



**Universidad del Desarrollo**  
Facultad de Ingeniería

**ANÁLISIS DE REDES PARA MARKET BASKET**  
Experiencia Practica con Big Data en una Empresa de Retail Chilena

**POR: MARTÍN ARNALDO GRUNERT GÓMEZ**

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad del Desarrollo para optar al grado académico de Magíster en Data Science

**PROFESOR GUÍA:**

**Dr. HEETAEE KIM**

**Noviembre 2020**

**SANTIAGO**

*Para mi padre, Mauricio Grunert Ayarza, a quien jamás podré retribuirle su esfuerzo por darme una buena educación, pero sobre todo, darme un amor incalculable.*

*Y mi primo Mario Jara, por ser mi guía, consejero y hermano.*

## AGRADECIMIENTO

Comienzo agradeciendo a mi profesor guía, Heetae Kim, quien con su profesionalismo y paciencia me acompaño en este proceso.

A los compañeros y amigos del magister, Diego e Ibar, gracias por enseñarme todo lo que se y acompañarme durante este proceso.

Por último, a mi familia, por todo el amor y cariño que me han entregado durante toda la vida.

## TABLA DE CONTENIDO

<b>RESUMEN.....</b>	<b>1</b>
<b>1. INTRODUCCIÓN .....</b>	<b>2</b>
<b>2. TRABAJO RELACIONADO .....</b>	<b>5</b>
<b>3. HIPÓTESIS Y OBJETIVOS .....</b>	<b>6</b>
<b>4. DATOS Y METODOLOGÍA .....</b>	<b>7</b>
4.1. DATOS .....	7
4.2. METODOLOGÍA .....	10
<b>5. LIMPIEZA Y PREPARACIÓN DE DATOS .....</b>	<b>16</b>
<b>6. RESULTADOS.....</b>	<b>16</b>
<b>7. CONCLUSIONES.....</b>	<b>22</b>
<b>BIBLIOGRAFÍA .....</b>	<b>24</b>

## **Resumen**

En este estudio se propone una metodología con análisis de redes para realizar un *Market Basket Analysis* a nivel de producto en una empresa de retail chilena. Se utiliza un algoritmo de detección de comunidades para estudiar los grupos que se generan durante tres trimestres del 2020 y como ha cambiado la red temporal a través del tiempo.

La metodología propuesta, contempla la creación de una red producto-producto, que va desde el procesamiento de los datos transaccionales, pasando por una matriz de co-ocurrencia, hasta la creación de un grafo con la aplicación de algoritmo de detección de comunidades.

Se observa que las comunidades del primer trimestre del 2020 tienen un cambio significativo comparado con el segundo y tercer trimestre por los efectos de la pandemia y el retiro de las AFP. También, es interesante observar que cada comunidad tiene una composición interpretable a las distintas ocasiones de consumo de los clientes.

Por último, a pesar de lograr una modularidad cercana a la literatura, esta sigue siendo baja y da para concluir que existe un patrón no modular en la red, donde los patrones de consumo de los clientes son todos distintos.

## 1. Introducción

Durante los últimos 10 años y de manera muy marcada por la última pandemia, el comercio vía e-commerce ha crecido de manera exponencial. Según estudios (Bhatti, Akram, Basit, Khan, Naqvi, Bilal, 2020), durante este año el 39% de las personas ha reemplazado el comercio físico por el electrónico. Esto conlleva no solo un mayor desafío logístico, si no también una oportunidad para las empresas de retail para poder ofrecer más productos con la gran cantidad de data de navegación y transaccional que se genera hoy en día. Esta tendencia viene desde principio de los años 90's, cuando Agrawal (Agrawal, Imielinski, & Swami, 1993; Agrawal & Srikant, 1994) estudió el uso de grandes volúmenes de datos transaccionales y data de consumidores en distintas organizaciones. Pero el crecimiento de este mercado y búsqueda de información con la data que este genera se ha visto facilitado por el avance en sistemas computacionales, tanto para almacenar estos grandes volúmenes de base de datos, como para procesarlos y obtener modelos analíticos de estos. Áreas de CRM y web analytics, que se encargan de tener relación online con cliente y analizar sus datos, son cada vez más comunes en empresas de retail. Uno de los análisis más comunes hoy en día es el de *Market Basket*, que estudia la composición de la canasta de compra que un cliente en una sola compra (G. Russell and A. Petersen, 2000). La idea general, es encontrar asociaciones de productos que tienden a ser vendidos juntos (Agrawal & Srikant, 1994). Estos patrones de consumo después pueden servir a áreas de CRM para realizar envíos complementarios a los productos que usualmente consumen los clientes, como también pueden ser utilizado por el área de e-

commerce para ofrecer productos complementarios a la canasta de productos que tiene el cliente en su momento u ofrecer productos apropiados a clientes los cuales no se tiene mucha información.

