

## 7 Análisis de resultados

La variedad de formas en las que se puede presentar la información proveniente de Twitter hace que resulte poco conveniente crear datos propios debido a que no reflejarían la estructura verdadera de un gorjeo, por ello se optó por ejecutar el sistema *Poblador DB* el tiempo necesario para contar con datos suficientes para que los resultados sean representativos.

El análisis se realizó para los tres temas definidos actualmente en el sistema (Ver tabla 6).

Tema	Numero de gorjeos recuperados
Tráfico	313
Homicidios	100
Asaltos	365
<b>Total</b>	<b>778</b>

Tabla 6 – Temas y gorjeos recuperados para el análisis.

La recuperación de gorjeos se hace, como se explicó en capítulos anteriores, a partir de una lista de términos truncados que se relacionan con cada uno de los temas antes listados (Ver apéndice C).

Como ejemplo podemos citar el tema de tráfico vehicular o simplemente *tráfico* donde los gorjeos contienen alguno de los términos truncados que se relacionan con tráfico, como lo son *trafic-* (tráfico), *congestion-* (congestión, congestionamiento, etc), *vialid-* (vialidad, vialidades, etc), *inund-* (inundación, inundaciones, etc), *manifest-* (manifestacion, manifestaciones, etc), *march-* (marcha, marchas, etc.), entre otros.

Es claro que un gorjeo que contenga alguno de estos términos truncados no necesariamente se va a relacionar con el tráfico. Por ejemplo en el gorjeo:

*Extrañe mucho el miercoles de cine y **trafico** de alimentos @ X ojala pronto todo se arregle para biemmmm jajaja creo haber visto indicios*

Es evidente que no está relacionado con el tema de *tráfico* vehicular, pero se apuesta a que la cantidad de gorjeos no relacionados al tema son pocos y que al calcular la distancia energética sobre ellos se alejarán lo suficiente para ser agrupados en niveles superiores, característica propia de un algoritmo de agrupamiento jerárquico.

Como análisis cualitativo se realizó la lectura de los grupos generados. Para esto se ejecutó el subsistema *Procesamiento de gorjeos* y se reportó el dendograma normalizado<sup>25</sup> del agrupamiento jerárquico para grupos con por lo menos dos gorjeos, dejando fuera gorjeos que no se alcanzan a incorporar en algún grupo en un determinado índice de fusión. Con el fin de que el lector no se sorprenda por una aparente *aparición* de nuevos gorjeos uniéndose a un grupo o formando uno nuevo, se invita a que se dé lectura del grupo final generado, el cual contiene todos los gorjeos que en un principio son grupos individuales y que han llegado a fusionarse todos en uno solo.

Los archivos generados para el análisis tienen la siguiente estructura:

```
----- Grupos para: Nivel=N, Distancia=X-----

///////////////////////////////// Número de gorjeos: Y //////////////////////////////////
IdGorjeo: gorjeo
...
///////////////////////////////// Número de gorjeos: Y //////////////////////////////////
IdGorjeo: gorjeo
...
W gorjeos en: Z grupos:
```

Las líneas de guiones (‘-----’) indican un aumento en el nivel *N* de agrupamiento e indican la distancia de corte o índice de fusión *X*. Los grupos contenidos en ese nivel están enlistados después de una serie de diagonales ‘////’ donde se indica igualmente el número de gorjeos *Y*, el identificador del gorjeo y el gorjeo en sí mismo. Al final del nivel se presentan los *W* gorjeos que componen al nivel *N* y los *Z* grupos que se formaron.

Algunos ejemplos de grupos generados para el tema *Tráfico* (el 31 de mayo de 2011) en el nivel 96 con un índice de fusión 0.572 son siguientes (se presentan solo algunos con el objetivo de dar un marco de referencia sobre los resultados):

```
----- Grupos para: Nivel=96, Distancia=0.5720524017467221-----

///////////////////////////////// Numero de gorjeos: 101 //////////////////////////////////

75259946027057152 -> @mhonividente ya en periodico metro te leo en el trafico del df saludos
75265723966046208 -> @Annie_Islas Y con el trafico es peor...
75265835916214272 -> Gana Trafico O Muere JAJAJAJ http://fb.me/unyCod5g
75266924686221314 -> Muero de calor y estoy atrapado en el trafico en periferico
75268413357948928 -> Noo!El tiempo se me viene encima! Mucho tráfico! Ni he llegado a constituyentes D:
```

---

<sup>25</sup> La normalización del dendograma es únicamente con fines demostrativos y se realiza tomando el índice de fusión máximo del dendograma para después dividir cada índice de fusión entre el valor máximo.

75269211961831424 -> Ay mucho trafico me extrezaaaa!

75270548594241536 -> Pinche trafico en la del valle, pinches niños ni les gusta estudiar, pa que los traen a la escuela??

75272448949489665 -> Calor + trafico = Yo estresada!

75273295116767232 -> Piden a EU parar tráfico de armas <http://t.co/3NuGrJS>

75273626311589890 -> ya de lo encabronado en este terrible trafico le acabo de recordar a su mama a varios cafes jaja

75273811871805440 -> Q belleza las calles de L.A Nada de trafico... #MemorialDay

75277175976247296 -> Y ya llegue :) hubo algo de trafico pero todo normal :3

75277742664466433 -> Telefónica niega que obstruya el tráfico de llamadas de Axtel <http://tinyurl.com/3e6bf7m>

75277933392039936 -> @MAL\_ARTON: Disfrutando del calor y e trafico !!! Jajajajaj be happy <wooralee eso es ACTITUD! Bien!!!

