

Práctica Análisis Redes Sociales

ANALISIS DE COMPONENTES EN COMESTICOS

Realizado por Araceli Macía Barrado



INDICE

ANALISIS COMPONENTES COSMETICOS	3
1. INTRODUCCIÓN	3
2. OBTENCION DE DATOS	4
3. CREAR GRAFO	7
4. ANALISIS DEL GRAFO	9
COMPONENTE GIGANTE	11
DEGREE	13
BETWEENNESS	16
CLOSENESS	18
ECCENTRICITY	19
DENSITY	20
MODULARITY	21
5. CONCLUSIONES	23

ANALISIS COMPONENTES COSMETICOS

Realizado por Araceli Macía Barrado

1. INTRODUCCIÓN

El objetivo del trabajo es realizar un análisis sobre la relación entre los componentes de diferentes productos cosméticos, como son: Cremas Antiarrugas, desodorantes, pastas dentífricas y champús.

Tras una primera impresión, se podría pensar que son productos que nada tienen que ver unos con otros, pero tras el análisis veremos que tienen componentes comunes entre ellos que establecen la relación entre los distintos productos.

Lamentablemente, no dispongo de conocimientos químicos que ayuden a entender los componentes de modo que fuera más fácil encontrar relación y sentido a los datos que vamos a analizar, pero, aun así, se pueden establecer conclusiones a raíz del estudio de los datos.

Además, tras la búsqueda de varios componentes en Internet, he llegado a varias páginas donde indican del peligro de ciertos componentes por clasificarlos como tóxicos:

- <http://www.ajedrea.com/blog/89-que-es-lo-que-llevan-nuestros-cosmeticos-100-ingredientes-de-cosmeticos-que-debes-evitar>
- <http://ecomaniablog.blogspot.com.es/2013/09/los-cosmeticos-mas-toxicos.html>.

De modo que he añadido esta información o clasificación a los componentes, para ver en qué tipo de productos nos los podemos encontrar.

2. OBTENCION DE DATOS

Como he comentado en el apartado anterior, la practica trata de analizar la relación entre los componentes de diferentes productos: Cremas, desodorantes, pasta dentífrica y champús.

Para ello, he accedido a una página para obtener los componentes de cada uno de los productos. La página donde he encontrado la composición de los componentes es:

<http://www.promofarma.com>

Tras la visualización de la página, he visto que los links a los detalles de cada una de las cremas se generaban dinámicamente, por lo que no ha sido posible hacer la recolección de links mediante un script, así que manualmente he copiado los enlaces de los productos a diferentes ficheros de texto. De modo que he generado 4 ficheros de texto, con las urls a los detalles de cada uno de los productos.

Ficheros: urlProductos.txt, UrlChampus.txt, UrlDentrificos.txt y UrlDesodorantes.txt.

Ejemplo de contenido del fichero:

```
http://www.promofarma.com/boreade-sl-emulsion-alisante-40ml
http://www.promofarma.com/btses-crema-hidratante-antiarrugas-50ml
http://www.promofarma.com/btses-inhibidor-de-arrugas-gel-15ml
http://www.promofarma.com/c-vit-locion-facial-revitalizante-30ml
http://www.promofarma.com/c-vit-radiance-fluido-luminoso-50ml-plus-c-vit-contorno-de-ojos-15ml-en-farmacia-ballesteros
http://www.promofarma.com/can-boada-crema-hidra-forte-uv-acido-hialuronico-50ml-en-farmacia-can-boada
```

El notebook entregado PT_1_ObtenerDATOS contiene el código que he utilizado para acceder a los componentes, donde se puede ver el detalle de todo el código. A grandes rasgos esto es lo que se hace:

- Leer el fichero de texto de las urls.

```
datos = pd.read_csv('UrlDesodorantes.txt', sep=" ", header=None, names=["url"])
```

- Desordenar los datos (ya que estaban ordenados alfabéticamente por nombre producto).

```
datosLeer1=datos.url
print datosLeer1[1:5]
np.random.shuffle(datosLeer1)
```

- Accedo a cada página para recoger el precio y los componentes:

```

xpath_expr = '//div[@class="bottom-spacer-small padded-small gray-box tab_composicion"]'
xpath_expr2 = '//span[@id="price"]'
composicion= list()
productos= list()
precios=list()
for uL in datosLeer1: #Se accede a cada URL pasada por parametro.
    response = requests.get(uL)
    content = response.text.encode('utf-8') #Codifico el texto a utf-8
    time.sleep(2)
    tree = html.fromstring(content)
    results = tree.xpath(xpath_expr)
    resultsPrice = tree.xpath(xpath_expr2)
    if (len(results)>0):
        #for elem in results:
        #    leido=elem.text_content()
        datos=results[0].text_content()
        datosPrecio=resultsPrice[0].text_content()
        composicion.append(datos)
        productos.append(uL)
        precios.append(datosPrecio)

```

con este código estoy guardando esta parte del html:

Composición

AQUA. PROPYLENE GLYCOL. OCTOCRYLENE. C12-15 ALKYL BENZOATE. VITIS VINIFERA EXTRACT. ALOE BARBADENSIS. JOJOBA ESTERS. TOCOPHERYL ACETATE. CARBOMER (CARBOXYPOLIMETILEN). TRIETHANOLAMINE. DMDM HYDANTOIN. METHYLPARABEN. PARFUM. PROPYLPARABEN. LIMONENE.

He tenido en cuenta que no todos los productos tienen detallada la composición.

d. Tratamiento de los datos.

Una vez que he recogido las composiciones de los productos, tengo estos datos:

```

In [15]: datosTratar=composicion#[1:3]
          datosTratar[1:5]

Out[15]: ['\n\n          Secreci\xfa de caracol, \xealcido fer\xfalico, retinol, silicio org\xneinico, glicosaminoglican\nos, resveratrol, prote\xededna hidrolizada de trigo, pantenol, pentella asi\xeltica, DMAE.\n\n          AQUA. BUTYLENE GLYCOL. BUTYLPARABEN. CARBOMER (CARBOXYPOLIMETILEN). CI 42051. COLLAGEN. ETHYLPA\nRABEN. GLYCERIN. GLYCOPROTEINS. ISOHEXADECANE. METHYLPARABEN. PARFUM. PEG-8. PENTAERYTHRITYL TETRACAPRYLATE/CAPRATE.\nPHENOKYETHANOL. POLYMETHYL METHACRYLATE. POLYSORBATE 20. PROPYLPARABEN. SORBIC ACID. TOCOPHERYL ACETATE. TRIETHANOLA\nMINE. XANTHAN GUM. ACRYLATES/C10-30 ALKYL ACRYLATE CROSS...\n\n          AQUA. ETHYLHEXYL METHOXYCINNAMATE. BUTYLENE GLYCOL. DIETHYLHEXYL CARBONATE. OCTOCRYLENE. GLYCE\nRETH-26. CYCLOPENTASILOXANE. ISONONYL ISONONANOATE. SETEARETH-2. BUTYL METHOXDIBENZOYLMETHANE. SILICA. STEARETH-21. D\nIMETHICONE. GLYCERIN. POLYMETHYL METHACRYLATE. PHENOKYETHANOL. AMMONIUM ACYLOYLDIMETHYLAURATE/VP.. CAPRYLYL GLYCOL. G\nLYCERYL STEARATE. CAPRYLIC/CAPRIC TRIGLICERIDE. PARFUM. XANTHAN GUM. EDTA DISODIUM. ETHYLHEXYLGLYCERIN. TOCOPHERYL AC\nETATE. \xe0DIPROPYLENE GLYCOL. TOCOPHEROL. BHT. DIPEPTIDE DIAMINOSUTYROYL BENEYLAMIDE... ASCORBYL PALMITATE. ASCORBIC\nACID. CITRIC ACID. HEXAPEPTIDE-9.\n\n          AQUA, GLYCERIN, ETHYLHEXYL METHOXYCINNAMATE, HYDROGENATED COCO-GLYCERIDES, BUTYLENE GLYCOL, BUT\nYROSPERMUM PARKII, BUTYL METHOXYDIBENZOYLMETHANE, DIMETHICONE, GLYCERYL STEARATE CITRATE, HOMOSALATE, STEARYL ALCOHO\nL, CETYL ALCOHOL, GLYCINE SOJA, SODIUM HYALURONATE, TITANIUM DIOXIDE, SODIUM CARBOMER, PHENOKYETHANOL, MICA, METHYLPA\nRABEN, CAPRYLYL GLYCOL, TRISODIUM EDTA, SILICA, ACRYLATES/C10-30 ALKYL ACRYLATE CROSSPOLYMER, PROPYLPARABEN, BHT, PAR\nFUM.\n']

```

Por cada fila, convierto los datos a minúsculas, y separo los componentes por “.”, “,” e “y” por “.”

