

## **Análisis descriptivo de los despachos realizados en Europa por una empresa de dispositivos médicos**

### Introducción

Contamos con un set de datos que tiene 760.377 observaciones y 269 variables cuantitativas y cualitativas. Estos registros representan 1 mes de trabajo de los centros de distribución ubicados en Europa de la compañía. Las variables que uso de todas las disponibles son las siguientes:

- Id: número de orden
- Tipo de orden
- Detalle de la orden
- Origen del almacén
- Destino de la orden
- Franquicia
- Subfranquicia
- Cumplimiento ON TIME
- Cumplimiento IN FULL
- Cumplimiento OT-IF
- Fecha de envío
- Fecha estimada de envío
- Cantidad de productos enviada
- Cantidad de productos pedidos
- Hora de confirmación del envío

OT-IF es una medida de la logística o del rendimiento de entrega dentro de una cadena de suministros, mide la frecuencia con la que el cliente obtiene lo que quiere en el momento que lo desea. Se calcula con las variables cantidad de producto (cuando es positivo tengo un HIT en el cumplimiento IN FULL) y la variable fecha de envío (cuando es positivo tengo un HIT en el cumplimiento ON TIME).

Este indicador clave de rendimiento es de gran valor para el negocio, ya que aumenta las ganancias operativas debido a la reducción de gastos operativos (mejor control del inventario, mejor toma de pedidos de los clientes, mayor confiabilidad en el almacenamiento y transporte, entre otras) y aumenta las ventas (debido a una mejor disponibilidad de productos para ventas)

Esto último es fundamental, ya que es importante tener los productos disponibles en los centros de distribución cuando el cliente los pida, es por esto, que decidí armar una red de los despachos realizados durante 1 mes y relacionarlo con el cumplimiento de OTIF por el tipo de orden, y su posible relación a la lejanía de los destinos, además de pensar un modelo predictivo de cumplimiento de OTIF basado en la distribución de los despachos.

## Metodología

Para la preparación de los datos se usó R studio, que me permitió transformar los datos de una forma cruda a una forma en la que pueden ser utilizados directamente en los modelos. Una vez completada la transformación se avanzó hacia la visualización de los datos y el modelado.

En cuanto a la visualización del grafo se usó Gephi y la librería de R studio igraph.

## Resultados y análisis de los resultados

Las variables que usé para realizar el grafo son el origen del almacén y el destino de la orden (los nodos) y el número de orden (arista), donde el peso es referente a las cantidades de despachos que se realizan desde el origen al mismo destino.

Por ende, es un grafo con aristas dirigidas porque hay una dirección en las interacciones y bipartito ya que ni los orígenes ni los destinos están conectados entre sí (se relacionan dos conjuntos de elementos diferentes). Además, es desconexo ya que, aun sabiendo que cada par de nodos está conectado por un camino, como podemos ver mas adelante hay una comunidad que no está conectada con el resto del grafo.

En un grafo la distancia entre dos nodos es el menor número de aristas de un recorrido entre ellos. El diámetro es la mayor distancia entre dos puntos de la misma, y en este caso el diámetro es 1

Número de nodos (N) = 183

Número de enlaces (L) = 390

Es un grafo dinámico, ya que, aunque no se crean nuevos nodos, se pudieran crear nuevos enlaces o rutas con el paso del tiempo.

Grado (k), en las redes dirigidas se toman dos parámetros para diferenciar el grado, uno es el grado entrante (k.en) el cual representa el número de enlaces que apuntan hacia el nodo y el grado saliente (k.in) que es el numero de enlaces que salen del nodo, entonces el grado total de un nodo (k) es:  $k.en + k.in$ . Tendremos nodos con grado k.en (los del destino) y nodos con grado k.in (los del origen).