No obstante, una de las grandes limitaciones de este tipo de análisis es encontrar los patrones de consumo con un alto volumen de transacciones, como cadenas de retail. Uno de los algoritmos mas conocidos es el Apriori (Agrawal & Srikant, 1994), sin embargo, dada su estructura es deficiente computacionalmente, donde la ejecución de este algoritmo requiere muchos recursos computacionales (memoria RAM) y un tiempo significativo de ejecución, y sus resultados son pobres. Es por esto que se propone realizar este tipo de análisis con técnicas de análisis de redes, utilizando algoritmos de detección de comunidades, para aislar asociaciones de productos significativas.

Durante los últimos años se ha vuelto cada vez más popular utilizar técnicas de análisis de redes para estudiar sistemas complejos, teniendo aplicaciones tan diversas como el estudio de vínculos comerciales en la biotecnología industrial (Sheldon Krimsky, James G. Ennis and Robert Weissman, 1991), la captura de Saddam Hussein (Brian J. Reed & David R. Segal, 2006) y proceso de desarrollo de softwares (Davis, Hossain, & Murshed, 2007). En nuestro caso, el análisis de redes para recomendaciones de productos en la industria del retail ha sido utilizada de diversas maneras, en la siguiente sección se ahondará en eso.

En este estudio, se utiliza data transaccional de una empresa de retail, tanto de web como de tienda, haciendo un análisis de redes de productos, lo cual ayudará a estudiar los

productos que son comprados juntos en la cadena comercial. Se utilizan distintas jerarquías de productos para ver cual genera resultados más consistentes, creando comunidades que son comparadas en el tiempo y realizando distintos filtros (*Thresholds*) en la red para buscar relaciones mas significativas.



## 2. Trabajo Relacionado

El presente estudio toma como referencia tres estudios relacionados al análisis de redes aplicado al *market basket* y uno descriptivo de esta técnica.

Sobre *market basket*, (Gupte & Mamtora, 2014) estudian esta técnica con algoritmos de asociaciones de reglas, que se utilizan comúnmente para encontrar asociaciones en altos volúmenes de datos. Acá, se estudian los patrones de consumo de los clientes para encontrar productos que se compran juntos y así realizar análisis para el negocio.

El primero de ellos es el de (Raeder & Chawla, 2009). En este, el grafo se construye a partir de productos linkeados por la misma boleta. Este tipo de red se llama *Co-Purchased Product Network (CPN)*. Para encontrar relaciones de productos usa algoritmos de detección de comunidades, tomando la modularidad como variable a optimizar. Este estudio fue uno de los primeros en tener un enfoque analítico usando teoría de grafos para Market Basket Analysis.

El siguiente estudio relacionado es el de (Kim, Kim & Chen, 2012), que generan un *Market Basket Analysis*, pero desde el punto de vista del cliente, siendo los nodos clientes y los enlaces los productos que compraron en común. Estos crean una red bipartita de cliente producto, desde donde nace la red del primer estudio, *Co-Purchased Product Network (CPN)*, pero también crean otro tipo de red llamado *Market Basket Network (MBN)*, donde se crea una red a partir de la relación de todos los productos usando análisis de redes, pero con definiciones de *Association Rules* (reglas de asociación). La gran limitación es que solo es creada a partir de clientes identificados, pero gran parte de las

transacciones del retail son con clientes no identificados, perdiendo información relevante. También es relevante destacar que el *threshold* que utilizan para filtrar la red densa es el promedio de enlaces de toda la red.

Por último, (Videla-Cavieres & Ríos, 2014) generan también una red de productos, pero para encontrar comunidades, utilizan algoritmos de *Overlapping Community Detection*. Para la limpieza de la red, toman como referencia los dos papers anteriores, pero generan una metodología más consistente llamada *three heavy edges threshold (tthet)*, que, a pesar de ser más efectiva en su contexto, no es útil para este estudio.

### **3. Hipótesis y Objetivos**

#### 3.1 Hipótesis de Trabajo

Es posible realizar un *Market Basket Analysis* con técnicas de análisis de redes y algoritmo de detección de comunidades, usando transacciones de una empresa de retail. Construyendo una matriz de co-ocurrencia se puede analizar las distintas asociaciones de productos y convertirla en un grafo. Todo esto, con herramientas de *big data*, para procesar y construir esta red, siendo capaz de estudiar las comunidades que esta red genera, de manera tanto estática como a través del tiempo.

#### 3.2 Objetivo General

El objetivo general de este estudio es mejorar el sistema de recomendación de productos complementarios.

### 3.3 Objetivos Específicos

- i) Probar y desarrollar método de análisis de redes para un *Market Basket Analysis* en una empresa de retail.
- ii) Analizar la estructura de la red con las transacciones para obtener conclusiones de esta, integrando métodos de detección de comunidades.
- iii) Analizar cambios de patrones de consumo con análisis de redes temporales en los distintos periodos de tiempo.

## 4. Datos y Metodología

### 4.1. Datos

Se utiliza datos de un supermercado chileno que esta presente de manera física en todas las regiones del país y también de manera online.

Para el estudio se utilizan 9 meses con alrededor de 12 millones de transacciones, 3 millones de clientes y 5 mil SKU 's<sup>1</sup> distintos.

