

ANÁLISIS DE LOS COMPROBANTES DE COMPRA DE UN MINORISTA APLICANDO REDES COMPLEJAS

ANALYSIS OF THE RECEIPTS OF A RETAILER APPLYING COMPLEX NETWORKS

Salvador Hernández-González^{1,♦}, Alejandro Estrada-Omaña², Idalia Flores-de-la-Mota³,
José Alfredo Jiménez-García⁴, Vicente Figueroa-Fernández⁵

RESUMEN

Las empresas, tiendas y negocios generan grandes cantidades de comprobantes de pago y con frecuencia son la única fuente de información sobre la forma en que se relacionan y agrupan los productos. Se analizaron los comprobantes de compra de una tienda minorista aplicando el enfoque de redes complejas. El catálogo consta de 272 productos pero los artículos se relacionan en promedio con 10 productos. A partir de la modularidad, se identificaron 39 familias de artículos; se muestran 2 ejemplos de ofertas de artículos para adquirirse juntos. Las compras forman una red tipo “mundo-pequeño” ordenada alrededor de tres productos relevantes. Esta red ha alcanzado un estado donde la probabilidad de que un artículo establezca una nueva conexión es baja. Los resultados se compararon con el sistema de clasificación ABC del inventario, los resultados muestran una clasificación que complementa el análisis y gestión de un inventario.

Palabras claves: Comprobantes de compra, relaciones entre productos, centralidad, redes complejas.

¹Departamento de Ingeniería Industrial, Tecnológico Nacional de México. Instituto Tecnológico de Celaya, Guanajuato, México. orcid.org/0000-0001-7237-9521

²Departamento de Ingeniería Industrial, Tecnológico Nacional de México. Instituto Tecnológico de Celaya, Guanajuato, México. al.estrada.o@outlook.com

³Universidad Nacional Autónoma de México, Ciudad de México, México. idalia.flores@unam.mx. orcid.org/0000-0002-5745-1756

⁴Departamento de Ingeniería Industrial, Tecnológico Nacional de México. Instituto Tecnológico de Celaya, Guanajuato, México. alfredo.jimenez@itcelaya.edu.mx. orcid.org/0000-0002-5293-4855

⁵Departamento de Ingeniería Industrial, Tecnológico Nacional de México. Instituto Tecnológico de Celaya, Guanajuato, México. vicente.figueroa@itcelaya.edu.mx. orcid.org/0000-0003-2824-9994

♦Autor para correspondencia: salvador.hernandez@itcelaya.edu.mx

ABSTRACT

Businesses, stores, and businesses generate large amounts of vouchers and are often the only source of information about how products are related and grouped. Vouchers from a retail store were analyzed using the complex networks approach. The catalog consists of 272 products; On average, a product is related to 10 other articles. From the modularity, 39 item families were identified, from these, 2 examples of offers with items that can be purchased together are shown. The sales form a "small-world" network ordered around three relevant products. His network has reached a state where the probability of a product establishing new connections is minimal. The results were compared with the ABC inventory classification system; the result shows a classification that complements the analysis and management of an inventory.

Keywords: Receipts, product relations, centrality, complex networks

INTRODUCCION

Las empresas, tiendas y negocios de toda índole generan grandes cantidades de comprobantes de pago y con frecuencia son la única fuente de información sobre la forma en que se relacionan y agrupan los productos y a partir de los cuales se pueden identificar relaciones e incluso hábitos de consumo (Videla & Ríos, 2014), esto es lo que se conoce como administración de las relaciones con los clientes (customer relationship management, CRM). Cuando un administrador conoce esta estructura y los artículos de mayor relevancia de su catálogo, entonces se encontrará en capacidad de generar estrategias que permitan incrementar sus ventas ofreciendo incentivos para motivar la compra, por ejemplo, ofreciendo paquetes de productos (Wan *et al.*, 2015; Hashemian *et al.*, 2017). Por ejemplo, en los sistemas de compras en línea es común observar una lista de recomendaciones, ésta se genera a partir de las consultas que realiza un cliente sobre el catálogo o bien a partir de la compra que realiza; la lista constará de aquellos productos del catálogo que tienen algo en común para el cliente (Zuo *et al.*, 2018).

Dada la cantidad de información que se genera, presentar la información en forma de tablas o reportes tradicionales no siempre es práctico (Rodríguez *et al.*, 2016). Un estudio de esta naturaleza está hoy en día al alcance dados los recientes avances en los sistemas computacionales para el manejo masivo de datos, aunque esto ha tenido un auge en las últimas dos décadas. El análisis de las compras y la relación entre productos, que de esta se pueden deducir se le conoce como análisis de la cesta de compra (Market Basket Analysis).

Por otra parte, los investigadores han comenzado a aplicar herramientas de análisis nuevas para determinar la relevancia de cada producto y las relaciones y grupos que forman, siendo una de las más recientes propuestas el enfoque de redes complejas, se mencionan a continuación tres bondades: existen métricas que cuantifican la relevancia (o centralidad) de cada nodo dentro de una red, existen algoritmos que permiten agrupar los nodos por familias a partir de las cuales pueden extraerse las relaciones entre artículos y el formato

para presentar los resultados permite manejar grandes cantidades de datos (Ceballos *et al.*, 2016; Aktas & Meng, 2017).

En lo que respecta al análisis de la canasta de mercado Agrawal *et al.* (1993) han propuesto los lineamientos para determinar las reglas de asociación entre productos, este trabajo considerado de los primeros en detectar los patrones de comportamiento. Por su parte, Huang *et al.* (2007) han aplicado el enfoque de redes complejas a una cadena de ventas on-line de libros asentada en Taiwan así como algoritmos para genera sugerencias a los clientes.

El enfoque de redes complejas se ha aplicado para determinar las relaciones entre productos y compras complementarias e identificar artículos relevantes (Kim *et al.*, 2012), estudiar las transacciones en tiendas departamentales (Pennacchioli *et al.*, 2014; Videla & Ríos, 2014) y para determinar las preferencias en el mercado de automóviles (Wan *et al.*, 2015).