Y las componentes que están formados por más de una palabra lo convierto a una sola palabra.

Además, tras la revisión de las palabras de los componentes, he unificado palabras, que correspondían al mismo componente.

Ejemplo:

```
if ("perfum" in ingParsear) | ("parfum" in ingParsear):
    ingParsear="perfum"
elif ("mosqueta" in ingParsear):
    ingParsear="rosamosqueta"
elif ("agua" in ingParsear) | ("agua" in ingParsear) | ("water" in ingParsear):
    ingParsear="agua"
elif ("camel" in ingParsear):
    ingParsear="camelia"
elif ("caprylic" in ingParsear):
    ingParsear="caprylic"
elif ("carbomer" in ingParsear):
    ingParsear="carbomer"
elif ("centella" in ingParsear):
    ingParsear="centella"
elif ("ceraalba" in ingParsear):
    ingParsear="ceraalba"
elif ("cetylalcoh" in ingParsear):
    ingParsear="cetylalcohol"
```

En el ejemplo, palabras que empiecen por caprylic las unifico al componente caprylic.

Los precios, también los tengo que tratar para limpiarlos de caracteres extraños.

```
#tratando los precios.
preciosFinal=list()
num=""
for i in precios:
    #print "..", i
    num=""
    for j in i:
        if (j.isdigit() or (j==".")):
            num=num+j
        #print num
    if (num!=""):
        preciosFinal.append(float(num))
    else:
        preciosFinal.append(float("0"))
```

e. Creación de Data Frame y volcado a fichero.

Una vez que ya tengo los datos, genero un dataframe y lo guardo en disco como un csv.

De forma que tengo una línea por producto, y de cada producto tengo los componentes, el precio y el número de componentes por producto.

```
d = {'producto': productos, 'composicion': composicionFinal, 'precio': preciosFinal, 'numIngredientes':totalIngre }
df = pd.DataFrame(data=d)
```

ejemplo:

	composicion	numIngredientes	precio	producto
0	[aloevera, centella, extractodesucaliptoorgnic...	14	9.50	http://www.promofarma.com/d-shila-desodorante-...
1	[agua, aluminumzirconiumoctachlorohydrexgly, a...	13	10.30	http://www.promofarma.com/germisdin-rx-hh-deso...

Los ficheros csv que he generado son: datosComponentes.csv, datosComponentesChampus.csv, datosComponentesUrIdentrificos.csv y datosComponentesUrIdesodorantes.csv

3. CREAR GRAFO

Para la generación del grafo, he creado el notebook llamado PT_2_CrearGrafo. En dicho notebook, se puede ver más en detalle el código que he creado para generar el grafo con los datos.

En este notebook, estoy utilizando también este fichero de texto:

ComponentesToxicos.txt

Que contiene una lista de elementos tóxicos según varias páginas web (mencionadas en la introducción).

A grandes rasgos:

Pasos para la generación del grafo:

a) Leo el fichero de datos.

```
#lectura de los datos.
datosCremas = pd.read_csv('datosComponentes.csv', sep=";")
datosChampus=pd.read_csv('datosComponentesChampus.csv', sep=";")
datosDientes=pd.read_csv('datosComponentesUrlDentrificos.csv', sep=";")
datosDesod=pd.read_csv('datosComponentesUrlDesodorantes.csv', sep=";")
```

Una vez que he leído todos los datos, voy a seleccionar 45 productos de cada producto, ya que he comprobado que con más datos Gephi se bloquea.

b) Generar nodos

Hay que tener en cuenta que los nodos son cada uno de los componentes existentes.

Leo los componentes de cada fila, generando un vector con todos los componentes, para luego eliminar duplicados y generar los nodos.

He creado una función para generar una lista con todos los componentes de los productos.

```
def CrearLista(datos):
    totalcomponentes=list()
    for comp in datos:
        #print "..", type(comp), "-", comp
        listacomp=nltk.word_tokenize(comp)
        for cpl in listacomp:
            #print "---", cpl
            if (cpl.isalnum() and (cpl>3)):
                totalcomponentes.append(cpl)
    return totalcomponentes
```

Para la generación de los nodos, tengo en cuenta de que producto son, y luego si están en la lista de componentes tóxicos o no para añadir esa información al nodo generado.

```
#creo los nodos, si el nodo esta en la lista de elementos toxicos lo pongo la marca.
toxico=""
#for e in nodosCrear:
for e in set(totalcomp):
    producto=""
    if e in comCrema:
        producto="Crema"
    if e in compChampu:
        producto=producto+"Champu"
    if e in compDientes:
        producto=producto+"Pasta"
    if e in compDesod:
        producto=producto+"Desodorante"
    if e in toxicoElem:
        toxico="SI"
    else:
        toxico="NO"
    #print "-pr", producto
    G.add_node(e, {"toxico":toxico, "producto": producto})
```

c) Generar Aristas

Reviso los componentes de cada producto, de forma que todos los componentes que pertenecen a un mismo producto (crema, pasta, etc.), están relacionados y por lo tanto debe generarse una arista no dirigida entre ellos.

```
totalcomponentes=list()
compProducto=list() #variable que tiene todos los componentes de un producto.
out_list_edges = []

for comp in totalComponentes: #recorro cada uno de los productos.
    #print "..", type(comp), "-", comp
    listacomp=nlk.word_tokenize(comp)
    compProducto=list()
    for cpl in listacomp:
        #print "---", cpl
        if (cpl.isalnum() and (cpl>3)):
            compProducto.append(cpl)
    #cuando salgo del bucle tengo la lista de componentes limpia del producto.
    in_list = compProducto
    for i in range(2, 3): #calcula las combinaciones de cada elemento de la lista,
        #creando relaciones entre cada uno de los componentes de un mismo producto
        out_list_edges.extend(itertools.combinations(in_list, i))
```

Ejemplo de las aristas:

```
out_list_edges[1:5]

[('jaleareal', 'algasangrededecristo'),
 ('jaleareal', 'extractodelinnydefrutadelapasin'),
 ('jaleareal', 'siliconascicloneticas'),
 ('jaleareal', 'vitaminae')]
```

A la hora de generar la arista, miro a ver el número de veces que se produce esa combinación, para añadirle peso en la creación de la arista en el grafo.

```
for elem in set(out_list_edges):
    peso=out_list_edges.count(elem) #añado el peso al grafo, como numero de veces que se repite la arista
    #print elem[0], ",", elem[1]
    G.add_edge(elem[0],elem[1], weight=peso)
```

Por último, guardo el grafo en disco.

```
nx.write_graphml(G, "GrafoComponentesComesticos.graphml")
```


4. ANALISIS DEL GRAFO

Una vez que tengo el grafo generado, lo importo con Gephi para empezar a analizarlo.
