Los datos del retail tienen una jerarquía como en casi todas las empresas con diversos tipos de productos. Existe una primera división entre los Productos *Food* y *Non-Food* y

---

<sup>1</sup> SKU: Stock Keeping Unit

baja a partir de distintas características. El último nivel es la subclase, mas abajo de eso está la marca y el SKU, pero como hay muchas marcas que se cruzan entre subclases, se considera esta como el ultimo nivel. ¿porque no nos interesa el SKU? Por que muchas veces este se diferencia con otros SKU 's por el formato, o sea, el mismo producto, pero de un litro, 200 c.c., 500 c.c., etc.

La figura 1 muestra la jerarquía de los productos con cinco niveles, donde cada nivel pertenece a su predecesor. Esto es importante ya que se realizaron pruebas con tres niveles distintos, clase, subclase y subclase – marca, donde finalmente se optó por subclase, ya que clase no capturaba muy bien las asociaciones y subclase – marca generaba información irrelevante, ya que no nos importan las asociaciones por marca, si no por producto en sí.

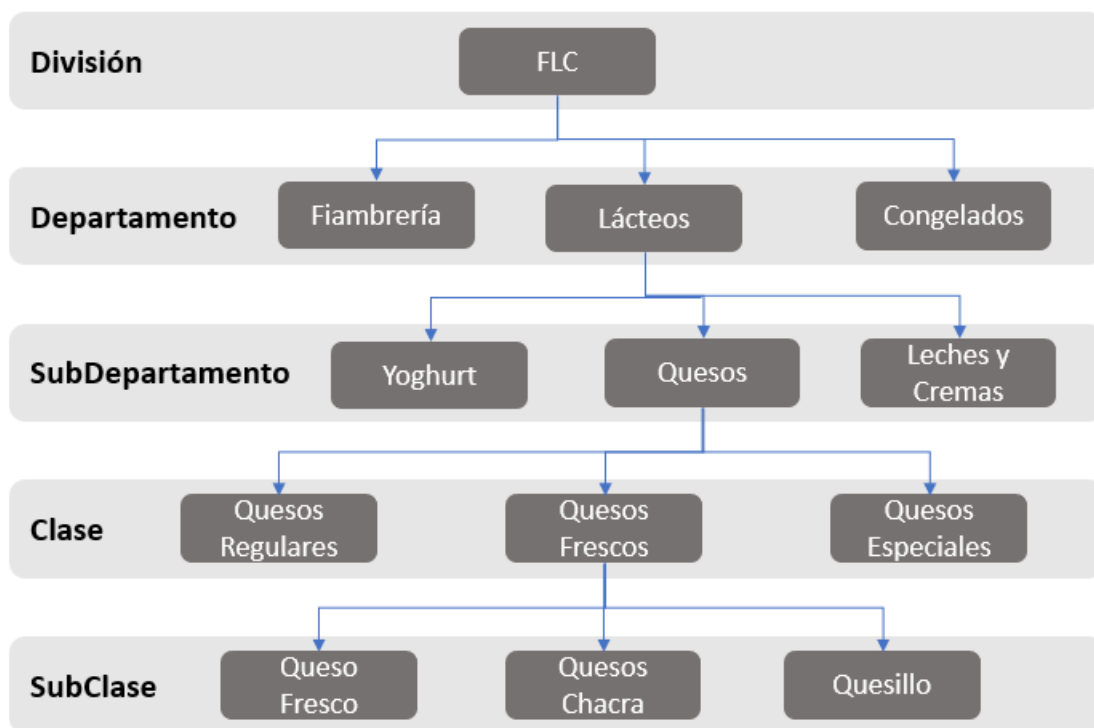


Figura 1. Jerarquía de Productos 1

La tabla 1 muestra el formato que se utilizó para hacer la primera captura de datos. Como se aprecia, cada fila es una boleta con las distintas subclases compradas en esa boleta y con la cantidad correspondiente. Esta tabla transaccional después se transformó en la tabla 2, donde cada fila sigue siendo la boleta, pero cada columna representa las distintas subclases que aparecen en la tabla 1. Cada intersección tiene un 1 o un 0, representando si en cada boleta se compró la subclase de cada columna o no.

Boleta	Producto	Cantidad
001	P1	1
002	P4	1
003	P1	1

*Tabla 1. Ejemplo Data Transaccional*

Boleta	P1	P2	P3	P4
001	1	0	0	0
002	0	1	0	1
003	1	1	1	0

*Tabla 2. Data Input para Matriz Co-Ocurrencia*

Esta matriz se generó trimestralmente, con la idea de evaluar las comunidades que se generan en los distintos periodos.

Por último, se hace el estudio solo con productos Food con el objetivo de encontrar asociaciones de productos que se compran usualmente, ya que los productos Non-Food son productos que se compran de manera intermitente y muy pocas veces tienen relación con otros productos.

## 4.2. Metodología

Una red se define como elementos interconectados unos con otros. La representación matemática de una red es un grafo. Estos grafos se componen de nodos, y la conexión entre estos que son los enlaces.

