplio soporte conceptual y metodológico en el análisis de redes sociales y gracias a aportes empíricos y teóricos de la física estadística.

En el desarrollo del formalismo de las redes complejas, los biólogos, sociólogos, economistas e ingenieros de diversas áreas se han concentrado en la recolección de información empírica y en la fenomenología de diversos sistemas sociales, naturales y artificiales, los matemáticos han hecho aportes en la teoría de grafos y los físicos han dado las bases teóricas para la definición de una concepción que integra trabajo empírico, análisis y modelación (Barabási, 2016).

Una red es una representación de los componentes de un sistema, los cuales son llamados *nodos* o *vértices*, las interacciones directas que hay entre ellos se representan mediante arcos, llamados también *enlaces* (figura 1). Con esta representación, mediante una red es posible estudiar sistemas que pueden diferir mucho en naturaleza, apariencia o alcance (Barabási, 2016).

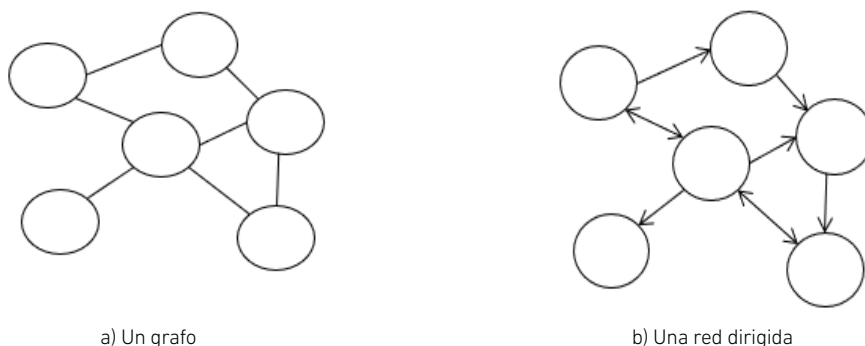


Figura 1. Diferencia entre a) grafo y b) red

Elaboración propia

Una red compleja se refiere a una red o grafo, la cual tiene propiedades estadísticas y topológicas cuantificables. En la actualidad, las redes complejas se estudian por su relación con muchos campos de la ciencia. Muchos sistemas en la naturaleza se pueden describir por medio de redes complejas.

4.1 Indicadores locales (centralidad)

Para cuantificar la relevancia de un nodo dentro de la estructura se emplean indicadores locales o medidas de centralidad. En este trabajo se consideraron las medidas grado del nodo y centralidad de la intermediación.

4.1.1 Grado de un nodo

Se define como el número de arcos o conexiones que contiene un nodo:

$$k_i = \sum_j A_{ij}, i = 1, 2, \dots \quad (2)$$

donde k_i es el grado del nodo i , A_{ij} toma valor 1 si existe un arco que parte del nodo i al nodo j ; 0 en otros casos. Si la gráfica es dirigida, entonces los nodos cuentan con grado de salida y grado de entrada, el grado del nodo será la suma del grado de entrada y el grado de salida (Freeman, 1977). El grado promedio de una red se obtiene como sigue:

$$k_{prom} = \frac{1}{N} \sum_i k_i \quad (3)$$

donde k_{prom} es el grado promedio del nodo, N es el número de nodos de la red y k_i es el grado del nodo i .

4.1.2 Centralidad de la intermediación

La importancia de un nodo por la frecuencia con la que aparece en la ruta más corta entre los nodos j y k se mide mediante el índice de intermediación (*betweenness centrality*). Cuando una mayor proporción de rutas cruzan un nodo específico, mayor relevancia tendrá como punto de conexión; la centralidad de intermediación del nodo i (b_i) se calcula como sigue (Barthelemy, 2011; Freeman, 1977):

$$b_i = \sum_{j \neq k \neq i} \frac{n_{jk}(i)}{n_{jk}} \quad (7)$$

donde $n_{jk}(i)$ es el número de rutas más cortas que conectan los nodos j y k y que pasan por el nodo i , n_{jk} es el número de total de todas las rutas más cortas que conectan los nodos j y k . A mayor valor de intermediación, mayor relevancia del nodo como punto de conexión entre distintas secciones de una red. Eliminar un nodo con un valor de intermediación alto lleva a problemas de comunicación entre secciones de una red.