El grafo resultante tiene estas medidas:

Nodes: 905

Edges: 23473

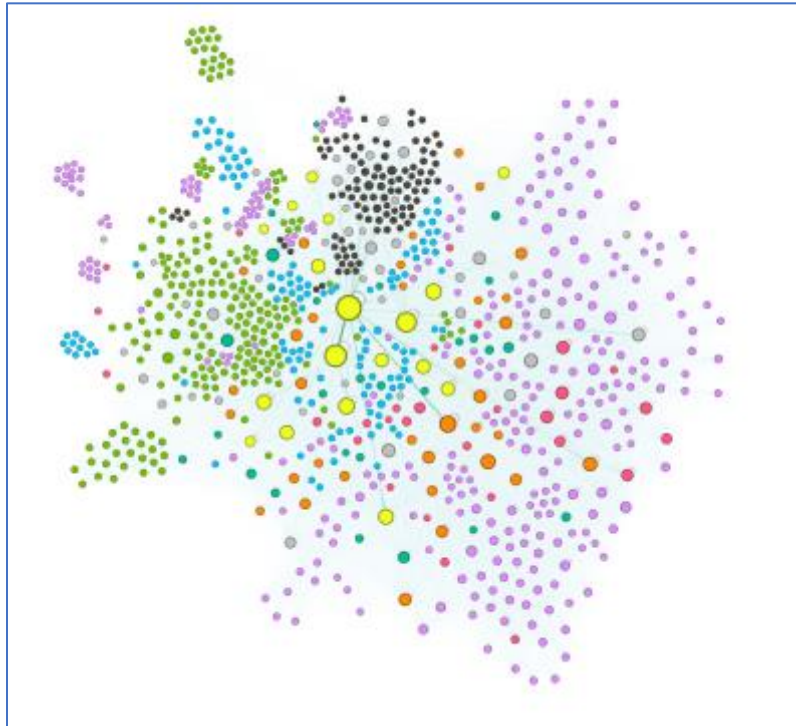
Undirected Graph

Para ver el nodo resultante, Lo primero que vamos a hacer es pintar los nodos de acuerdo al tipo):

	Crema	(34.36%)
	Champu	(24.2%)
	Desodorante	(13.48%)
	Pasta	(11.16%)
	CremaChampuDesodorante	(3.76%)
	CremaDesodorante	(3.2%)
	CremaChampu	(2.98%)
	CremaChampuPastaDesodo...	(1.99%)
	ChampuDesodorante	(1.55%)
	CremaChampuPasta	(0.99%)
	CremaPasta	(0.77%)
	ChampuPasta	(0.55%)
	PastaDesodorante	(0.44%)
	ChampuPastaDesodorante	(0.33%)
	CremaPastaDesodorante	(0.22%)

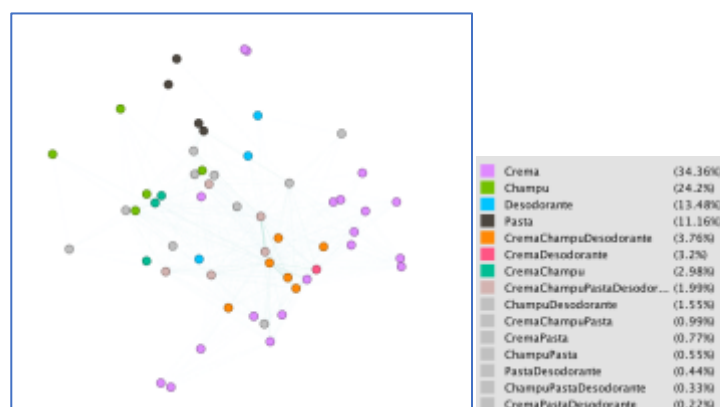
Con esta tabla podemos ver el % de nodos "puros", lo que solo se dan en Cremas, Champú, Desodorante y pasta Dentífrica. Y luego ya vemos los componentes que aparecen en diferentes productos, que son menos frecuentes.

También en las aristas, se pinta el grosor de acuerdo al peso que tienen (número de veces que aparece dicha combinación en los diferentes productos).



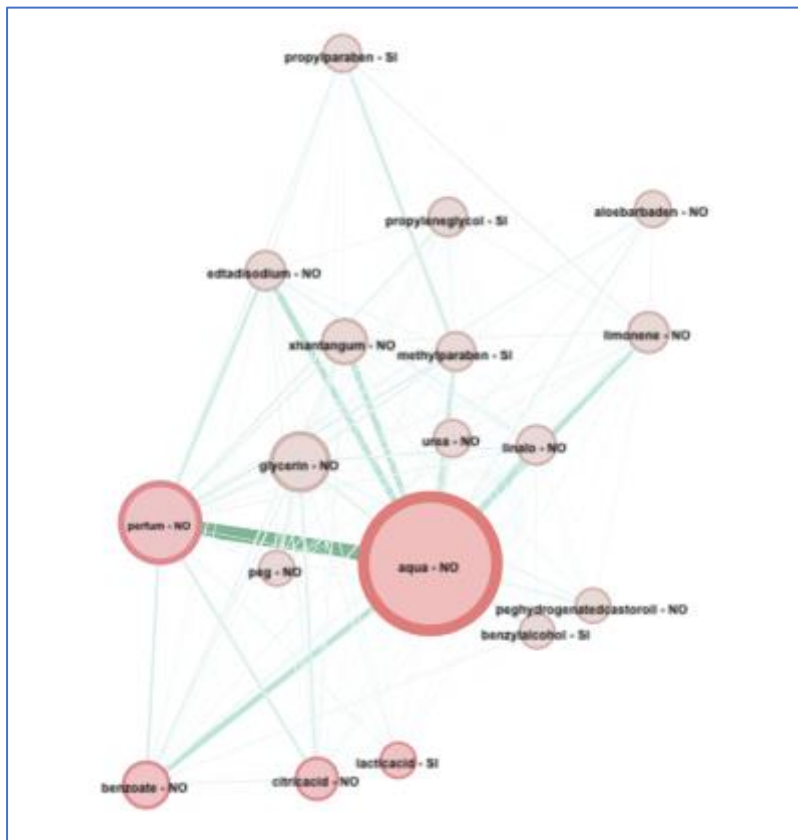
Con layout de Force Atlas, podemos ver como los componentes de los diferentes componentes se distribuyen alrededor de los componentes comunes a todos los componentes, así como ya podemos ver cuáles son las uniones que más peso tienen.

Vamos a ver ahora el grafo, pero seleccionando solamente los nodos que están marcados como tóxicos.



Lamentablemente vemos que los nodos tóxicos aparecen en los diferentes productos, tanto en cremas, desodorantes, champús y pastas de dientes, si bien parecen estar en Pasta Dentífrica en menor proporción.

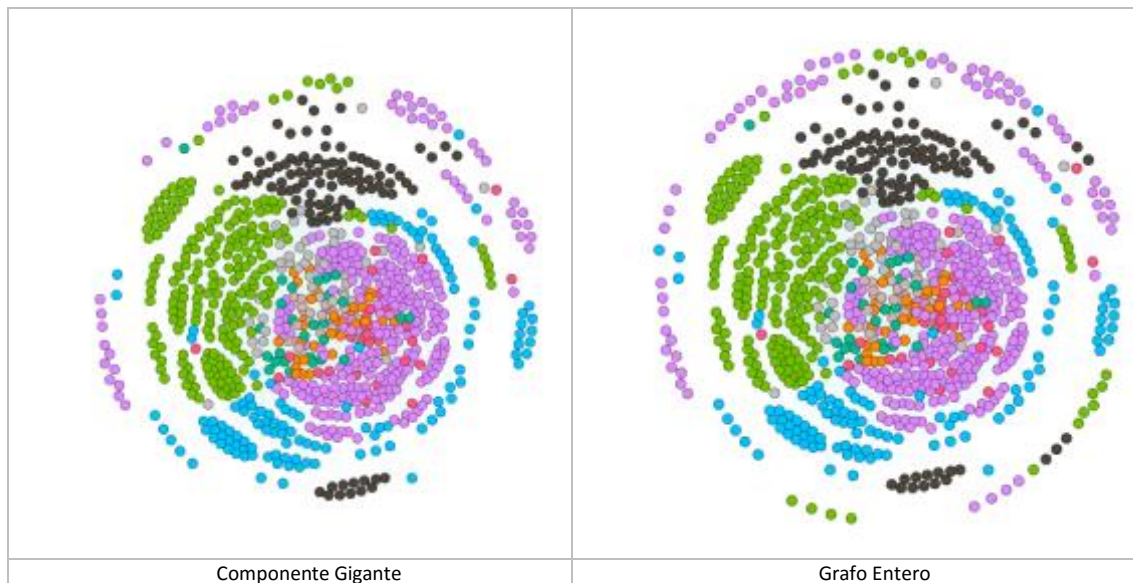
Los componentes que aparecen en todos los productos, marcando las aristas según el peso que tiene:



Con lo que vemos que agua y perfume son los que más relación tienen, por lo que esa combinación se da con mucha frecuencia en todos los productos. Como detalle añadir, que detrás del nombre de cada componente aparece el indicador NO (No es tóxico) o SI (Si es tóxico).