Para nuestro estudio, crearemos una red de productos, donde cada nodo es una subclase y la conexión entre cada nodo se da cuando estas dos subclases fueron compradas en una misma boleta. Por lo tanto, nuestra red es una representación de la data transaccional. Este

tipo de red se denomina red de productos co-comprados (*Co-purchased Product Network*).

Los datos y construcción de la red se trabajan en *Google Cloud Platform* (GCP), aprovechando sus ventajas en el procesamiento de grandes volúmenes de bases de datos con distintas herramientas de *big data*. Uno de los mayores desafíos fue trabajar con el volumen de data transaccional, tanto en la consulta y análisis en *BigQuery*, como el procesamiento y creación de la red en *AI-Platform* utilizando la librería de Python *NetworkX*.

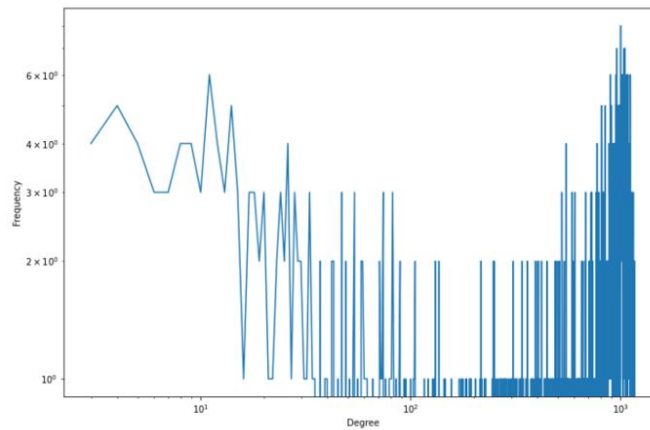
La construcción de esta red parte de la tabla 2, donde se crea una matriz de co-ocurrencia. Esta matriz es representada por la tabla 3, donde filas y columnas son cada una de las subclases que se han comprado en el periodo de tiempo analizado. Así, la intersección de cada elemento de la matriz se interpreta como las veces en que ambas subclases fueron compradas en una misma boleta. Es importante recordar que la diagonal de esta matriz debe ser cero para tener una apropiada interpretación. Por último, esta matriz se normaliza con respecto al total de boletas del periodo para así tener la relación relativa al total de boletas. Por ejemplo, si la intersección de P1 y P4 es 0.01, significa que solo el 1% de las boletas se compraron estas dos subclases juntas.

## 5. Limpieza y preparación de datos

	P1	P2	P3	P4
P1	0	1	1	0
P2	1	0	1	1
P3	1	1	0	0
P4	0	1	0	0

*Tabla 3. Matriz Co-Ocurrencia de Productos*

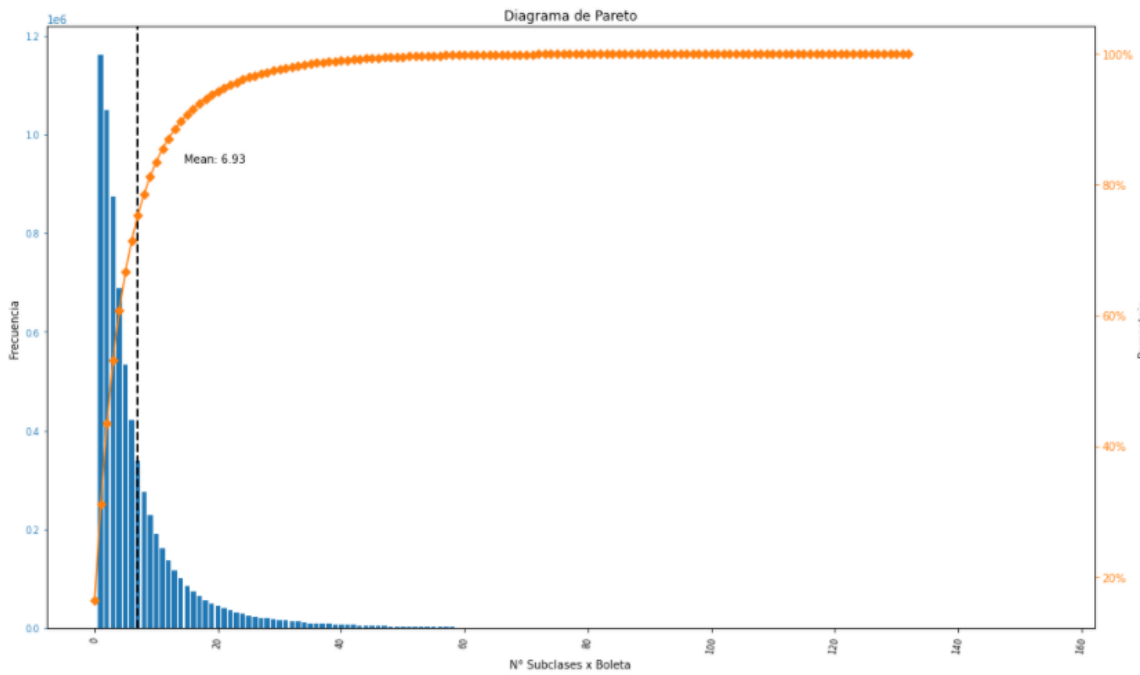
Una de las problemáticas al igual que (Raeder & Chawla, 2009; Videla-Cavieres & Ríos, 2013) es que se encontró una red demasiado densa, con nodos con un muy alto grado. En la Figura 2 se aprecia la densidad de esta red y la Figura 3 también muestra la distribución de grados de la red completa. Esto es importante ya que a nosotros nos importa encontrar relaciones representativas y descartar relaciones espurias. Es por esto que se generan distintas limpiezas para poder encontrarlas.