Grado promedio es  $L/N = 390/183 = 2,13$

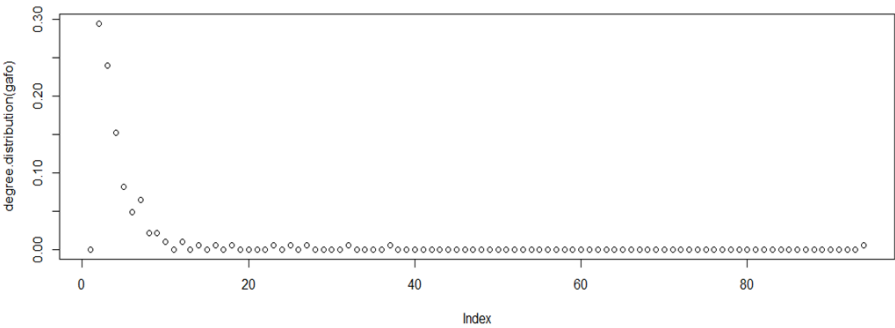
Distribución de grado ( $p_k$ ) da la probabilidad de que un nodo elegido al azar de la red tenga grado k. Siendo  $p_k$  una probabilidad, entonces para una red con grado k, la distribución de grado será  $p_k = N_k / N$ . donde  $N_k$  es la cantidad de nodos con grado k. En la siguiente tabla podremos observar la cantidad de nodos con los grados correspondientes y la probabilidad de encontrarnos con ellos en los datos.

**Tabla1. Cantidad de los nodos presente en los distintos grados con su distribución.**

Cantidad de Nodos	Grado	pk %
54	1	29,51
44	2	24,04
28	3	15,30
15	4	8,20
9	5	4,92
12	6	6,56
4	7	2,19
4	8	2,19
2	9	1,09
2	11	1,09
1	13	0,55
1	15	0,55
1	17	0,55
1	22	0,55
1	24	0,55
1	26	0,55
1	31	0,55
1	36	0,55
1	93	0,55

En lo datos, es más probable encontrarnos con nodos de grado 1 o 2 que más altos y es menos probable encontrarnos con nodos de grado mayor. Además, esos únicos nodos que tienen grado superior al promedio, tienen mayor número de HUB, que en la teoría se refiere a un nodo que tiene un numero de enlaces que excede ampliamente al promedio. En la siguiente grafica podemos observar la distribución de los grados, la cual se ajusta a una exponencial invertida.

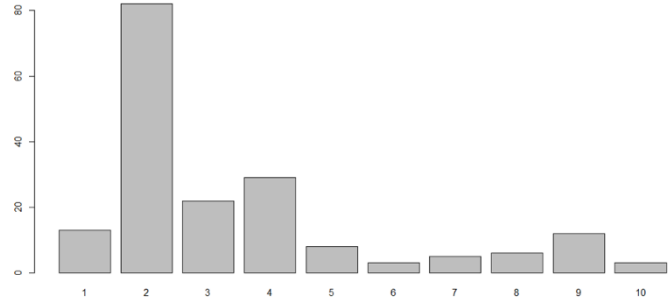
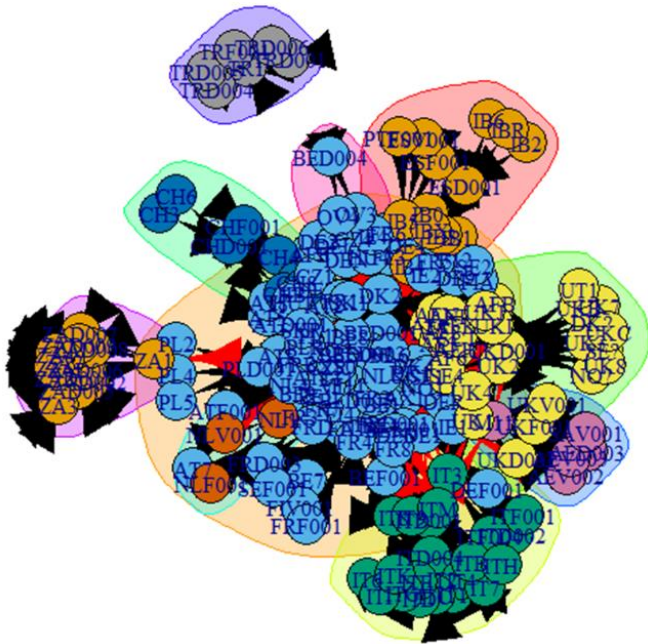
**Grafico 1. Distribución de los grados**



Siguiendo con el análisis se midió la modularidad (m) del grafo con dos algoritmos distintos. Primero se usó cluster walktrap en R studio, el cual trata de encontrar las comunidades densamente conectada a través de caminatas aleatorias, la idea es que los paseos cortos tienden