COMPONENTE GIGANTE

A continuación, se muestra cuál es el componente gigante del grafo:



Se puede ver que existen “islas” o nodos en el grado de la derecha que no están relacionados con el resto.

Por ejemplo, haciendo zoom en una de las “islas”:



He buscado estos componentes y son componentes de origen vegetal, que solo he aparecen en 3 de las cremas del estudio, y por su composición o tipo de crema no se relaciona con el resto.

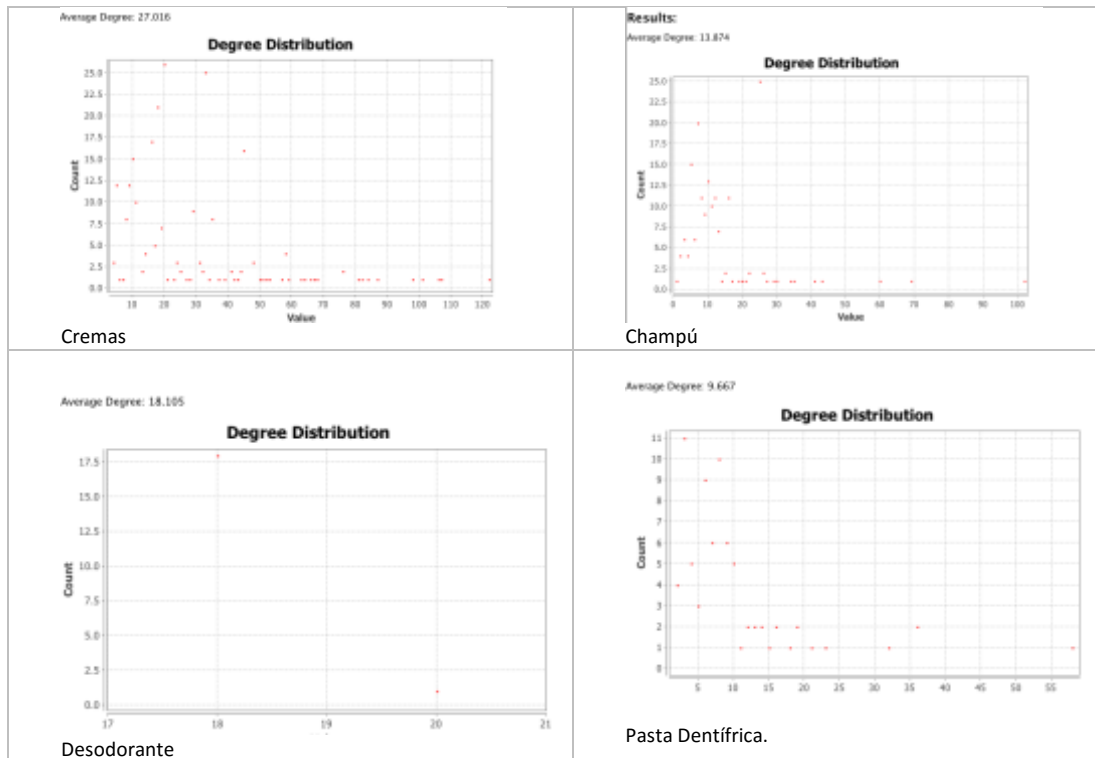
Puede ser que, analizando la información, pudiéramos llegar a la conclusión de que existen menos productos en el mercado con compontes de origen vegetal, o que existen y simplemente son menos comerciales. También hay que tener en cuenta que la página de donde he obtenido la información no se caracteriza por tener productos naturales, como sería el caso de los productos más de herbolario. También hay que añadir que no tengo conocimientos químicos como para poder realizar un estudio de los tipos de componentes según el origen.

Dicho lo cual, no podemos hacer un estudio de tipo de compontes, ni extraer conclusiones sobre productos o no naturales.

Para realizar las distintas mediciones solo voy a tener en cuenta el grafo gigante

DEGREE

Vamos a ver el número de grados o conexiones que tiene los nodos.

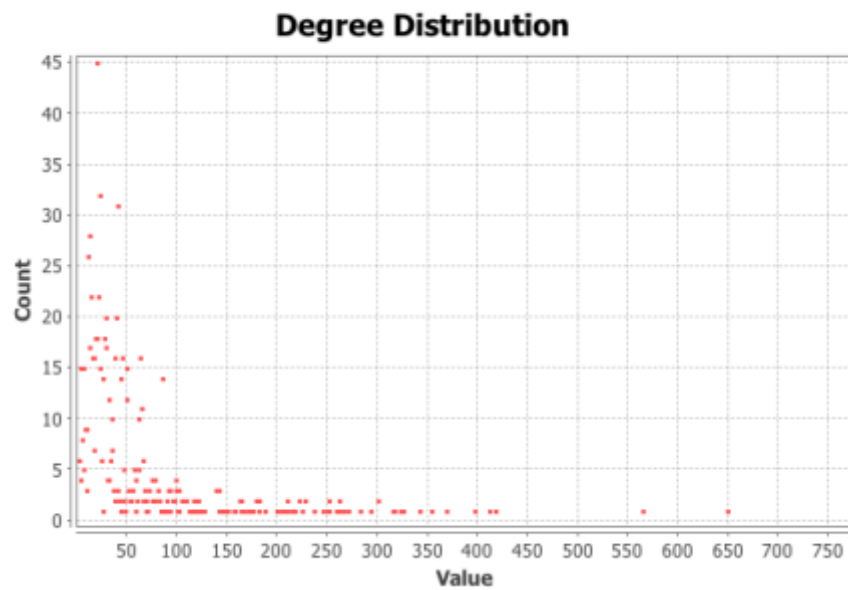


Vemos que, por producto, la media de grados en la crema es superior al resto. Es decir que tiene componentes que están muy relacionados, como se ve en la gráfica hay componentes que tienen más de 100 relaciones). El que menos relaciones tiene es el de la pasta dentífrica, porque tiene elementos muy poco relacionados en contraste con alguno que tiene más de 35 relaciones.

Esto significa, que en las cremas existen diferentes productos con más elementos en común que en el resto de productos.

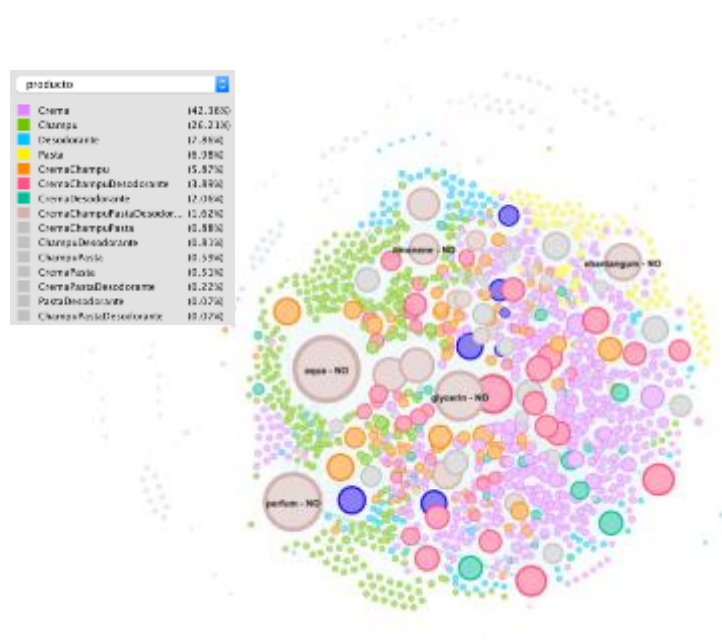
Veamos la medida del grafo entero.