*Figura 2. Distribución de Grados de la Red*



La primera es más a nivel comercial. Durante cada trimestre de análisis tenemos alrededor de 4 millones de boletas, con distintas cantidades de subclases en cada una de ellas. En este caso, solo se toman las boletas que tengan más de una subclase, ya que buscamos más de una subclase por boleta para realizar asociaciones. Por otro lado, el límite superior debe cubrir el 90% de las boletas, que en el caso del primer trimestre son 16 boletas. Esto significa que el percentil 90 tiene 16 subclases por boleta. Para los trimestres siguientes también se utilizó este percentil para realizar el corte superior, pero este era de 24 y 26 cada uno. Esto se puede interpretar que los clientes compraban más subclases por boleta los trimestres siguientes, ya que tenían restricciones para ir a las tiendas a comprar, por lo que debían aprovechar el viaje a la tienda. Ahora bien, se genera un límite superior ya que hay muchas boletas que tienen desde 30 a 160 subclases, lo que indica que muchas de estas no existen patrones claros, por lo que, para tener una medida comparable en los tres trimestres, se realiza un corte con el percentil 90 en los tres. La figura 4 y tabla 4 muestran la distribución del número de subclases por boleta, siendo la mayor cantidad de boletas con una subclase y la menor cantidad con 160 subclases.



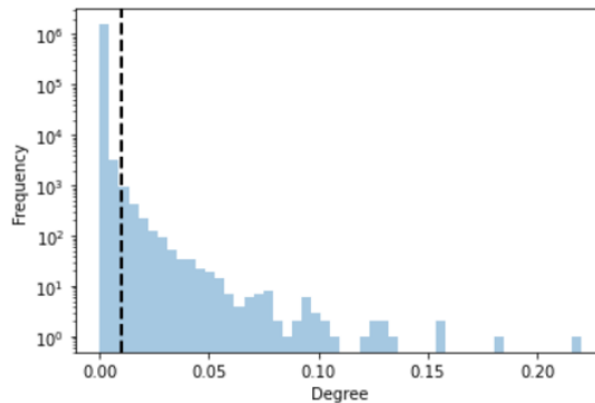
*Figura 3. Distribución de Cantidad de Subclases Distintas por Boleta. En promedio, el Primer Trimestre se Compraron 6,9 Subclases por Boleta.*

N° de Distintas Subclases x Boleta	N° Boletas	Frecuencia Relativa %	Frecuencia Relativa % Acumulada
1	1.160.035	16%	16%
2	1.048.434	14%	31%
3	874.801	12%	43%
4	688.855	9%	53%
....	....	....	....
15	74.822	1%	90%

*Tabla 4. Distribución de Cantidad de Subclases por Boleta. 15 Subclases por Boleta Acumula el 90% de las Boletas del Primer Trimestre.*

La segunda limpieza es encontrar un *threshold* de corte utilizando un grafico de distribución de pesos de la red. Como se ve en la figura 5 gran parte de la red tiene menos

de 0.01 de peso, por lo que se realiza un corte bajo ese valor para mantener solo las relaciones importantes. Es decir, todas las relaciones que tengan menos del 1% de las compras totales del periodo, se dejan como cero.



*Figura 4. Gráfico Distribución de Pesos. Se Realiza un Corte para los Tres Trimestres para Valores Menores a 0.01 para Hacerlos Comparables.*

Después de realizar esta limpieza podemos comenzar a realizar la búsqueda de comunidades. Las comunidades en el mundo de la teoría de grafos son grupos de nodos que están fuertemente conectados entre sí. Una de las medidas cuantitativas más usadas es la modularidad. Esta, mide la fuerza de la división de la red en comunidades. Si nuestra red tiene alta modularidad, quiere decir que existen conexiones densas entre los nodos dentro de la comunidad, pero conexiones escasas entre nodos en diferentes comunidades. Nuestro objetivo es tener una modularidad lo más cercana a 1 (va entre -0.5 y 1), para así mostrar la independencia de cada comunidad de la red. Como referencia (Videla-Cavieres & Ríos, 2014) obtiene un modularidad máxima de 0.617 y una mínima de 0.394.