En Zhao *et al.* (2017) se ha desarrollado un sistema para proponer sugerencias de compra a los clientes basado en la construcción de una red de productos sustitutos y productos complementarios. Sha *et al.* (2018) han aplicado el enfoque de redes complejas para modelar las decisiones de compra en supermercados de China. En el caso de tiendas de conveniencia o franquicias, el enfoque de redes complejas se ha aplicado para entender el comportamiento de los clientes al momento de realizar sus compras, Zuo *et al.* (2018) y Valle *et al.* (2018) realizan un análisis de la cesta de supermercado empleando arboles de expansión en una tienda de abarrotes.

Las transacciones realizadas mediante tarjetas de crédito son otro ejemplo de grandes cantidades de información analizadas con enfoque de redes complejas. Hashemian *et al.* (2017) construyen la red de transacciones realizadas mediante tarjeta de crédito en España, posteriormente con la información obtenida predicen información socioeconómica de la población. En Di Clemente *et al.* (2018) el análisis de las transacciones se realiza empleando la información de la Ciudad de México.

En la literatura consultada, se observa que el empleo de redes complejas se enfoca a determinar reglas de asociación, dejando de lado la cuantificación de la relevancia de los artículos dentro del catálogo, así como la identificación de aquellos productos que funcionan como puentes (“hubs”) y alrededor de los cuales se estructuran la red de compras. En las fuentes consultadas no se hace comparación con una herramienta común en la gestión de inventarios conocida como clasificación ABC el cual particiona los artículos de acuerdo con su participación en el costo del inventario. Finalmente, el caso de estudio es un pequeño distribuidor de México para ejemplificar la aplicación del enfoque de redes complejas.

Redes

La red es una herramienta que permite representar sistemas de diversos ámbitos (Newman, 2003):

- a. Sociedad: Cuando se representan las relaciones entre personas o grupos de personas.
- b. Información: Cuando la relación entre entidades se basa en compartir algún tipo de conocimiento, por ejemplo, la red formada por las relaciones establecidas entre autores a través de las citas de trabajos científicos.

c. Tecnología: Cuando la relación implica la distribución, transporte o traslado de algún bien o recurso entre entidades, por ejemplo, una cadena de suministro o bien una red de distribución de electricidad.

d. Biología: Se construye una red con base en las reacciones químicas que se presentan en presencia de componentes como proteínas y enzimas.

Mediante el enfoque de redes se identifican aquellos nodos que tienen mayor relevancia (centralidad) dentro de la estructura, así como comunidades o bien nodos que funcionan como puentes que conectan distintas clases de nodos (Barábasi, 2016).

Una red es un sistema formado por un conjunto de elementos llamados nodos los cuales están conectados entre mediante líneas llamadas arcos. Los nodos pueden representar ciudades, un arco representa una relación que conecta a un par de nodos como los son las carreteras que conectan pares de ciudades. En ocasiones se utilizan flechas para indicar la dirección en que se mueve un flujo; en este caso se les conoce como gráficas dirigidas. Los arcos pueden tener un peso que indica un costo como lo son las distancias entre ciudades (Barábasi, 2016).

Propiedades de las redes

En este trabajo se utiliza una representación donde los nodos corresponden a los productos ofertados por el vendedor y los arcos representan la relación entre ellos que se genera cuando pertenecen a una misma compra y que se encuentra registrada en un comprobante; una red es un conjunto $G = (N, A)$ donde N representa los productos y A la relación entre ellos debido a que se adquirieron juntos.

Las redes se representan matemáticamente mediante las matrices de adyacencia, las matrices de incidencia o bien las listas de aristas. En este trabajo se construyó una lista de aristas, el producto i tiene conexión con el producto j si pertenecen al mismo comprobante, la arista será:

$$\text{Queso-Jamón} \quad (1)$$

En una red dirigida, la ruta es una secuencia de nodos conectados entre sí. Si en la ruta todos los nodos son diferentes entonces se trata de un camino; si el nodo de inicio coincide con el nodo final de la secuencia entonces se le conoce como ciclo. Existen dos tipos de indicadores que cuantifican la estructura de una red: Globales y locales (Barthelemy, 2011).

Indicadores globales de la red.

Esta clase de indicadores cuantifican propiedades de la estructura de la red, en este trabajo se utilizaron la longitud promedio del camino y el diámetro de la red (Newman, 2003).

Longitud promedio del camino y diámetro de la red.

La longitud promedio del camino es el número de nodos de las rutas más cortas de la red; el diámetro se refiere a la ruta más corta de mayor longitud en la red (Newman, 2003). El diámetro cuantifica la capacidad de comunicación entre dos nodos de la red, entre más pequeño, la red está mejor comunicada. En las redes complejas se ha encontrado que, aunque el número de nodos sea muy grande la longitud promedio entre pares de nodos es de

unos cuantos nodos, a estos sistemas se les conoce como mundo-pequeño (Wang & Chen, 2003).

Indicadores locales (centralidad)

Para cuantificar la relevancia de un nodo dentro de la estructura se emplean indicadores de locales, mejor conocidos como medidas de centralidad. En este trabajo se consideraron las siguientes medidas: grado del nodo y su distribución de probabilidad, centralidad de la intermediación y centralidad de la cercanía.

Grado de una red

Es el número de arcos que contiene un nodo. En una gráfica dirigida hay tanto grado de salida como grado de entrada (Newman, 2003; Freeman, 1977).

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

El grado promedio de una red se obtiene como sigue:

$$\bar{k} = \frac{1}{N} \sum_i k_i \quad (3)$$

En una red real el grado de un nodo se distribuye de forma heterogénea. En términos de probabilidad, p_k es la fracción de nodos con grado k , en otras palabras, es la probabilidad de que un nodo seleccionado aleatoriamente sea de grado k . En las redes reales se ha determinado que la distribución de probabilidad se ajusta a funciones de tipo potencial o bien de tipo exponencial; a diferencia de las redes construidas de forma aleatoria, en las redes reales se encuentran nodos llamados "hubs", nodos con un gran número de conexiones (Barábasi, 2016).