Average Degree: 53.848



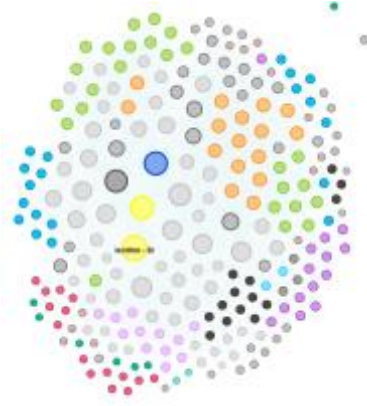
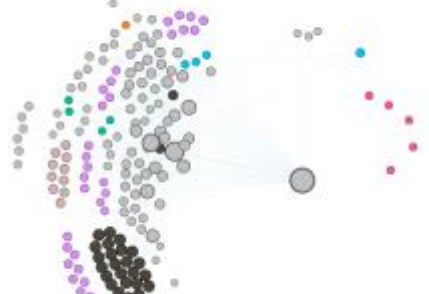

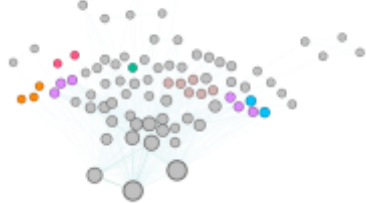
La media ha subido considerablemente, porque tenemos componentes que forman parte de distintos productos, por ello, ahora tenemos más conexiones.

Visualmente, vamos a mostrar los nodos con más tamaño según el número de grados:



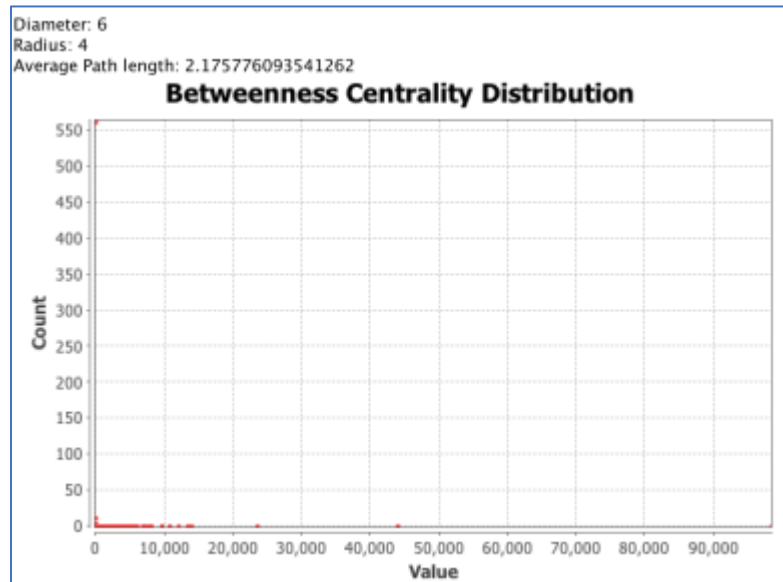
Ya vemos que los nodos con más grados, corresponde a componentes que están en varios grupos y son el agua, el perfume y la glicerina (ninguno tóxico), lo que tiene toda la lógica, puesto que son los componentes que “unen” los diferentes productos.

A continuación, vemos cuales son los componentes por tipo de producto que tienen más relaciones. Son lo que normalmente nos vamos a encontrar en los productos.

	<p><i>Cremas</i>, el componente con más grados es Lecithin, que es tóxico.</p>
	<p>Champú, es polyquaternium el que tiene más grado, que no es toxico.</p>
	<p>En Desodorante, es aluminumchlorohydrate, que no es tóxico.</p>
	<p>De la pasta dentífrica, es sodiumsaccharin (sacarina).</p>