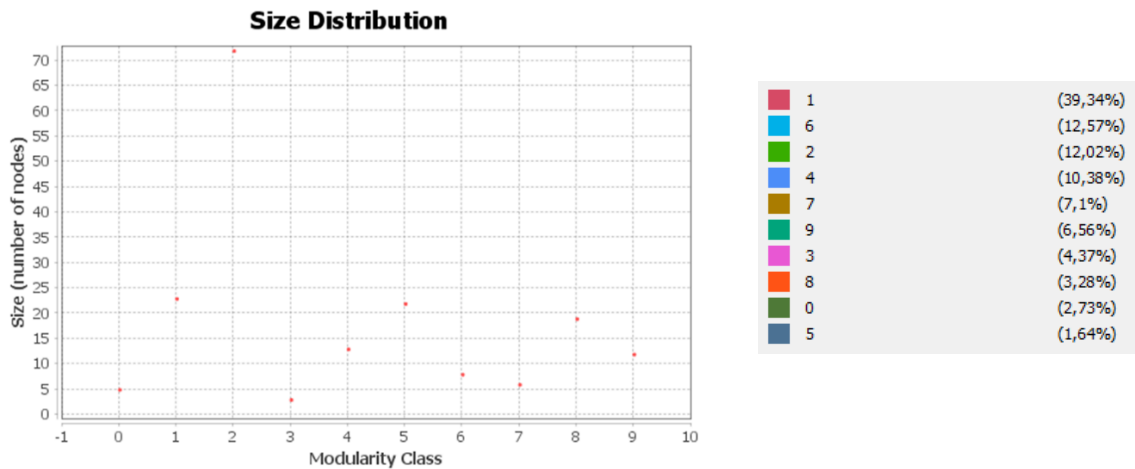
a permanecer en la misma comunidad, se obtuvo  $m = 0,45$ . Encontrando 10 comunidades con diversas cantidades de nodos como podemos observar en las siguientes graficas.

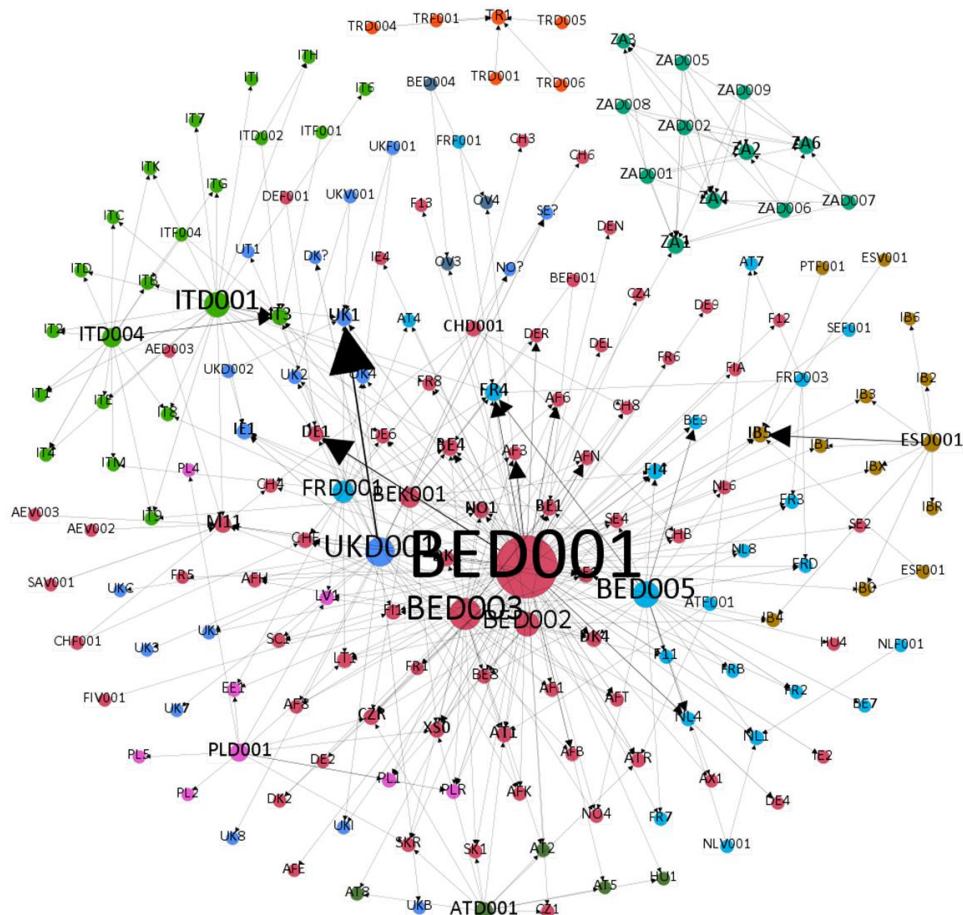
**Graficas 2 y 3. Comunidades y cantidad de nodos por comunidad usando R studio.**