Coficiente de agrupamiento (clustering)

En una red si los nodos A y C, y los nodos B y C están conectados, hay una alta probabilidad de que los nodos A y C también estén conectados formando un grupo o aglomeración (cluster). Existen dos tipos índices de agrupamiento, global y local. El local mide la conectividad de los nodos con sus vecinos. El global mide el número total de triángulos cerrados en una red. El coeficiente global se obtiene con la siguiente expresión:

$$C_{\Delta} = \frac{3x \# \text{Triángulos}}{\# \text{de tripletas conectadas}} \quad (4)$$

El coeficiente de agrupamiento es la proporción de triángulos con respecto al número de tripletas de nodos conectados. Se han encontrado redes reales donde la longitud promedio de un viaje requiere atravesar por un número pequeño de nodos, estas redes se caracterizan por tener un C alto. Se recomienda calcular este índice en casos donde existen grandes variaciones en el grado de los nodos (Barthelemy, 2011).

Modularidad

Mediante la formación de comunidades dentro de una red, se identificará cómo se agrupan

las compras de los artículos, a cada comunidad se le otorga el nombre de módulo. 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.

Para el cálculo de la modularidad de la red se utilizó un algoritmo heurístico que construye comunidades pequeñas a las que agrega nuevos nodos, el proceso es iterativo y el resultado es un conjunto de comunidades dentro de la red (Blondel *et al.*, 2008).

MATERIALES Y METODOS

Con el apoyo y colaboración del negocio Cremería ISSA se contó con el acceso a su base de datos de los movimientos en inventarios y la demanda diaria de sus productos. Cremería ISSA es un negocio ubicado en la ciudad de Jaral del Progreso, Guanajuato, con venta de productos de canasta básica; los productos ofrecidos se catalogan de la siguiente forma: crema, queso, leche, bebidas gaseosas, jugos, agua, vasos, cubiertos, embutidos, productos enlatados, aderezos y condimentos, botanas, dulces y galletas, bolsas y derivados, pan, chiles y salsas, yogurt (Estrada, 2018).

Dentro del sistema de control de ventas, se utilizó la base de datos con los productos ofrecidos; en este estudio se analizó el movimiento de salidas (ventas realizadas) de seis días consecutivos del mes de febrero (Figura 1). Se asume que cada nodo representa un producto que fue vendido durante el periodo estudiado, dando un total de 272 nodos, los productos vendidos que aparecen en el comprobante son de la misma transacción y se genera una conexión no dirigida entre los integrantes (nodos) de dicho comprobante.

Las relaciones entre artículos se descargaron en una lista de pares de nodos conectados, a partir de esta lista de relaciones, con el paquete Gephi se construyó la respectiva red, (Bastian *et al.*, 2009); posteriormente, los cálculos de las medidas de centralidad de la red se realizaron igualmente empleando dicho paquete; a partir de las medidas de centralidad se obtienen conclusiones acerca de la relevancia de un producto dentro de la estructura. Como paso final se comparó una de las métricas de redes contra la clasificación de inventario ABC.

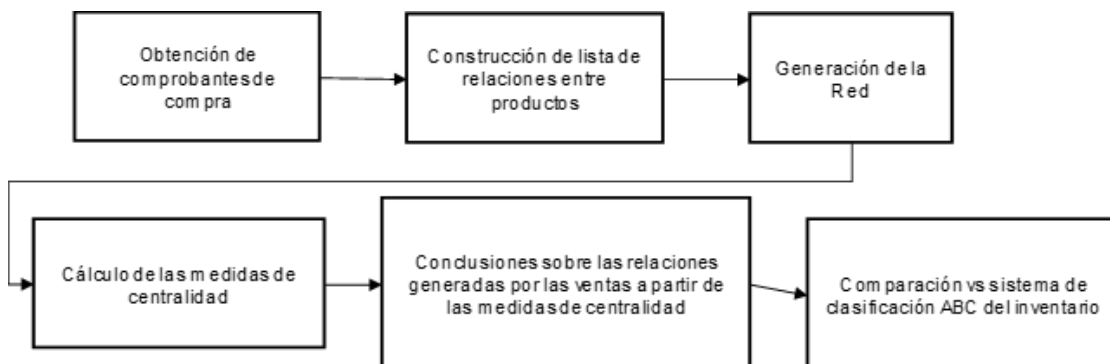


Figura 1. Pasos del método para la construcción y análisis de la red.

En la Tabla 1 se muestra un detalle del precio por unidad y las unidades en inventario al momento de realizar el análisis. Los artículos se encuentran codificados para no mostrar marcas.

Tabla 1. Ejemplo de productos ofertados, inventario y precio de venta (los nombres se encuentran codificados).

Descripción	Precio Venta (US\$)	Inventario	Descripción	Precio Venta (US\$)	Inventario
		(unidades)			(unidades)
Crema Natural	1,80	14,2	Jamón Ame Mex 5,4	3,00	5,32
Mayonesa Suelta	2,74	32,77	Salchicha Go	1,07	20,94
Jamón CampT. 4.8	2,74	30,74	Jamón Espaldilla L 3,5	1,71	25,52
Queso Asadero PN	3,00	14,73	Jamón De Cuad 3,5 kg	2,57	11,4
Huevo Sn Jn	1,63	8,78	Chiles Suelos	1,50	0,08
Queso Asadero M Suelto	3,85	5,02	Queso Bio Amarillo Suelto	2,14	7,35
Salchicha APvo	1,50	11,49	Catsup Suelta	0,86	12,1
Queso Asadero DN CH	3,85	15,03	Queso Asadero Mj 250	0,99	26

RESULTADOS Y DISCUSIÓN

Con el método descrito anteriormente se obtuvo una red con 272 nodos y 1080 aristas, sin repetir y sin pesos.

Centralidad del grado

El grado promedio es $7,941 \approx 8$, el grado promedio indica que en las transacciones realizadas durante el período de estudio los productos formaron en promedio conexiones con otros 8 productos del catálogo, en otras palabras, cuando un cliente compra un producto, existen 8 posibilidades que pueden incluirse en la compra. En la Figura 2, se muestra un detalle del grafo resultante (el tamaño de los nodos es proporcional al número de arcos que lo conectan), el nodo de mayor tamaño corresponde a la crema natural (#64), le siguen en tamaño la mayonesa (#141) y a continuación una marca de jamón (#109).