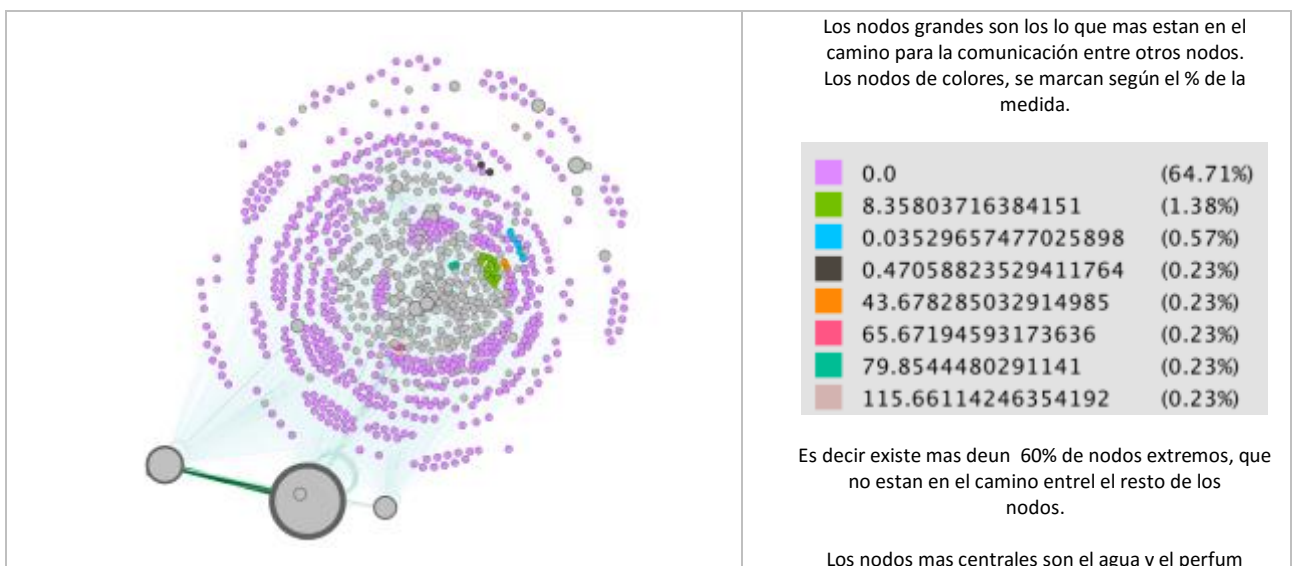
BETWEENNESS

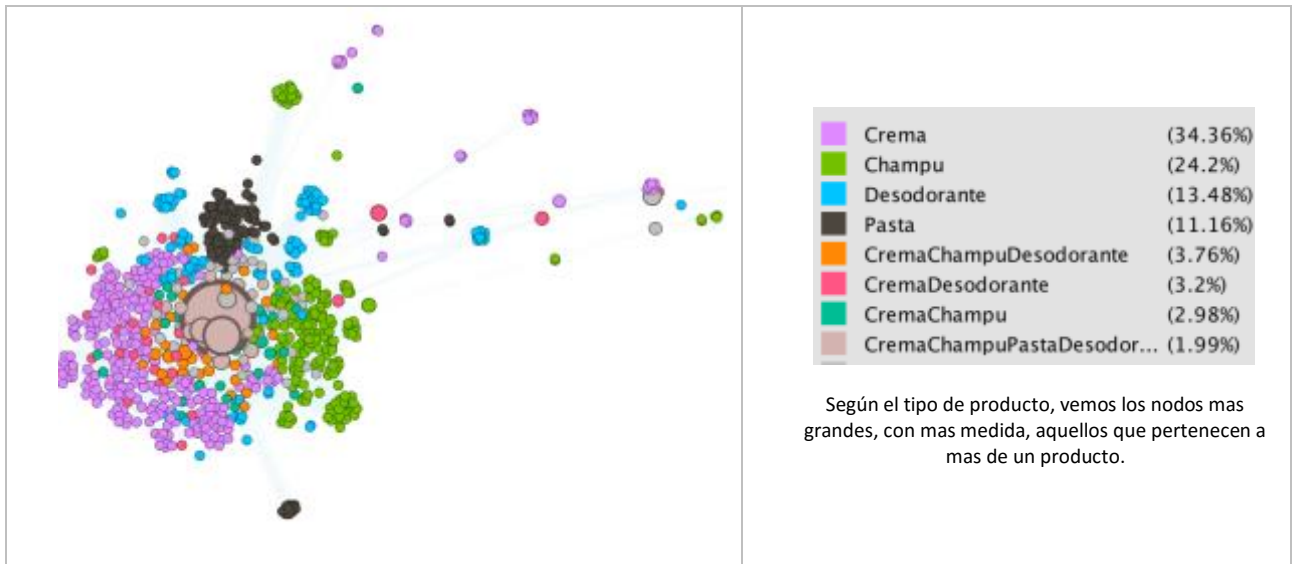
Esta medida nos dice cuántas veces un nodo aparece en el camino de conexión de otros nodos, es decir nos indica la centralidad que tiene un nodo.



Como punto a destacar vemos que hay más de 550 nodos que no aparecen en el camino de ningún nodo

También nos indica que el diámetro es 6, eso es la máxima distancia entre dos nodos.





¿ Cuales son los nodos con mas Betweenes según el tipo de producto “puro” ?

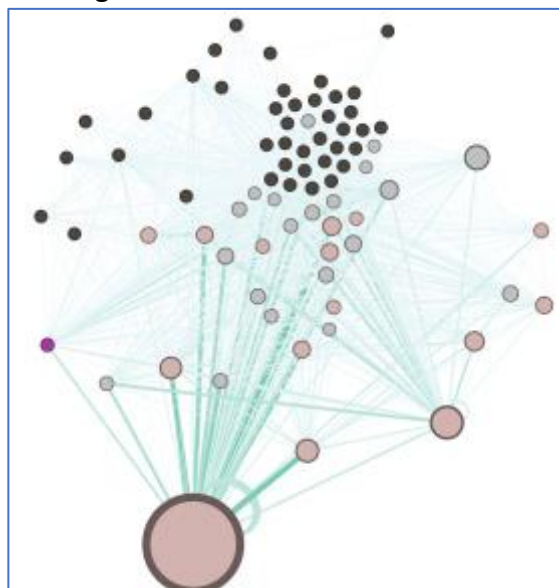
Vamos a ver si cambia con respecto al numero de grados de cada nodo que hemos visto antes.

Filtrando en el grafo por componentes gigante y tipo de producto, y luego visualizando los datos, podemos llegar a la información:

- Pasta Dentrifica : sodiumfluoride
- Desodorante : aluminiumchlorohydrate
- Champu: guarhydroxypropyltrimoniumchloride
- Crema: behenylalcohol

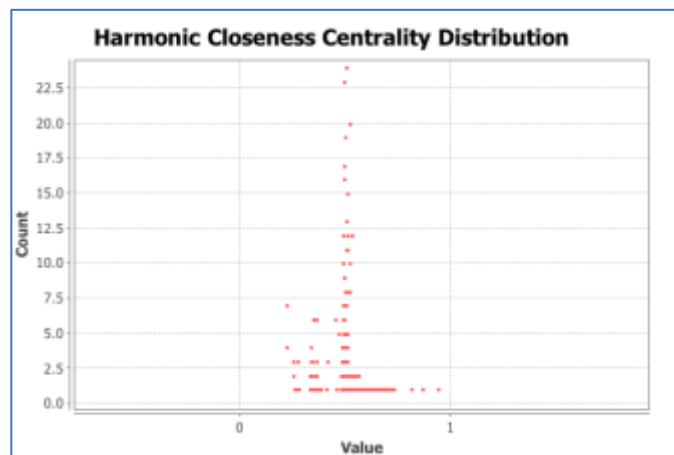
Estos son los componentes por tipo de producto, que estan en el camino mas corto entre el resto de componentes del grafo.

Por ejemplo, en el caso del componentes de la pasta : SodiumFluoride, vamos a buscar sus relaciones en el grafo:

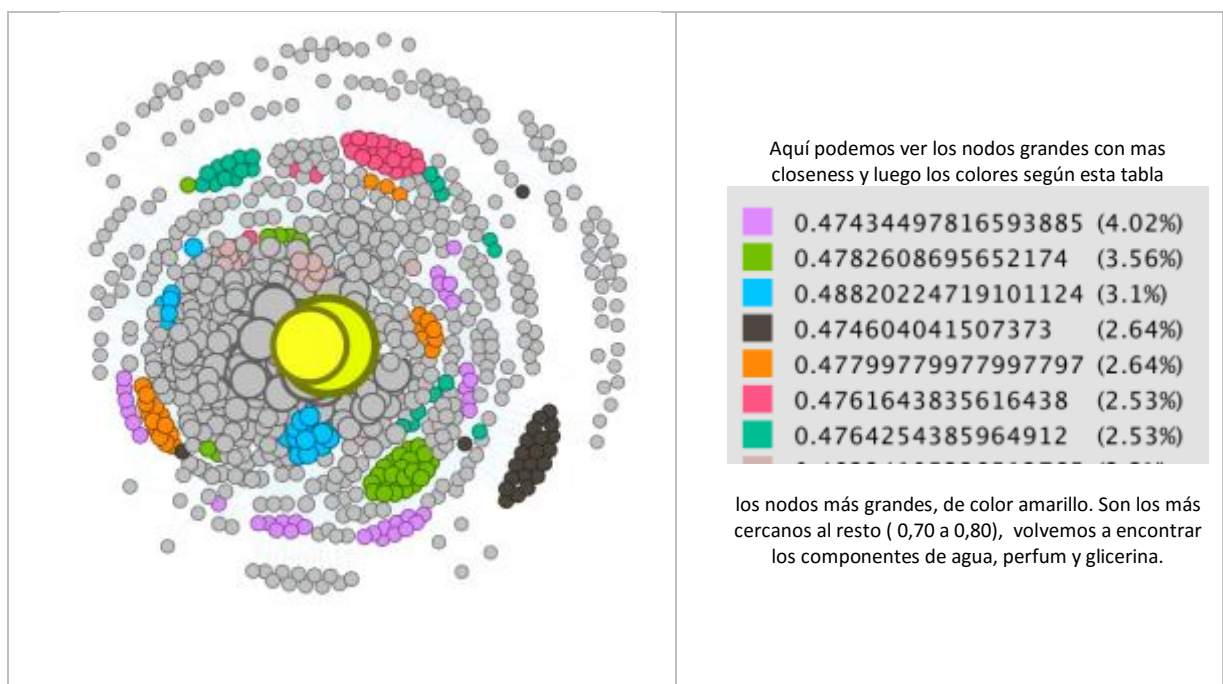


Es el nodo de la izquierda, que he coloreado de color rosa. Vemos que ese nodo es el que esta relacionado con el nodo grande (recordemos que es agua, y que esta en todos los productos), por lo que digamos que ese nodo de sodiumfluoride, es el que por estar relacionado con el agua, y con los demas componentes de las pastas, el que relaciona los componentes de la pasta (que solo estan en pasta) con el resto de componentes de los cosmeticos.