4.1.3 Modularidad

Identificar comunidades dentro de una red proporciona información sobre la estructura y sobre subsistemas o grupos (módulos). Para detectar una comunidad es necesario particionar la red agrupando los nodos de conjuntos densamente conectados y diferenciándolos de aquellos nodos con sólo algunas conexiones. El valor del módulo identifica una comunidad o familia en la red, sin embargo, es necesario estudiar las comunidades para identificar rasgos o características comunes. Cuando se aplican las medidas de centralidad a las comunidades, se cuantifica la cohesión entre los nodos que las forman.

Por lo regular se utilizan algoritmos heurísticos que construyen comunidades pequeñas a las que se van agregando nodos, siguiendo un proceso iterativo hasta que todos los nodos han sido agregados a una comunidad (Gephi, 2010).

5. MATERIALES Y MÉTODOS

Como caso de estudio se analizaron los comprobantes de una semana de ventas (11 al 17 de febrero de 2019) del área de telefonía que fueron proporcionados por una tienda departamental ubicada en la ciudad de Querétaro, México. El método consistió en 7 pasos los cuales se describen brevemente a continuación (figura 2).

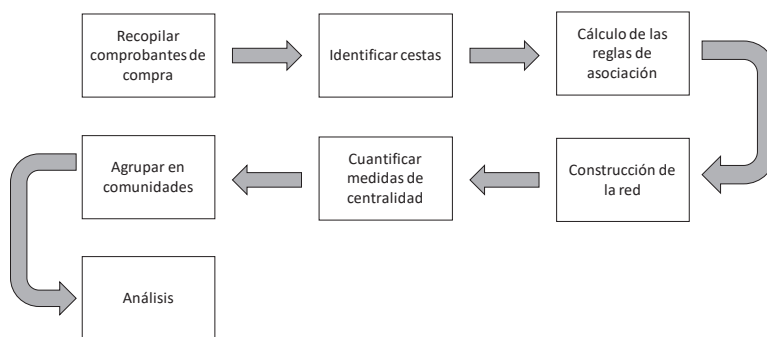


Figura 2. Fases para el análisis de las ventas

Elaboración propia

Cada comprobante de venta corresponde a una cesta de supermercado con lo que se obtuvo la siguiente tabla de relaciones con los nombres codificados de los productos (tabla 1):

Tabla 1
Relación de ventas de teléfonos celulares

Comprobante	PRODUCTO 1	PRODUCTO 2	PRODUCTO 3	PRODUCTO 4
1	TEL SYJ8	CHIP TX		
2	TEL SYJ4	CHIP TX		
3	EL MOMOG6PL	CHIP MY		
4	TEL NX	CHIP TX		
5	TEL AL	CHIP TX		
6	TEL SYJ4LAV	CHIP TX	MICAS BL SY J4	
7	TEL UNHIL675 PRO	CHIP UW		
8	TEL SYA7	CHIP TX		
9	TEL SYLJ4	CHIP TX	CARGADOR BL CARRO USB	CABLE BL MICRO 2M
10	TEL SYALJ8	CHIP TX	MEMORIA MICRO SD K 32 GB	
11	TEL NXXN306NGP R6	CHIP TX		
12	TEL SGJ6	CHIP MZ 4G		
13	TEL ATHUP20LIT AZU	CHIP AV		
14	TEL ATSAMJ6	CHIP AV		
15	TEL ATHIL675 PRO	CHIP AV		
16	TEL ATHUP20LITNEG	CHIP AV		
17	TEL MOM0500ID	CHIP TX		
18	TEL CELMOONE	CHIP MZ 4G		
19	TEL AIPH632GB	CHIP MZ 4G		
20	TEL SGJ6	CHIP MZ 4G	PROM TEL SAM CAMISETA ENAMO	
21	TEL CELZUONS	MLT CHIP		
22	TEL ANILIM9	CHIP MZ 4G		
23	TEL SYJ610GJ632G	CHIP MZ 4G		
24	TEL UNHIL675PRO	CHIP UW		
25	TEL SYPRI16GB	CHIP MZ 4G		
26	TEL ZTBLL7A DRD	CHIP MZ 4G		
27	TEL ATSYJ4	CHIP AV		
28	TEL ZTBL7AAZU	CHIP MZ 4G		
29	TEL SYA7AZU	CHIP MZ 4G		