Figura 2. Red generada a partir de los comprobantes de compra, el tamaño del nodo es proporcional al número de conexiones (grado)

Destaca que la crema natural tiene conexión con $(75/ 272) \times 100 = 27,57\%$ de los productos ofertados por el almacén; la mayonesa suelta se relaciona con 24,63% de los 272 productos considerados (Tabla 2). Cabe señalar que por el momento no se está determinando si existe una motivación común para adquirir dos productos, pero se aprecia por otra parte que existen artículos en el catálogo alrededor de los cuales se forma una estructura ordenada y que no es un arreglo aleatorio (Barábasi, 2016; Hearnshaw & Wilson, 2013).

Tabla 2. Artículos y grado ordenados de mayor a menor (se muestran los 10 primeros).

Id	Producto	Grado	Id	Producto	Grado
64	Crema Natural	75	187	Queso Asadero MJ	49
141	Mayonesa Suelta	67	211	Salchicha An Pavo	46
109	Jamón Camp T. 4.8	66	107	Jamón AmMex 5.4	40
188	Queso Asadero PN	61	184	Queso Asadero DC	40
104	Huevo Sn Jn	55	215	Salchicha Go	36

Por otra parte, se detectó que 9,19% de los productos aparecen como única transacción o compras aisladas (grado igual a 0), se encuentran en la periferia del grafo en forma de islas:

pasas, yogurt, chocolate en polvo y sal (Estrada, 2018).

Al eliminar el efecto de las sub-redes para determinar el grado promedio del componente principal o componente gigante, filtrando los datos eliminando las islas, la nueva red tiene 230 artículos con 1061 aristas siendo el grado promedio de 9226 con desviación estándar de 9,82 y un coeficiente de correlación de 1,32 que indica una variabilidad moderada y una distribución heterogénea de valores de k (Figura 3).

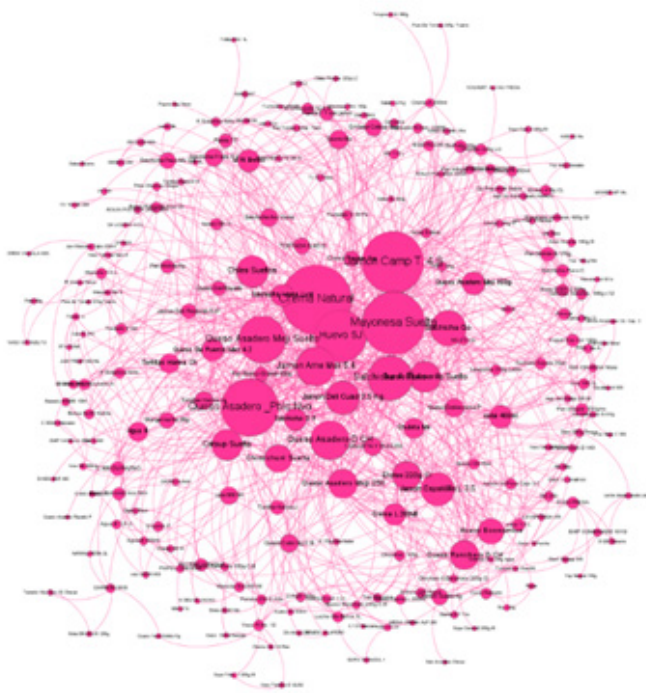


Figura 3. Detalle del componente gigante. El tamaño del nodo es proporcional a valor de k

Las gráficas de frecuencia y dispersión muestran la existencia de una correlación entre el grado del nodo (k) y la probabilidad (p_i) de ser seleccionado aleatoriamente (Figura 4 y Figura 5).

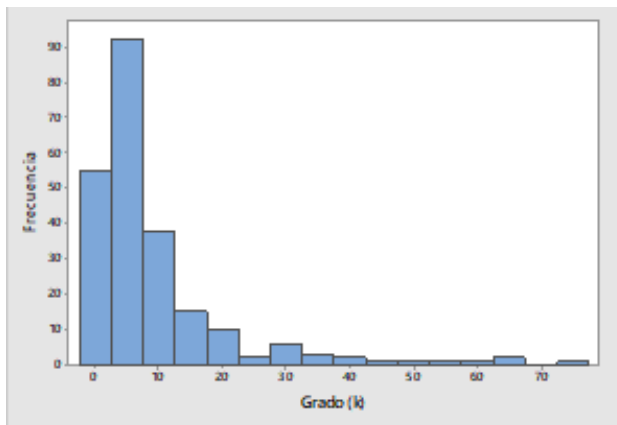


Figura 4. Gráfica de frecuencias del grado de los nodos (Componente gigante)

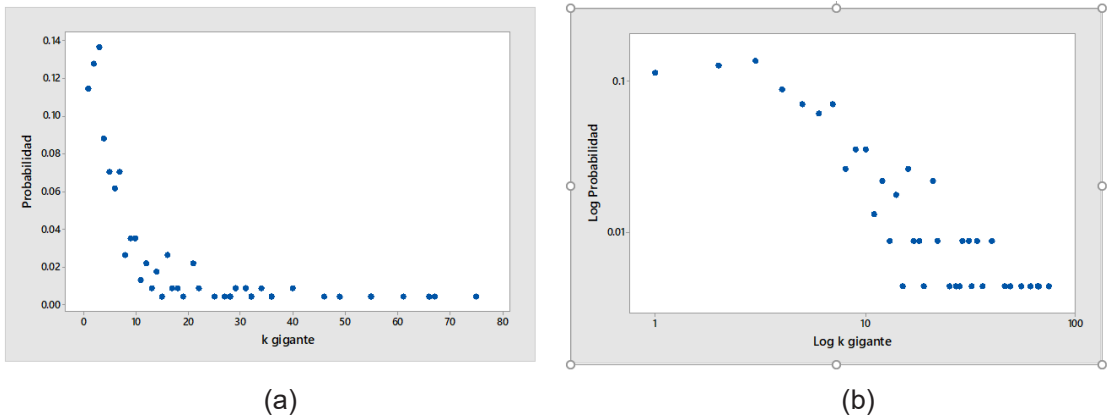


Figura 5. Probabilidad de k vs grado del nodo: a. Escala decimal; b. Escala logarítmica.