CLOSENESS



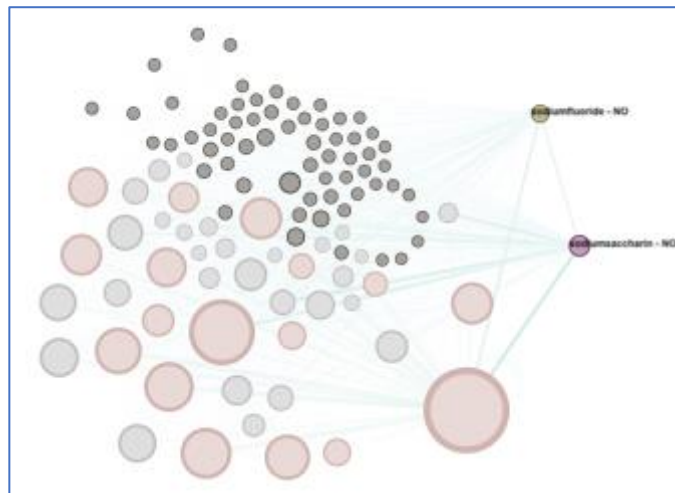
La media de distancia media entre un nodo y los demás. Cuanto más se acerca a 1 mas están los nodos relacionados.
Como hemos visto anteriormente, hay pocos nodos que estén muy relacionados.



Sin embargo, si filtramos por cada uno de los productos “puros”, los componentes cambian con respecto a la métrica anterior:

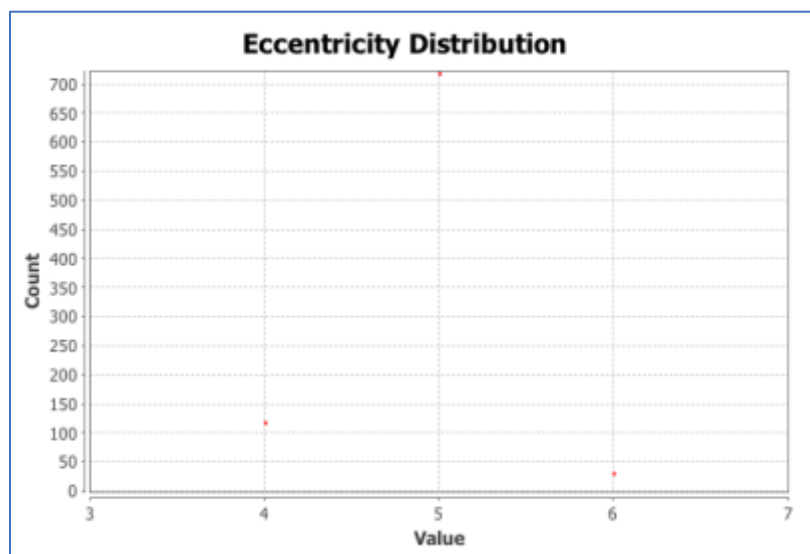
- Cremas: butyrospermumparkii
- Champú: polyquaternium
- Desodorantes: aluminumchlorohydrate
- Pasta: sodiumsaccharin

Ejemplo, viendo el componente de Pasta, el elemento ahora más cercano al resto es sodiumsaccharin , sin embargo el que tiene más Clossenenes es sodiumfluoride, la razón por la creo que es así, es porque sodiumsaccharin tiene más grados, por lo tanto tiene más relaciones directas que sodiumfluoride.

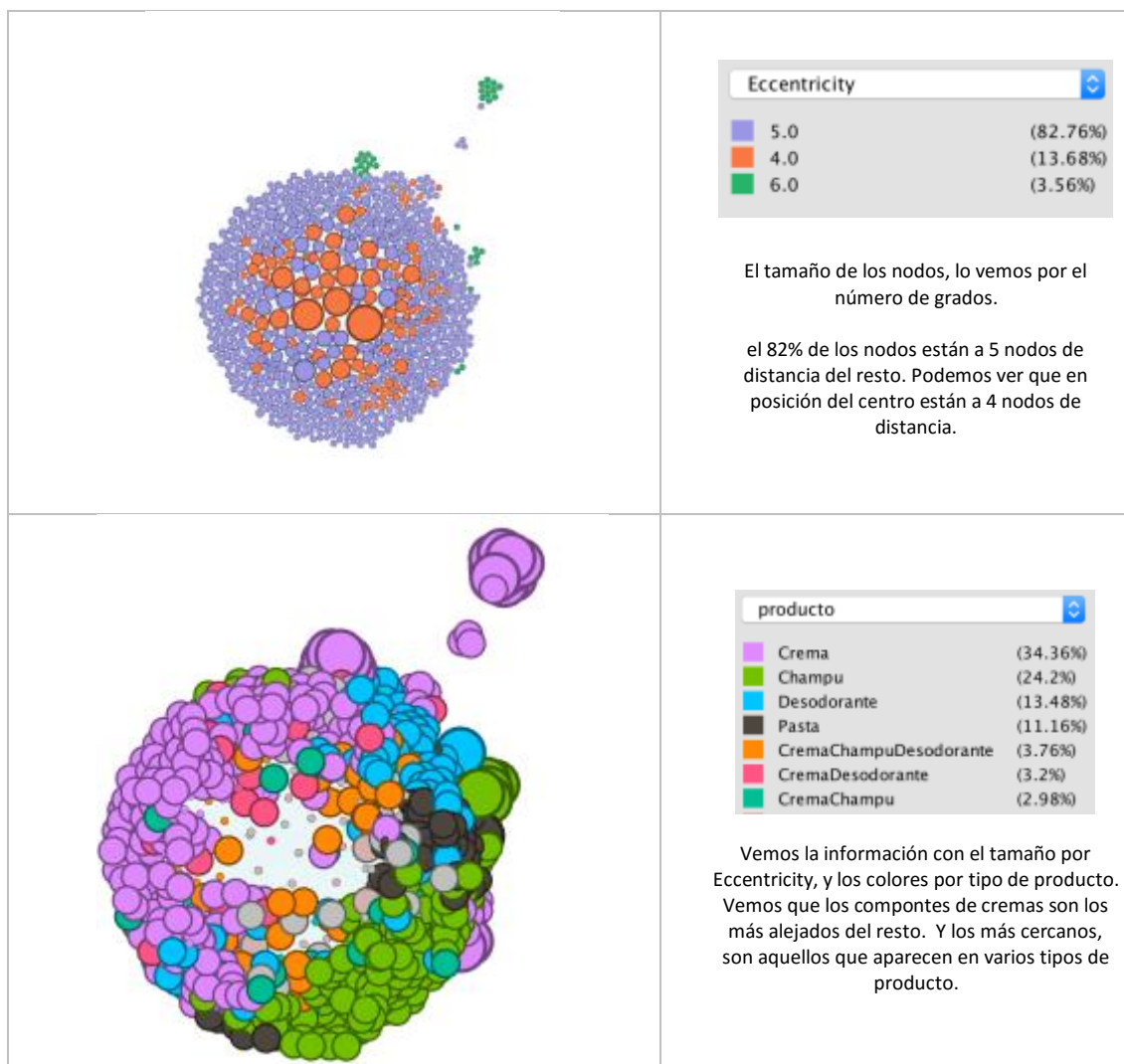


ECCENTRICITY

El número de nodos necesario para conectar un nodo con el más lejano.



Vemos, que lo más habitual son 5 enlaces de distancia.
Hemos visto que los nodos no están fuertemente relacionados entre si y que el diámetro del grafo es de 6.