(continúa)

(continuación)

30	TEL MOG6PLAAZU	CHIP MZ 4G	PROTECTOR BL MO G6P
31	TEL SYJ415G32GB	CHIP MZ 4G	
32	TELX210HM K9 16 G	CHIP MZ 4G	
33	TEL SYJ632G	CHIP MZ 4G	
34	TEL ATHUMAT20LNEG	CHIP AV	

Elaboración propia

Por ejemplo, el comprobante 1 contiene un teléfono celular (identificado por la clave TEL) y un chip; el comprobante 6 corresponde a un teléfono celular, un chip y micas. En total se revisaron 34 comprobantes de compra.

Posteriormente, se realizó el cálculo de las reglas de asociación determinando primero el número de cestas en las que aparece el producto, a continuación, se calculó el “soporte” como el cociente del total de cestas en las que A y B se adquieren juntos, del total de cestas analizadas, enseguida se obtuvo la “confianza” que muestra la fracción con la que los artículos en B aparecen o se venden en conjunto con los artículos A y, finalmente, se evaluó la “confianza esperada”, que cuantifica la confianza o “certeza” de que se compre un artículo B, dado que ya se compró un artículo A, en este caso, a mayor valor de la confianza esperada, mayor es la certeza de que A y B se compren juntos.

En la siguiente fase, y con los pesos dados por la frecuencia de compra, se generó la red de ventas en donde cada nodo de la red corresponde a un artículo. Una vez que se generó la red, el siguiente paso consistió en obtener las medidas de centralidad y, finalmente, se agruparon los nodos y sus relaciones en comunidades o módulos; la construcción de la red, el cálculo de medidas de centralidad y la agrupación en módulos se ejecutaron empleando el paquete Gephi diseñado para construir modelos de redes de sistemas.

6. RESULTADOS

Una muestra de las reglas de asociación calculadas se presenta en la tabla 2. Se encontraron cestas con un nivel de confianza bajo en su relación con otros artículos, por ejemplo, la relación TEL NX - CHIP TX y la relación TEL NXXN306 NGP R6 - CHIP TX, en ambos, el nivel de confianza es 0,34, se obtiene este valor porque existen varios modelos de teléfonos celulares que se adquirieron junto con el chip TX, se obtiene el mismo valor para otras 6 relaciones, esto indica una probabilidad uniforme de adquirir el chip TX y un teléfono celular de cualquier modelo.

En la relación 4, formada por el teléfono TEL SYJ4LAV y las MICAS BL SY J4, la cual obtuvo un nivel de confianza de 34, el resultado indica que existe un nivel de certidumbre muy alto de que ambos artículos se adquieran juntos, esto se explica porque el tipo de mica es exclusiva para este modelo. Cada una de las reglas de asociación cuantifican la certidumbre de una compra que incluye los artículos A y B; en este caso de estudio, los productos considerados forman relaciones obvias y permiten establecer una agrupación dentro de esta sección del departamento.

Sin embargo, con este enfoque resulta complicado identificar la relevancia de un producto dentro de la red, ya sea por el número de artículos diferentes con los que se relaciona o bien visualizar su relevancia como puente entre artículos que no están conectados directamente.