Cabe señalar que la gráfica de dispersión muestra una meseta para valores de $k > 9,22 \approx 10$, debido a que en dicha región existe un solo nodo para varios casos de k y refleja una baja probabilidad de que un artículo se relacione con más de 10 productos del catálogo. Se construyeron nuevos intervalos de clase aplicando el método propuesto por Barábasi (2016) y posteriormente se llevó a cabo un ajuste de tipo no lineal, el modelo propuesto es $P(k) = \beta_0 k^{\beta_1} \exp(-\beta_2 k)$ (Tabla 3). En este caso se trabaja con el componente gigante.

Tabla 3. Intervalos de clase de k y frecuencia y grado medio

k	Frecuencia	k	Frecuencia
1	26	16	24
2	60	32	12
4	61	64	3
8	44		

Los cálculos se realizaron con el paquete Minitab, el algoritmo utilizado para la regresión es el Marquardt llevando a cabo 31 iteraciones, los valores iniciales son $\beta_1 = \beta_2 = \beta_3 = 0,1$ Donde $\beta_1 = 0,1817$, $\beta_2 = 0,7689$ y $\beta_3 = -0,1763$. La función exponencial predice la probabilidad de que un producto (nodo en la red) seleccionado al azar pertenezca la clase k (Tabla 4 y Figura 6).

Tabla 4. Datos reales vs Predicción del modelo

k	Probabilidad	Predicción
1	0,1130	0,1523
2	0,2608	0,2176
4	0,2652	0,2607
8	0,1913	0,2194
16	0,1043	0,0912
32	0,0521	0,0092
64	0,0134	0,00005

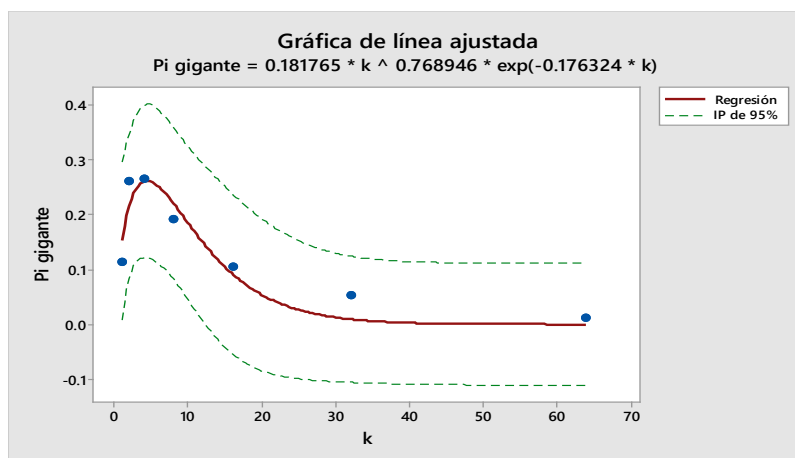


Figura 6. Curva de regresión ajustada del grado del nodo e intervalo de confianza de la predicción.

A partir de la función de probabilidad se aprecia que: 1. Si bien el catálogo es de 272 productos, al eliminar las islas en el componente gigante quedan 230 productos (84,56% del total), dentro del componente gigante la probabilidad de seleccionar un artículo con grado $k \leq 10$ es 0,8304; 2. La probabilidad de seleccionar un artículo con $k > 10$ es 0,1696; 3. La probabilidad de seleccionar un artículo con $k \geq 64$ es 0,01304, es decir, existe un conjunto de artículos que los clientes compran frecuentemente junto con el resto de los productos del catálogo y son en este caso la “Crema natural”, “Mayonesa suelta” y “Jamón Camp T 4.8”.

Por ejemplo, en este negocio, si un cliente compra “Crema Natural” existen $k = 75$ alternativas para complementar la compra, con la “mayonesa suelta” existen 64 alternativas para complementar la compra ($k = 64$), por otra parte, si un cliente lleva pasas ($k=3$) sólo existen tres alternativas (coco rallado, jamón y levadura R).

En la red generada por las compras los 3 artículos mencionados se encuentran en la cola de la distribución (llamados “hubs” o nodos de conexión), y alrededor de estos se forma la estructura ordenada; una red como ésta es sensible a la desaparición (fallas) de los “hubs”, equivalente a la falta del correspondiente artículo, lo que puede modificar el comportamiento de un cliente (Navarro *et al.*, 2018). Por el momento el trabajo se limita a la identificación de dichos artículos relevantes,

Diámetro de la red, longitud promedio del viaje y coeficiente de agrupamiento (clustering)

El “viaje promedio” dentro de la red gigante es 2,69 y el diámetro de la red es de 5, es decir, la ruta más corta desde los nodos más alejados de la red es de 5 nodos. Tanto el diámetro como la ruta promedio más corta indican que las compras de los clientes forman un arreglo compacto a pesar de que el número de productos del catálogo es 272 artículos. El valor promedio del coeficiente de agrupamiento es 0,656, en concordancia con Raeder y Chawla (2009); Videla y Ríos (2014).

Las redes de mundo pequeño se caracterizan porque sus equivalentes de tipo aleatorio tienen

un coeficiente de agrupamiento menor, para efectos de comparación mediante el paquete Gephi, se generó un grafo aleatorio con el mismo número de nodos y una probabilidad promedio de 0,05 de conexión, el resultado es un grafo con 1862 aristas y un coeficiente de agrupamiento promedio de 0,028. Comparando los valores se aprecia que C_{ventas} cumple con las características de una estructura de mundo pequeño.

En la Tabla 5, se muestra un resumen de los resultados obtenidos en este trabajo sobre la estructura de la red, resalta que sólo el trabajo de Raeder y Chawla (2009) reporta un grado medio menor al reportado en Wan *et al.* (2015) y en este trabajo. Por otra parte, la función de probabilidad que es relevante para determinar el tipo de red sólo se reporta en dos trabajos. En el caso del coeficiente de agrupamiento, en todos los casos se obtiene un resultado que descarta que la red forme una estructura aleatoria.