Por último, utilizaremos el algoritmo de *Louvain* (Blondel; Guillaume; Lambiotte; Lefebvre, 2008) para la detección de comunidades. Este algoritmo aplica un método para optimizar el modularidad, comparando la densidad de enlaces dentro y fuera de los datos

de la red. Uno de los parámetros a ajustar es la resolución, la que por defecto es 1, y a medida que es más alta produce un mayor número de comunidades. Como nosotros queremos la mayor cantidad de comunidades, pero al mismo tiempo tengan el mayor valor de modularidad posible, graficamos esta relación para encontrar el valor de resolución para aplicar al algoritmo de *Louvain*. La figura 6 muestra lo descrito.

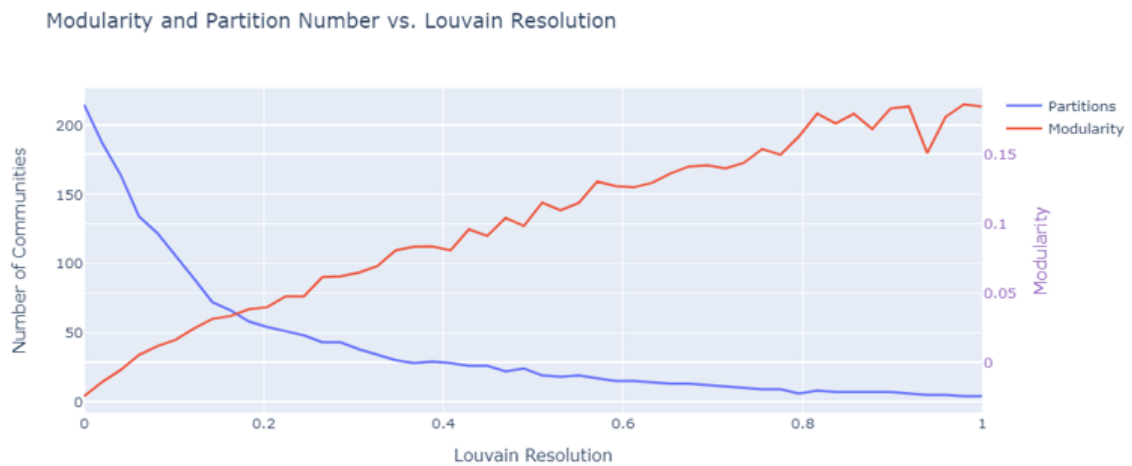


Figura 5. Gráfico que Compara Distintos Valores de Resolución y Evaluar Cuales Generan una Mayor Modularidad.

Por lo tanto, se utiliza un valor de modularidad de 1 en cada trimestre. En la siguiente sección se comentarán los resultados de las comunidades generadas.

## 6. Resultados

Como se comentó anteriormente, el análisis de comunidades se realizó para los primeros tres trimestres del 2020.

En forma de resumen, la tabla 5 muestra el número de comunidades por trimestre, la modularidad obtenida, el número de subclases, enlaces y grado promedio. Es llamativo el segundo trimestre, donde a pesar de mantener una alta modularidad, la cantidad de

subclases que pasan el filtro de 0.01 es baja. El repunte del tercer trimestre se puede deber en su mayoría por septiembre y el aumento del consumo por el retiro de dinero de las AFP<sup>2</sup>.

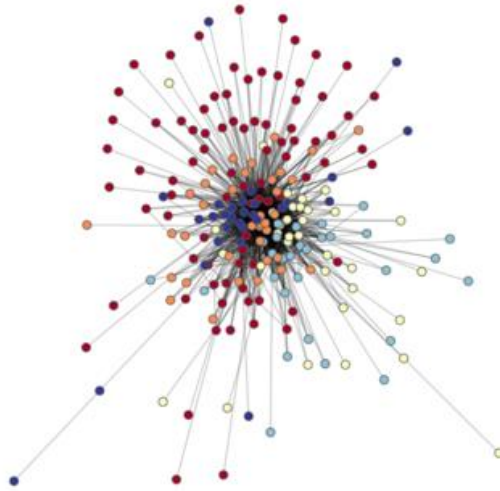
	N° Comunidades	Modularidad	N° de Subclases (Nodos)	N° de Enlaces	Grado Promedio
T1	5	0.25	215	1687	15,6
T2	3	0.23	57	202	7.07
T3	5	0.14	289	3497	24,2

*Tabla 5. Resumen de Grafos Trimestrales*

Como se mencionó en la sección de datos, se realizaron las pruebas a nivel de clase, subclase y subclase-marca, sin embargo, se optó a nivel de subclase ya que la modularidad que se encontraba en los otros dos niveles no llegaba a más de 0.1, lo que era bastante baja. En el caso de subclase alcanzamos una modularidad de 0.25, lo que es cercano al *benchmark* que tienen las otras investigaciones.

A modo de ejemplo, la figura 7 tiene el grafo que se obtiene, coloreando los nodos por comunidad, del primer trimestre.

<sup>2</sup> <https://www.diariooficial.interior.gob.cl/publicaciones/2020/07/30/42718/01/1791258.pdf>



*Figura 6. Grafo Trimestre 1*

A nivel visual no se puede apreciar muchas diferencias entre los nodos. Sin embargo, si graficamos solo una comunidad, como en la figura 8, podemos apreciar la relación entre los productos.