Tabla 2
Ejemplo de cálculos para las relaciones

PRODUCTO	¿CUÁNTAS CESTAS CONTIENEN EL PRODUCTO? (A&B)	SOPORTE (FRECUENCIA) (A&B)/34	CONFIANZA (A&B)/A	CONFIANZA ESPERADA [(A&B)/(A)]/[B/ TOTAL]	FAMILIA
CHIP TX->TEL SGJ8	2	0,0588	0,2	0,68	5
TEL SYJ8, MEMORIA MICRO SD 32 GB	1	0,0294	0,5	8,5	5
TEL SYJ4, CHIP TX	2	0,0588	0,2	0,68	5
TEL SYJ4, CARGADOR BL CARRO USB, CABLE BL MICRO 2M	1	0,0294	0,5	8,5	1
TEL NX, CHIP TX	1	0,0294	0,1	0,34	5
TEL AL, CHIP TX	1	0,0294	0,1	0,34	5
TEL SYJ4LAV, CHIP TX	1	0,0294	0,1	0,34	5
TEL SYJ4LAV, MICAS BL SAM J4	1	0,0294	1	34	5
TEL SYA7, CHIP TX	1	0,0294	0,1	0,34	5
TEL NXXN306 NGP R6, CHIP TX	1	0,0294	0,1	0,34	5
TEL MOMOT500ID, CHIP TX	1	0,0294	0,1	0,34	5
TEL SYJ6, CHIP MZ 4G	2	0,0588	0,1429	0,3469	0
TEL SYJ6, PROM TEL SAM CAMISETA ENAMO	1	0,0294	0,5	8,5	0

Elaboración propia

A continuación, las relaciones se modelaron como una red para observar el ordenamiento de las ventas, las comunidades existentes y la relevancia de los artículos, el resultado se observa en la figura 3.

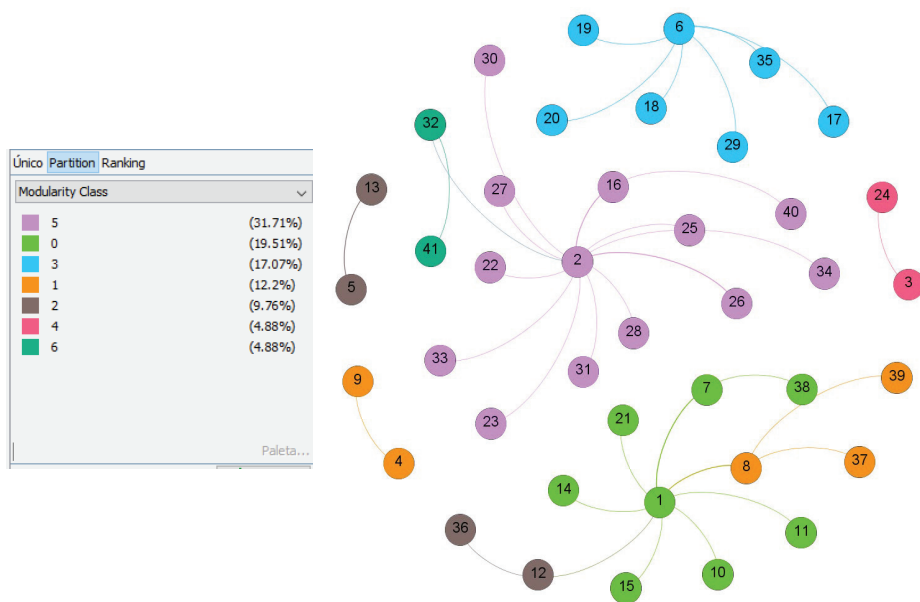


Figura 3. Red de ventas y comunidades identificadas en del departamento de telefonía celular
Elaboración propia

Una vez obtenida la red, se calcularon las medidas de centralidad del grado y de la intermediación, así como las comunidades o módulos, todos los cálculos se realizaron empleando el paquete Gephi (2010).