En este trabajo la distribución exponencial de la red indica que se trata de un sistema en donde conforme para el tiempo, la probabilidad de que un nodo establezca una nueva conexión es baja, en otras palabras, en esta tienda se trata de un sistema que ha alcanzado un estado donde la probabilidad de que un artículo establezca una nueva conexión es baja y su efecto no será significativo en las métricas de la estructura (Amaral *et al.*, 2000).

Tabla 5. Resultados de la estructura

Autor	Tipo de negocio	# productos	Grado medio	Diámetro	Coef. agrupamiento	Función de probabilidad
Raeder y Chawla (2009)	Tienda de conveniencia de Universidad de Notre Dame	2248	2,41	5	0,518	Potencial
Wan <i>et al.</i> (2015)	Sector automotriz	No reporta	No reporta	No reporta	No reporta	No reporta
Valle <i>et al.</i> (2018)	Pequeño distribuidor	3240	No reporta	No reporta	No reporta	No reporta
Kim <i>et al.</i> (2012)	Pequeño distribuidor	255	8,58	4	0,155	Potencial
Videla y Ríos (2014)	Cadena de supermercados	5359	No reporta	No reporta	No reporta	No reporta
Presente trabajo	Pequeño distribuidor	272	9,226	5	0,656	Exponencial

Modularidad.

En este trabajo, la obtención de las comunidades se realizó calculando la modularidad de la red y se utilizaron por el momento los valores sugeridos por el paquete Gephi.

Se obtuvieron 39 comunidades, en cada una quedaron agrupados varios artículos, de la inspección de estas comunidades se observa que es viable formar “paquetes” con algunos de estos productos y que se pueden ofrecer a los clientes; en la Tabla 6 se muestra dos ejemplos de sugerencias.

Tabla 6. Paquetes con productos sugeridos.

Paquete	Productos	Módulo
1	Atún, ensalada de verduras, galletas saladas, mayonesa	5
2	Aceite, queso, leche, huevo	6

En la Figura 7, se muestra el detalle de la comunidad 5, en esta se aprecian por ejemplo las relaciones correspondientes al atún, la mayonesa, la ensalada de verduras y las galletas saladas, de la experiencia adquirida por los administradores de la tienda se ha observado que son ingredientes de un alimento preparado y por lo tanto pueden ofrecerse como paquete; otro ejemplo es el caso del paquete 2 se trata de una compra de canasta básica urbana definida por CONEVAL (Consejo Nacional de Evaluación de la Política de Desarrollo Social, 2018).

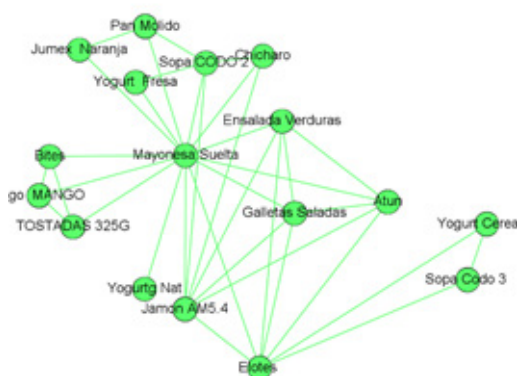


Figura 7. Detalle de la comunidad 5.

Comparación contra el sistema ABC de clasificación del inventario

El sistema de clasificación del inventario ABC es una herramienta utilizada para catalogar los productos, el objetivo es ubicar al 20% de los productos a los que les corresponde el 70-80% del valor monetario almacenado, para efectos de control, los responsables de la toma de decisiones vigilan de cerca los productos del tipo “A” ya que representan la mayor parte del costo total. En el otro extremo se encuentran los productos de la clase “C” los cuales consisten en productos cuyo valor monetario es alrededor del 5% y abarca hasta el 50% de los productos de un almacén, en este caso su seguimiento es de tipo periódico; el sistema ABC puede considerarse como una manera de identificar los artículos relevantes de un catálogo.

En la Tabla 7, se muestra la clasificación ABC para los productos del almacén, los productos de la clase “A” representan el 65,75% del valor monetario de las existencias.

Tabla 7. Clasificación ABC del sistema.

Clasificación	Cantidad de productos	Valor monetario del inventario(US\$)
A	154	1894,78
B	232	642,87
C	386	343,98
Total general	772	2881,62

Se encontró que en la clasificación de productos tipo “A” aparecen los artículos de la Tabla 8; se debe señalar que se detectó una discrepancia entre ambos métodos: los chiles secos son un producto con $k= 32$ sin embargo quedó catalogado como producto de la clase C, la explicación es que este producto tiene un valor monetario bajo y por lo tanto su contribución al valor monetario total tiene como resultado una clasificación como producto de revisión periódica.

Tabla 8. Clasificación ABC y grado k del artículo (se muestran los primeros 14 ordenados de mayor a menor).

Descripción	Grado	Precio Venta US\$	Inventario	Clasificación
Crema Natural	75	1,80	14,2	A
Mayonesa Suelta	67	2,74	32,77	A
Jamón Camp 4,8	66	2,74	30,74	A
Queso Asadero P Nv	61	3,00	14,73	A
Huevo SJ	55	1,63	8,78	A
Queso Asadero M Suelto	49	3,85	5,02	A
Salchicha An Pavo	46	1,50	11,49	A
Queso Asadero DC	40	3,85	15,03	A
Jamón Ame Mex 5,4	40	3,00	5,32	A
Salchicha Go	36	1,07	20,94	A
Jamón Espaldilla L 3,5	34	1,71	25,52	A
Jamón Deli Cuad 3,5 kg	34	2,57	11,4	A
Chiles Sultos	32	1,50	0,08	C
Queso Bi Amarillo Suelto	31	2,14	7,35	A

En la Figura 8, se aprecian en color rojo los productos de la clase “A” encerrados en un círculo, por otra parte, el tamaño del nodo corresponde al valor de k ; la clasificación ABC y la centralidad del grado coinciden en clasificar los productos.