El otro algoritmo empleado para calcular la modularidad es el de Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000 en Gephi  $m = 0,62$ , el algoritmo es un método heurístico que se basa en la optimización de la modularidad y lo que nos da un poco más alto que la modularidad anterior pero encontrando también 10 comunidades distintas

**Graficas 4, 5 y 6. Comunidades y cantidad de nodos por comunidad usando Gephi.**





La modularidad es la fuerza de la división de una red en módulos, su valor reside en la gama de  $[-1,1]$ , ambos algoritmos nos dan una alta modularidad lo que significa tener conexiones densas entre los nodos de la comunidad, pero conexiones dispersas entre los nodos en diferentes comunidades.

En grafos dirigidos, cada camino debe seguir la dirección de las flechas. Y para este caso donde la red tiene caminos dirigidos y es bipartita, tiene la particularidad que de los dos subconjuntos que tiene la red, uno solo realiza despachos (ósea que nos solo salen los links) y el otro subconjunto solo recibe (solo recibe los links dirigidos), esto explica los resultados de las siguientes métricas calculadas con R studio:

- Diámetro de la red es la distancia mas grande entre todos los pares de nodos de la red

`diameter(grafo, directed = TRUE, unconnected = TRUE, weights = g$V3) = 101915` (medido en despachos realizados)

Pero para no pesado es 1

- Coeficiente de clustering

Nivel de agrupamiento que existe entorno a un nodo

`transitivity(grafo,type="global") = 0`

Esta métrica nos explica que los vecinos no están conectados

- Camino mínimo entre los nodos

`average.path.length(grafo, directed=TRUE) -> para no pesado = 1`

`coreness(grafo, mode = "out") = 0`

`coreness(grafo, mode = "in") = 0`

`coreness(grafo, mode = "all")`

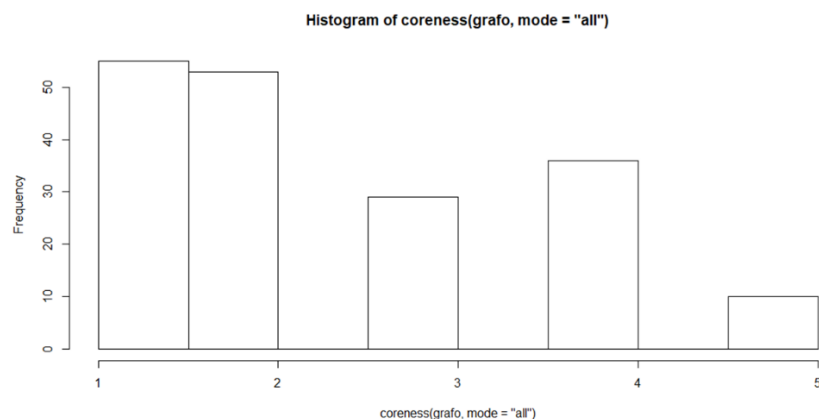
`plot(coreness(grafo, mode = "all"))`

all: se considera el gráfico no dirigido correspondiente

Usando Gephi calculé la densidad del grafo = 0,012, lo que nos dice que es un grafo disperso ya que es mas cercano de 0 la densidad, lo que nos dice que el grafo tiene muy pocas aristas. Y el grado medio con pesos 3554,53.

En la figura 7 podemos observar como veíamos en los grados de la tabla 1, que la mayor cantidad de nodos tienen entre grado 1 y 2. Y para finalizar en la figura 8 se observa el grafo sobre el mapa para representar con las coordenadas de los orígenes y los destinos, como se ven los despachos realizados desde Europa.

**Figura 7. Histograma de coreness considerando el grafo no dirigido.**



**Figura 8. Plot del grafo con las coordenadas de los envíos realizados desde Europa.**