En total se obtuvieron 7 comunidades: la comunidad 5 tiene el mayor número de nodos agrupados alrededor del chip de la compañía MZ (nodo 2), la familia 0 muestra un agrupamiento en torno del chip de la compañía TX (nodo 1) y la familia 3 se aglomera alrededor del chip de la compañía AV (nodo 6). Como resultado, estas tres comunidades concentran el 68,29 % de los nodos, también se encontró que tres artículos de la comunidad 1 están conectados a la familia 0 a través de un teléfono celular (nodo 8).

6.1 Centralidad del grado

Se encontró que en las familias 5, 0 y 3, el nodo de mayor grado es el chip que da acceso al servicio de cada compañía de telefonía (nodos 1, 2 y 6), alrededor de estos chips se adquieren los teléfonos celulares con sus respectivas características, en otras palabras, el resultado

indica que las ventas tienden a formar una red con un arreglo ordenado alrededor de los chips que ofrecen las compañías; cabe señalar que la variedad de teléfonos celulares es mayor comparada con el número de compañías que ofrecen el servicio. Este resultado es interesante porque regularmente se cree que el producto principal es el teléfono.

En el caso de la familia 5, el chip (nodo 2) cuenta con 11 arcos y todos conectan a teléfonos celulares, el nodo 40 corresponde a una playera patrocinada por la compañía del chip y está enlazada al nodo 16. También se encontró que existe una conexión con la comunidad 6 formada por un teléfono celular (nodo 32), un protector.

En la familia 0, por ejemplo, el chip de la compañía TX (nodo 1) tiene grado 8, los nodos 7, 8, 10, 11, 12, 14, 15 y 21 están conectados a este chip y corresponden a modelos de teléfonos celulares; enlazado al nodo 7 se encuentra una tarjeta de almacenamiento SD (nodo 38); al nodo 12 se enlazaron micas protectoras (nodo 36); el nodo 8 tiene, a su vez, enlazados un cable (nodo 37) y un cargador (nodo 39).

6.2 Intermediación

En la tabla 3 se muestra la comunidad 5, con el grado de cada nodo y el valor de intermediación calculado: la inspección muestra que después del nodo 2, el nodo 16 es un artículo (teléfono) donde el valor de intermediación es el más alto; al inspeccionar esta comunidad, se puede observar que enlaza con una playera (nodo 40) que no se encuentra en ninguna otra cesta de esta familia.

Tabla 3
Datos de centralidad en la familia 5

Nodo	Grado	Intermediación	Modularidad
2	12	0,1141	5
16	2	0,0166	5
22	1	0	5
23	1	0	5
40	1	0	5
25	1	0	5
26	1	0	5
27	1	0	5
28	1	0	5
30	1	0	5
31	1	0	5
33	1	0	5
34	1	0	5

Elaboración propia

Considere ahora la familia 1 formada por 5 nodos, en esta familia el nodo 8 es relevante porque a través de ese artículo hay conexión con la familia 0, su valor de intermediación es el más alto dentro de esta comunidad, ya que conecta un cargador y un cable. Un caso especial es la familia 2, está formada por 4 nodos, de estos, el artículo 12 es el de mayor relevancia porque es el que conecta a esta familia con la familia 0 (tabla 4).

Tabla 4
Datos de centralidad de las familias 1 y 2

Nodo	Grado	Intermediación	Modularidad
8	3	0,3181	1
4	1	0	1
9	1	0	1
37	1	0	1
39	1	0	1
12	2	0,1667	2
36	1	0	2
5	1	0	2
13	1	0	2

Elaboración propia

Finalmente, en la familia 0, después del nodo 1, el siguiente nodo de mayor relevancia es el 7, que corresponde a un teléfono celular, este conecta la comunidad con una memoria SD (nodo 38) (tabla 5).