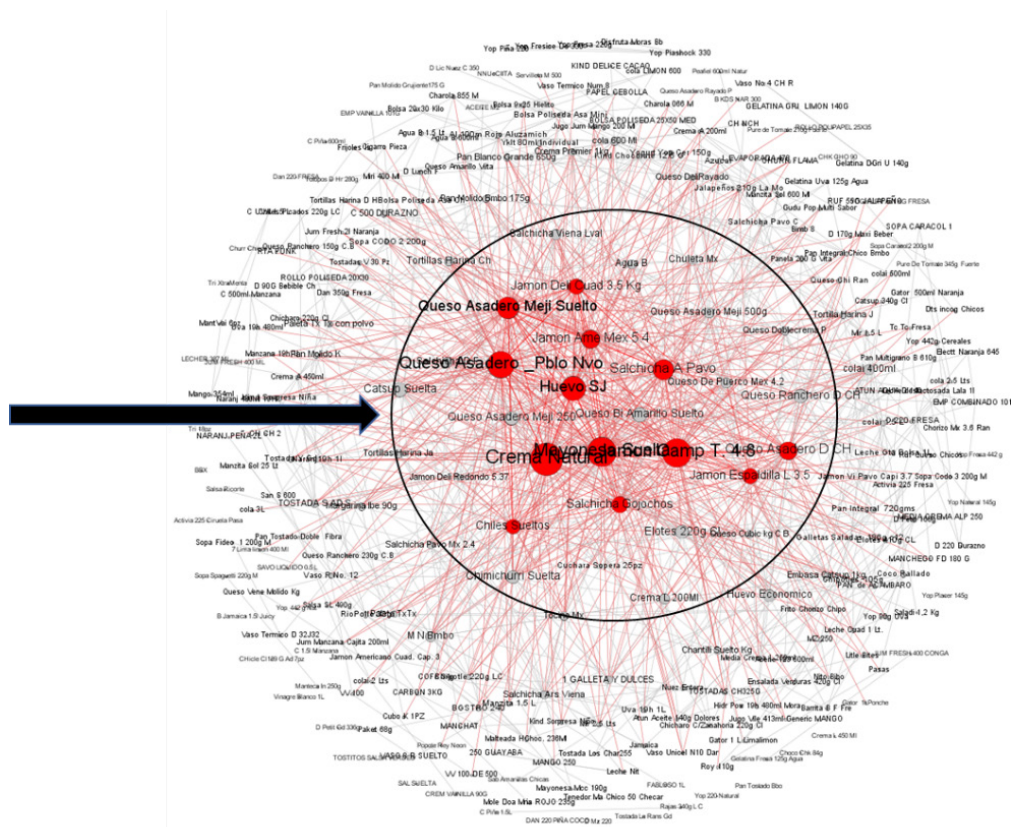


Figura 8. Red de transacciones. El círculo agregado encierra los productos de la clase A

CONCLUSIONES

Las características de esta red corresponden a una estructura conocida como mundo-pequeño; además, la probabilidad de que un artículo establezca una nueva conexión con otro producto del catálogo es baja. Existe también un conjunto de artículos alrededor de los cuales las compras forman un arreglo ordenado.

Mediante el valor de modularidad se identificaron comunidades de productos, del análisis se obtuvieron ejemplos de “paquetes” de artículos que pueden ofertarse juntos con la intención de incentivar las ventas.

Al comparar este paradigma de análisis contra el sistema de clasificación ABC del inventario, el resultado muestra una clasificación similar de los productos detectándose únicamente una discrepancia con un producto, la razón es que el sistema ABC se basa en la contribución monetaria de cada producto al valor total del inventario; el enfoque de redes se basó en el

número de conexiones establecidas (grado del nodo) con otros productos.

Este trabajo muestra una forma de aprovechar la información generada en los negocios derivada de la centralidad del grado y la modularidad; dado que el caso de estudio es un negocio de abarrotes se pone al alcance un método de análisis de las ventas con un enfoque de redes complejas. En caso de que alguna empresa de esta naturaleza empiece a ofrecer sus productos por Internet, le servirá para armar estrategias de oferta que de alguna manera complementen las necesidades de sus clientes y que incrementen sus ventas.

Trabajos posteriores pueden explorar de forma puntual las redes formadas por las familias de artículos, también comparar los resultados con la técnica conocida como reglas de asociación e identificar si existen similitudes entre ambos paradigmas. Finalmente convendrá trabajar con bases de datos que correspondan a un horizonte temporal mayor a un mes.

REFERENCIAS

AGRAWAL, R., IMIELINSKI, T. y SWANI, A. (1993) Mining association rules between sets of items in large data bases. *ACM SIGMOD Record* [en línea], 22(2), 1 – 10. Disponible: <https://doi.org/10.1145/170035.170072>

AKTAS, E., y MENG, Y. (2017). An Exploration of Big Data Practices in Retail Sector. *Logistics* [en línea], 1(2), 1-28. Disponible: <https://doi:10.3390/logistics1020012>

AMARAL, L.A., SCALA, A., BARTHELEMY, M. STANLEY, E. (2000). *Classes of small-world networks*. Proceedings of National Academy of Sciences of USA, 97(21), 11149-11152. Disponible: <https://doi:10.1073/pnas.200327197>

BARÁBASI, A. *Complex Networks*. 152 – 153. Cambridge: Cambridge University Press. 2016
BARTHELEMY, M. (2011). Spatial Networks. *Physics Reports* [en línea], 499(1-3), 1-101. Disponible: <https://doi:10.1016/j.physrep.2010.11.002>

BASTIAN, M., HEYMANN, S. JACOMY, M. (2009). Gephi: an open source software for exploring and manipulating networks. En: Proceedings Third international AAAI conference on weblogs and social media [en línea]. pp.361-362. 2009. Disponible en: <https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154/1009>

BLONDEL, V., GUILLAUME, J.L., LAMBIOTTE, R., y LEFEBVRE, E. (2008) Fast unfolding of communities in large networks. *Journal of Statistic Mechanics: Theory and Experiments* [en línea], 1-12. Disponible: <https://doi:10.1088/1742-5468/2008/10/P10008>