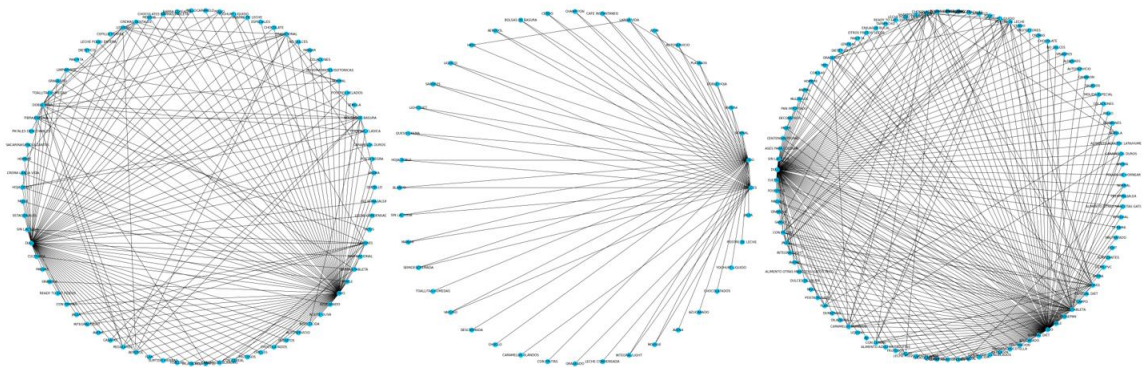
del doble de subclases que las demás comunidades, es menos probable que todos sus nodos estén conectados entre sí.

<b>Comunidad</b>	<b>N° Subclases</b>	<b>Grado</b>	<b>Centralidad Promedio</b>	<b>Descripción</b>
0	84	791	0.044	Dulces
1	37	812	0.102	Once
2	35	626	0.083	Fiesta
3	29	455	0.073	Ensaladas
4	30	690	0.107	Otros

*Tabla 6. Descripción Comunidades Trimestre 1*

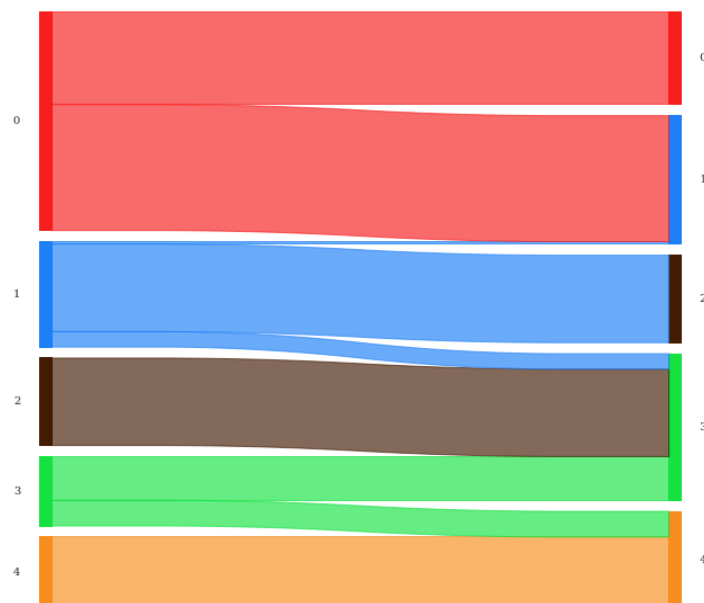
Un punto llamativo, es analizar la consistencia de estas comunidades a través del tiempo, sobre todo entendiendo el efecto de la pandemia en el patrón de consumo. Una de las comunidades mas llamativas es la que nosotros nombramos *Dulces*. Esta, que tiene como subclases principales los dulces y los batidos, se mantiene en el tiempo. Como se ve en la figura 9, el primer trimestre esta comunidad tiene mas de 20 subclases, donde los nodos que tienen la mayor cantidad de enlaces son las subclases nombradas anteriormente. Sin embargo, esta comunidad se reduce a más de la mitad el segundo trimestre, entendiendo como efecto de la pandemia, donde se podría intuir que gran parte de las familias optaron por productos de primera necesidad. El tercer trimestre esta comunidad recobra su forma del primer trimestre, entendiendo que ese trimestre el país vivió una recuperación económica, sumado al retiro de las AFP, lo que les dio a las familias mayor liquidez.





*Figura 8. Grafo Comunidad de Dulces en los Tres Trimestres*

Sin embargo, a pesar de que esta comunidad recobra la cantidad de nodos, estos no son los mismos que el primer trimestre. Con un diagrama de Sankey, podemos ver en la figura 10 como el primer y tercer trimestre, la comunidad 0 no mantiene sus subclases. Esto se debe en su mayoría a que hay muchos enlaces dentro de una comunidad que de un trimestre no tienen la misma “fuerza”. Pero también, hay nodos que son mas estables y estos son los que se mantienen en cada trimestre, como dulces y batidos.