Tabla 5
Datos de centralidad de la familia 0

Nodo	Grado	Intermediación	Modularidad
1	8	0,9242	0
7	2	0,1667	0
38	1	0	0
10	1	0	0
11	1	0	0
14	1	0	0
15	1	0	0
21	1	0	0

Elaboración propia

Los nodos con valor de intermediación 0 corresponden a nodos que no son relevantes para conectar pares de nodos dentro de la red.

Si bien la centralidad de un nodo cuantifica la relevancia de un artículo, también mide el efecto en caso de que dicho nodo sea eliminado de la red, en el contexto de redes complejas a este fenómeno se le conoce como "falla", por ejemplo, considerando nuevamente la familia 1, si se elimina del catálogo el artículo 8 se rompe la conexión entre las familias 1 y 0.

7. DISCUSIÓN

De los resultados obtenidos, se deduce que las reglas de asociación y su respectivo nivel de confianza se limitan a cuantificar la certeza de que un par de productos se adquieren juntos, por ejemplo, se puede mencionar el caso de la cesta formada por TEL SYJ4 - CARGADOR BL CARRO USB - CABLE BL MICRO 2M (nodos 8, 37 y 39) donde el nivel de certeza es 8,5 que, en comparación con otras cestas, indica una fuerte relación entre estos tres artículos.

Ahora bien, al modelar las relaciones como red se encuentra que el teléfono TEL SYJ4 está relacionado con el chip TX (nodo 1) y, además, forma un grupo con otros 7 modelos.

Por otra parte, las medidas de centralidad muestran con quién se relacionan los artículos (grado) y qué artículos forman puentes (intermediación): el protector para teléfono celular (nodo 41) se adquirió a través del teléfono celular TEL MOG6PLAAZU (nodo 32), que a su vez se adquirió con el teléfono celular TEL SYPRI16GB.

Finalmente, se encontró que las ventas se pueden catalogar en 7 familias; de estas, 3 tienen una estructura ordenada alrededor del chip de la compañía telefónica que proporciona el servicio, en otras palabras, la red indica qué compañías celulares controlan la venta de un teléfono.

Representar las relaciones como red, además de reforzar los resultados del nivel de confianza de las reglas de asociación, muestra el sistema en una forma que permite identificar relaciones potenciales entre artículos y que pueden ofertarse juntos.

Cabe señalar que una limitante en este trabajo es la cantidad de datos, en este caso, los comprobantes corresponden a una semana de ventas y, además, se enfocó en un solo departamento, lo que afecta el nivel de confianza de la cesta de supermercado y, en consecuencia, los pesos de los enlaces entre pares de nodos. Por lo anterior, las bondades de los modelos en red serán más evidentes al aplicarse a historiales de ventas con un horizonte temporal mayor.

8. CONCLUSIONES

El enfoque del análisis de la cesta de supermercado sirve para identificar las relaciones entre los artículos que se ofertan en un negocio, sin embargo, al analizar la cesta con un enfoque de red compleja, se identifican las comunidades o familias que se forman y, además, se cuantifica la relevancia de cada artículo mediante las medidas de centralidad.

La red permite identificar que las ventas del departamento de telefonía tienen como producto central a los chips, alrededor de estos se venden los diferentes modelos de teléfonos celulares, igualmente, se observó que compañías celulares controlan la venta de teléfonos celulares.

La centralidad de la intermediación ayuda a identificar los artículos relevantes dentro de la estructura y a identificar con qué otros artículos se relacionan; en este caso, se encontró que los artículos relevantes resultaron ser los chips y, en segundo lugar, los teléfonos celulares que forman puentes hacia cables, cargadores y micas, objetos que tienen el potencial de ofrecerse en paquete.