CEBALLOS – CARDONA, Y., ROBLEDO – GIRALDO, S. y DUQUE – MÉNDEZ, N. (2016). Network marketing en microempresas. Incidencia de la estrategia de referidos utilizando redes complejas y modelado basado en agentes. *Revista Vínculos* [en línea], 13(1), 6-16. Disponible en: <https://revistas.udistrital.edu.co/index.php/vinculos/article/view/11659/12790>

CONSEJO NACIONAL de EVALUACIÓN de la POLÍTICA de DESARROLLO SOCIAL. CONEVAL. *Evolución del valor de la canasta alimentaria*. 2018. [En línea]. Disponible en: <https://www.coneval.org.mx/Medicion/MP/Paginas/Lineas-de-bienestar-y-canasta-basica.aspx>

DI CLEMENTE, R., LUNEGO-OROZ, M., TRAVIZANO, M., Xu, S., VAITLA, B. y GONZÁLEZ, M. (2018). Sequences of purchases in credit card data reveal lifestyles in urban populations. *Nature Communications* [en línea], 1-8. Disponible: <https://doi:10.1038/s41467-018-05690-8>

ESTRADA, A. *Análisis de sistemas de servicios aplicando métricas del análisis de redes sociales para el diagnóstico y una mejor toma de decisiones*. Tesis de Maestría. HERNÁNDEZ – GONZÁLEZ, Salvador (dir.) Celaya: Tecnológico Nacional de México en Celaya. 2018.

FREEMAN, L. (1977). A set of measures of centrality based on betweenness. *Sociometry* [en línea], ISSN 2325-7938, 40(1), 35-41. Disponible: <https://doi:10.2307/3033543>

HASHEMIAN, B., MASSARO, E., BOJIC, I., MURILLO – ARIAS, J., SOBOLEVSKY, S. y RATTI, C. (2017). Socioeconomic characterization of regions through the lens of individual financial transactions. *PLOS One* [en línea], 12(11), 1-20. Disponible: <https://doi:10.1371/journal.pone.0187031>

HEARNSHAW, E. y WILSON, M. (2013). A complex network approach to supply chain network theory. *International Journal of Operations & Production Management* [en línea], 33(4), 442-469. Disponible: <https://doi:10.1108/01443571311307343>

HUANG, Z., ZENG, D. y CHEN, H. (2007). Analyzing consumer-product graphs: Empirical findings and applications in recommender systems. *Management Science* [en línea], 53(7), 1146-1154. Disponible: <https://doi:10.1287/mnsc.1060.0619>

KIM, H., KIM, J., y CHEN, Q. (2012). A product network analysis for extending the market basket analysis. *Expert Systems and Applications* [en línea], 39(8), ISSN 7403-7410. Disponible: <https://doi:10.1016/j.eswa.2012.01.066>

NAVARRO, O., FERRER, W. y BURGOSO D. (2018). La calidad como factor estratégico en el desarrollo competitivo de las Pequeñas y medianas empresas. *Universidad y Sociedad* [en línea], 10(2), 171-174. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202018000200171

NEWMAN, M. E. (2003). The structure and function of complex networks. *SIAM Review* [en línea], 45(2), 167-256. Disponible: <https://doi:10.1137/S003614450342480>

PENNACCHIOLI, D., COSCIA, M., RINZIVILLO, S., GIANOTTI, F. y PEDRESCI, D. (2014). The retail market as a complex system. *EPJ Data Science* [en línea], 3(33), 1-27. Disponible: <https://doi:10.1140/epjds/s13688-014-0033-x>

RAEDER, T. y CHAWLA, N. Modeling a store's product space as a social network. En: *Proceedings of International Conference on Advances in Social Network Analysis and Mining* [en línea]. Athens: Greece, 2009. Disponible: <https://doi:10.1109/ASONAM.2009.53>

RODRÍGUEZ, F., FLORES, L. y DÁVILA, E. (2016) Inteligencia de negocios y minería de datos aplicado a la industria refresquera. *Research in Computing*

Science [en línea], 126, 63–71. Disponible en: <https://pdfs.semanticscholar.org/ed31/f60175e0b6d3efd82789b1778cfbb363c369.pdf>

Sha, Z., Huang, Y., Sophia Fu, J., Wang, M., Fu, Yan., Contractor, N. y Chen, W. (2018). A network-based approach to modeling and predicting product coconsideration relations. *Complexity* [en línea], 2018, 1-13. Disponible: <https://doi:10.1155/2018/2753638>

VALLE, M., RUZ, G. y MORRÁS, R. (2018). Market basket analysis: Complementing association rules with minimum spanning trees. *Expert Systems and Applications* [en línea], 97(1), 146-162. Disponible: <https://doi:10.1016/j.eswa.2017.12.028>

VIDELA, I., y RÍOS, S. (2014). Extending market basket analysis with graph mining. *Expert Systems and Applications* [en línea], 41(4), 1928-1936. Disponible: <https://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 AutoMarket Using Choice Modeling and Network Analysis. *SAE International Journal of Materials Manufacturing* [en línea], 8(3), 1-10. Disponible: <https://doi:10.4271/2015-01-0468>

WANG, X. y CHEN, G. (2003) Complex networks: small world, scale-free and beyond. *IEEE circuits and systems magazine* [en línea], 3(1), 6-20. Disponible: <https://doi:10.1109/MCAS.2003.1228503>

ZHAO, T., McAULEY, J., LI, M. y KING, I. Improving Recommendation Accuracy using Networks of substitutable and complementary products. En: *International Joint Conference on Neural Networks* [en línea]. Anchorage: AK, 2017. Disponible: <https://doi:10.1109/IJCNN.2017.7966315>

ZUO, Y., YADA, K., LI, T. y PHILIP, C. Application of Network Analysis Techniques for Customer In-store Behavior in Supermarket. En: *Proceedings IEEE International Conference on Systems, Man, and Cybernetics* [en línea]. Miyazaki: Japan, 2018. ISBN 978-1-5386-6650-0. Disponible: <https://doi:10.1109/SMC.2018.00322>