DENSITY

Esta medida nos indica si el número de aristas está cercano al número de máximo de aristas. La densidad de nuestro grafo es la siguiente:

Results:

Density: 0.062

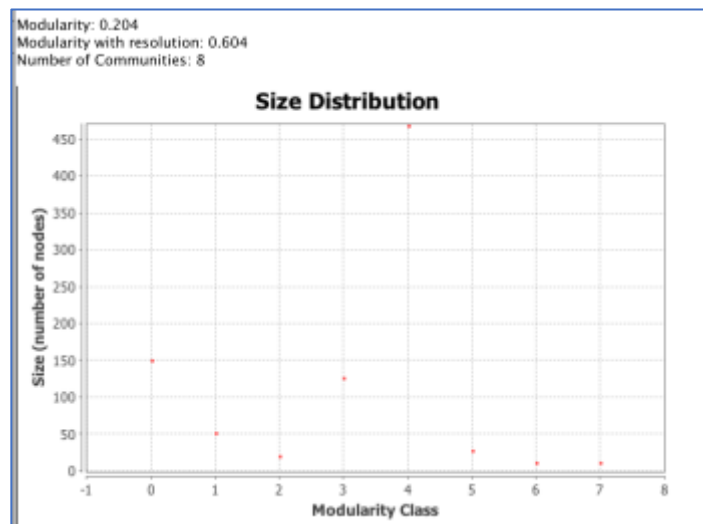
Tiene poca densidad, ya que la densidad máxima es 1, y este dato es muy bajo.
La densidad de los componentes seleccionando todos los productos que aparecen:

- en las cremas es: 0,172
- en champú, es 0.140
- en desodorante, es 0.164
- en pasta, 0,230

Lo que quiere decir, que son los componentes de pasta los que más relación tienen entre ellos.

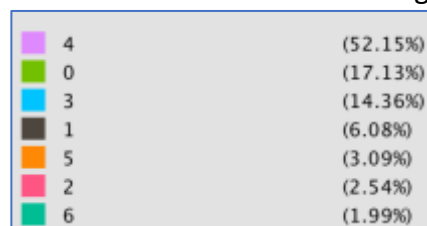
MODULARITY

Vamos a ver ahora los grupos que existen según los datos que tenemos.
Jugando con los parámetros, para que el número de comunidades se acerque al número de productos que tenemos.

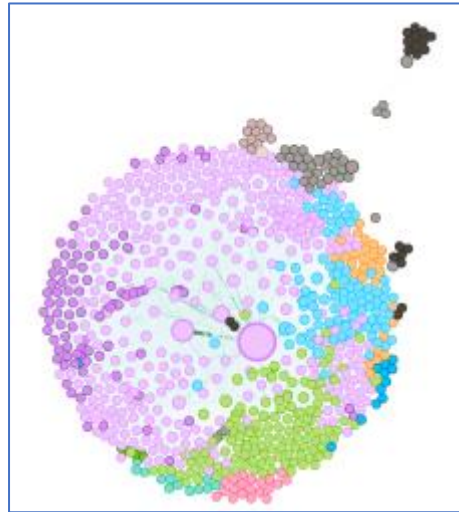


Ha encontrado 8 grupos, siendo el más numeroso el 4

En el siguiente grafico vemos los nodos de acuerdo a la siguiente leyenda:

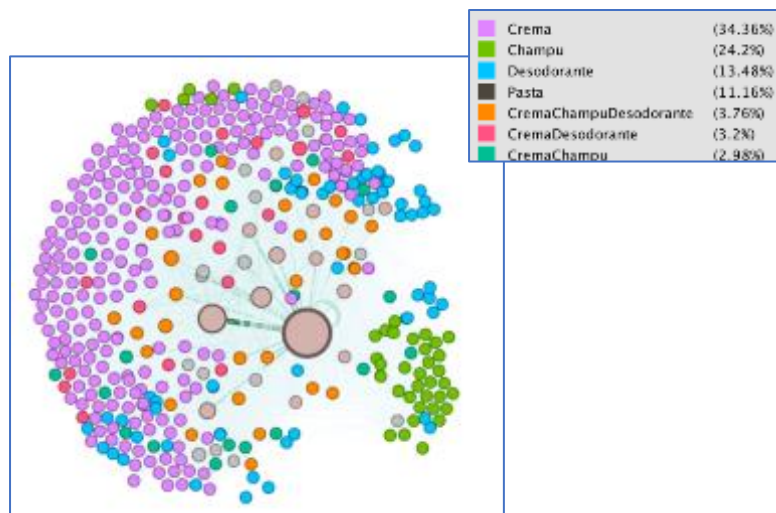


El tamaño de los nodos es por la medida de Betweenes.



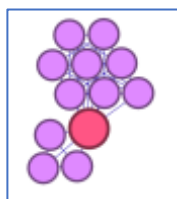
Revisando los datos, vemos que los grupos que ha encontrada nada tienen que ver con los tipos de producto.

Ejemplo, grupo 4:



Está formado por elementos de distintos productos.

Y luego, por ejemplo, el grupo 7:



Está formado por manteca, macademia, camelia, granada soja.

La impresión que tengo, es que los grupos se realizan en función de los elementos químicos que pueden y suelen juntarse para formar un producto. Por lo que la

combinación de soja, granada.. que se da en algunas cremas, no son elementos que pueden unirse a otros más químicos como el triethylcitrate.

5. CONCLUSIONES

Después del análisis con la herramienta gephi de los datos obtenidos, a modo de resumen podemos extraer las siguientes conclusiones.

Los elementos que según varias páginas web se mencionan como tóxicos aparecen en todos los productos que hemos analizado.

El grafo resultante no tiene mucha densidad, y también hemos visto que por ejemplo la métrica de eccentricity, que los componentes no están altamente relacionados, si bien, existen unos pocos elementos que se encuentran con muchísima frecuencia y que relaciona los distintos componentes, como son agua, perfume y glicerina.

Los productos analizados son en su mayoría formados por elementos puramente químicos, de forma que aquellos que tenían elementos más naturales (como lavanda, proteína de trigo) somos capaces de detectarlos, porque no se relacionan con el resto.

De los datos obtenidos tras el scrapping, con el número de productos y el número de ingredientes, podemos ver las medias de componentes por producto:

- cremas: 30
- champús: 18
- desodorantes: 19
- pasta: 14

De acuerdo al número de grados que hemos visto de cada componente, las cremas que son las que más componentes tiene de media, son las que más grados tienen sus componentes. Y luego hemos visto que la pasta dentífrica tiene menos componentes, pero menor número de grados, y sin embargo tiene más densidad que el resto.

También hemos visto que hay muchos componentes con 0 de betweenness, esa información de centralidad y la densidad del grafo, nos indica que es verdad que los productos son distintos en su mayoría de componentes.

También hemos podido ver cuáles son los componentes que más veces nos vamos a encontrar en los distintos productos, por lo que, a igualdad de componentes, habría que fijarse en el precio a la hora de comprar ;-)

En cuanto a la modularidad, hemos visto los diferentes grupos que se han generado, que no se parecen a los grupos por tipo de producto. La verdad es que no podemos extraer una conclusión de esos datos, porque no tengo conocimientos de química para poder realizar un análisis. Como he comentado, la impresión después de ver los

datos, es que los grupos se han generado según los elementos químicos que siempre tienen que estar juntos por sus propiedades.

Es decir, que en principio nos tendremos que creer el gran número y variedad de productos cosméticos existentes, que verdaderamente son distintos, y que eso explica las diferentes calidades y distintos precios. Otra cosa que no se puede medir aquí, es la efectividad de acuerdo a los productos, o que efecto tiene en realidad en el organismo los elementos que calificados como tóxicos.

Podemos ver también, la importancia que tiene el agua, sin ella no se podrían la inmensa mayoría de productos cosméticos.

Me pregunto a título personal, de donde y como obtienen el agua las diferentes industrias cosméticas, y cómo influye el agua en el precio final del producto. En relación a esto, buscando un poco por Internet, he encontrado este artículo :

<http://cosmeticwaterprocess.blogspot.com.es/2008/11/introduccion-la-industria-quimica.html>

que habla acerca del agua, la contaminación de la industria cosmética, así como los peligros en el uso de cosméticos.

Bueno, pues después de esta práctica, he aprendido a ver como a través del análisis de la relación de los datos podemos llegar a entender mucho mejor la información que nos rodea, y a pesar de que he descubierto más cosas en relación a los cosméticos de las que me gustaría, tengo que decir, que pienso seguir usándolos.