Este trabajo muestra cómo el enfoque de redes complementa una técnica de análisis de las ventas, proporcionando información sobre la estructura que forma el comportamiento de consumo de los clientes; se identificaron comunidades y se cuantificó la relevancia de cada artículo. En una extensión de este trabajo, debe aplicarse dicho análisis a una base de datos de mayor tamaño, en este caso, se analizaron únicamente 34 comprobantes, aun así, la información obtenida permite apreciar el potencial para analizar bases de datos de mayor tamaño.

REFERENCIAS

- Agrawal, R., y Srikant, R. (1994). Fast algorithms for mining association rules. En *Proceedings of the 20th VLDB Conference* (pp. 487-499). Santiago: Almaden Research Center.
- Ayse Nur, S. y Ayvazb, B. (2018). Determination of association rules with market basket analysis: an application in the retail sector. *Southeast Europe Journal of Soft Computing*, 7(1), 10-19. doi:10.21533/scjournal.v7i1.149
- Barabási, A. L. (2016). *Network Science*. Cambridge: Cambridge University Press.
- Barthelemy, M. (2011). Spatial networks. *Physics Reports*, 499(1-3), 1-101. doi:10.1016/j.physrep.2010.11.002
- Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 35-41. doi:10.2307/3033543

- Fruchterman, T., y Reingold, E. (1991). Graph drawing by Force-directed Placement. *Software: Practice and Experience*, 21(11), 1129-1164. doi:10.1002/spe.4380211102
- Gephi (2010). Gephi Tutorial Quick Start. Recuperado de https://gephi.org/tutorials/gephi-tutorial-quick_start.pdf
- Huang, Z., Zeng, D., y Chen, H. (2007). Analyzing consumer-product graphs: empirical findings and applications in recommender systems. *Management Science*, 53(7), 1146-1154. doi:0.1287/mnsc.1060.0619
- Kaur, M., y Kang, S. (2016). Market basket analysis: identify the changing trends of market data using association rule mining. *Procedia Computer Science*, 1(85), 78-85. doi:10.1016/j.procs.2016.05.180
- Kim, H., Kim, J., y Chen, Q. (2012). A product network analysis for extending the market basket analysis. *Expert Systems with Applications*, 39(8), 7403-7410. doi:10.1016/j.eswa.2012.01.066
- Pennacchioli, D., Coscia, M., Rinzivillo, S., Gianotti, F., y Pedreschi, D. (2014). The retail market as a complex system. *EPJ Data Science*, 3(33), 1-27. doi:10.1140/epjds/s13688-014-0033-x
- Raeder, T., y Chawla, N. V. (2011). Market basket analysis with networks. *Social Network Analysis and Mining*, 1(2), 97-113. doi:10.1007/s13278-010-0003-7
- Sha, Z., Huang, Y., Fu, J., Wang, M., Fu, Y., Contractor, N., y Chen, W. (2018). A network-based approach to modeling and predicting product co-consideration relations. *Complexity*, 2018, 1-13. doi:10.1155/2018/2753638
- Valle, M. A., Ruz, G. A., y Morrás, R. (2018). Market basket analysis: complementing association rules with minimum spanning trees. *Expert Systems with Applications*, 97(1), 146-162. doi:10.1016/j.eswa.2017.12.028
- Videla-Cavieres, I., y Ríos, S. (2014). Extending market basket analysis with graph mining. *Expert Systems with Applications*, 41 (4), 1928-1936. doi:10.1016/j.eswa.2013.08.088
- Wan, M., Chen, W., Fu, Y., y Yang, Y. (2015). Analyzing and predicting heterogeneous customer preferences in China's auto-market using choice modeling and network analysis. *International Journal of Materials and Manufacturing*, 8(3), 1-10. doi:10.4271/2015-01-0468
- Zhao, T., McAuley, J., Li, M., y King, I. (2017). Improving recommendation accuracy using networks of substitutable and complementary products. En *2017 International Joint Conference on Neural Networks (IJCNN)* (pp. 3649-3655). Anchorage, AK: IEEE. doi:10.1109/IJCNN.2017.7966315