*Figura 9. Diagrama de Sankey con Movimiento de Subclases en Comunidades entre Trimestre 1 y Trimestre 3*

## 7. Conclusiones

Este trabajo estudia la aplicación de técnicas de análisis de redes y algoritmo de detección de comunidades para el problema *Market Basket Analysis* en una empresa de retail para tener una alternativa de sistema de recomendación para esta empresa.

Esto, se realiza en tres trimestres distintos con productos Food de la empresa, tratando de buscar asociaciones entre subclases.

Nuestro objetivo general de obtener un nuevo sistema de recomendación para la empresa de retail se cumple, ya que, a pesar de obtener una baja modularidad, se observan asociaciones fuertes entre cada comunidad de cada trimestre. Este recomendador puede tener múltiples usos como enviar productos complementarios a los que siempre compra el cliente, estudiar modelos de pricing para ofertas de productos empaquetados o modelos para supply-chain para posicionar productos complementarios en góndola.

Nuestros objetivos específicos también se logran. Se prueba y desarrolla un método de análisis para un *Market Basket Analysis* en la empresa de retail, encontrando una metodología practica para encontrar asociaciones relevantes con la matriz de co-ocurrencia, filtrando los enlaces menos relevante y construyendo las distintas comunidades.

También, se observa que la red no sigue un patrón modular, donde no hay una separación clara entre subclases. Esto se puede explicar entendiendo que cada cliente sigue un patrón de compra distinto, a pesar de que existan asociaciones comunes como los dulces, en

general a la red completa tiene muchas combinaciones distintas sin un patrón fuerte. Por lo que el segundo objetivo se cumple, ya que se crean comunidades, pero dada las características de los datos, las comunidades no son representativas.

Finalmente, se analizan las redes temporales de los tres trimestres, encontrando que hay comunidades que mantienen cierta forma, alterando algunos nodos. Sin embargo, resulta interesante destacar el efecto de la pandemia en la construcción de estas comunidades, dejando a la vista que hay subclases que se llevan muchas transacciones asociadas (grados del nodo) pero que, al comenzar la pandemia, pierden asociaciones porque no son productos de primera necesidad.

Por último, como se mencionó en la revisión de literatura, teníamos un benchmark de modularidad del estudio de (Videla-Cavieres & Ríos, 2014), donde la menor modularidad era de 0.3. Acá se alcanza un modularidad de 0.25, la que se acerca a la del estudio, pero sigue siendo baja. Esta podría mejorar incluyendo otras técnicas de limpieza de enlaces y que se analizara un periodo de tiempo más largo, ya que se estudiaron solo tres trimestres, incluyendo una pandemia que cambio mucho los patrones de consumo. También sería interesante probar otros algoritmos de detección de comunidades, para hacer de estas mas consistentes.

## Bibliografía

1. Kim, H. K., Kim, J. K., & Chen, Q. Y. (2012). A product network analysis for extending the market basket analysis. *Expert Systems with Applications*, 39(8), 7403–7410.
2. Raeder, T., & Chawla, N. V. (2009). Modeling a store's product space as a social network. In *2009 International conference on advances in social network analysis and mining* (pp. 164–169). IEEE.
3. Ivan F. Videla-Cavieres, Sebastián A. Ríos. (2014). Extending market basket analysis with graph mining techniques: A real case. *Expert Systems with Applications*, 41(4), 1928-1936.
4. Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceedings of 20th international conference on very large data bases, VLDB* (pp. 1– 32).
5. Agrawal, R., Imielinski, T., & Swami, A. (1993). Database mining: A performance perspective. *IEEE Transactions on Knowledge and Data Engineering*, 5(6), 914–925.

6. G. Russell and A. Petersen. Analysis of cross category dependence in market basket selection. *Journal of Retailing*, 76(3):367–392, 2000.
7. S. Krinsky, J. Ennis & R. Weissman. Academic-Corporate Ties in Biotechnology: A Quantitative Study. *Science, Technology, & Human Values*. 16(3):275-287, 1991.
8. B. J. Reed & D. R. Segal. Social Network Analysis and Counterinsurgency Operations: The Capture of Saddam Hussein. *Sociological Focus*. 19(4):259-264, 2006.
9. Davis, J., Hossain, L., & Murshed, S. H. Social network analysis and organizational disintegration: The case of Enron corporation. In *Proceeding of international conference on information systems*. Montreal, Quebec, Canada, 2007
10. Vincent D Blondell, Jean-Loup Guillaume, Renaud Lambiotte & Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory and Experiment*, 2008.

11. Bhatti, Akram, Basit, Khan, Naqvi, Bilal. E-commerce trends during COVID-19 Pandemic. International Journal of Future Generation Communication and Networking, 2020.
12. <https://www.ecommerceccs.cl/comercio-electronico-sigue-en-trayectoria-de-crecimiento-exponencial/>
13. Gupta & Mamtora. A Survey on Association Rule Mining in
14. Market Basket Analysis. International Journal of Information and Computation Technology, 2014.