75281324159680512 -> y todo esto de las obras publicas, el calor ,la gente odiosa, el trafico y tu me tienen hasta la ma#!&

75285789055004672 -> Me esta dando miedo. A donde se fue el trafico?

75287231908155392 -> Para variar, a vuelta de rueda.. (@ Trafico Tuneles de Santa Fe) <http://4sq.com/j4AvFs>

75288757510737920 -> esse papo de que querem liberar o uso da maconha pra inibir o comércio dos traficantes, liberar ou não, o traficante vai continua vendendo.

75291744924741632 -> @jorolverao no se te olvide el del trafico, revisa a quien sigo

75292453250400256 -> @HOSTINGWEBPRO buen sistema el de Site Analytics para medir el tráfico de mi sitio web [www.hostingwebpro.com.mx](http://www.hostingwebpro.com.mx)

75293030025920512 -> Este tráfico que da tiempo de pensar y ya tengo nombre artístico: Tamalta Guadarrama alias Tamalguada... Jajajaja #estoydesimple!!! Jajajaja

75293435120197632 -> @Elizabeth\_boni y con el trafico se siente mas se encierra el calor en el coche y es sofocante

75299290376978433 -> RT @Profepa\_Informa: Como resultado del 2do Operatino Nacional Contra el Tráfico Ilegal de Vida Silvestre se logró asegurar 1,899 ejemplares de flora y fauna

75300533107294208 -> Sorprendido d q no agarre nada d tráfico

75302383638417408 -> RT @malenahurtado:Antes me molestaba mucho el tráfico.. ahora tengo BlackBerry / con razon todo mundo tiene BB en el DF #enmiranchonoesasi

75304429175324672 -> Trafico, calor, tengo sueño, tengo hambre, arrrgggghhhh!

75305510307508224 -> Los oficiales del trafico #pormiquese mueran

75310274411900928 -> #DeboConfesar que odio el trafico y el calor de la cd de mexico

75313957048565760 -> @Formato21GRC Gracias y por fa como se encuentra el trafico en constituyentes hacia Toluca

75315669331222528 -> ámonos, voy a plaza satelite, alguien que tenga reporte del trafico? no? naidien? que gachos bye

75316161465683968 -> #PENSANDOENELFUTURO Obra que desquicio el trafico de #HUIXQUILUCAN..... CLARO EL TRIANGULO @colonosteca @AlexFerCa

75317154643329024 -> I'm at Trafico de Zaragoza (Calle Ignacio Zaragoza, Mexico City) <http://4sq.com/k24Nk4>

75317615270166528 -> A que tráfico

75319332502110208 -> No! No! No es normal el trafico de hoy! Esta cañon!

75320024612610048 -> Q calor q trafico y cuanta gente en el banco esto esta cabron

75321013398806528 -> Haha Q belleza necesitaba trafico para hacer tiempo en el Periferico y no me toco Nada !! Ley de Murphy

75321876569800704 -> @Luis\_arturoRS #GRRR SUERTE EN EL TRÁFICO! Abrazoooo :)

75322624271585280 -> Malvado calor ya no lo soporto!!! Trafico y calor, pesima combinacion!!!

75325031244238849 -> Tarde para dar clase odio el trafico!!

75325727658086400 -> @stabilimentum ha de ser porque siguen a gente que sólo escribe "estoy en el tráfico" o "ya va a empezar mi telenovela" ¡Compréndelos!

75326816277118976 -> Trafico <http://twitpic.com/54shxe>

75326999979237376 -> @alfredodelmazo cuando van a poner foco en el tráfico en Jesús del Monte y el Paso a Santiago? #fail #huixquilucan

75328916390285312 -> Debido al gran trafico y a la dilatación de las vías del metro, no se como regresar a casa?

75330855115038720 -> @guznava @thanissalazar yo digo que si. Es ejemplo de inseguridad de trafico de caminos gachos. En fin tantos y tantos. #ecatepec s ejemplar

75333009749970944 -> @labeledwell en el trafico reinis

75333277770203136 -> @Mrbrown16 ah, es el de tráfico del Universal! Jajaja...

75334787530899456 -> La Supervía solución a 1.5% del tráfico <http://bit.ly/jpGEOp>

75337874286985218 -> saliendo de la oficina a junta en santa fe esperemos el trafico este leve !!

75338579328503808 -> Q dice mi taxi q sta atorado entre el tráfico y la lluvia... O sea ya valió y no va a llegar.... #muymal

75338566468775936 -> @Formato21GRC y en este momento? como va el tráfico en el circuito en la misma zona?

75360927062568960 -> En el trafico, tomando the con mi papa, q fue por mi a la Of , platicando de lo mas agusto, amo a mi padre es un hombre extraordinario!

75361077759717377 -> "@Karyamm: Tráfico te AMO!"// no mames!!! :|#]

[#|2011-06-06T17:15:50.045-0500|INFO|glassfish3.1|javax.enterprise.system.std.com.sun.enterprise.server.logging|\_ThreadID=371;\_ThreadName=Ejb-Timer-Thread-10;|75361231359311872 -> Las 3 gotas de lluvia que cayeron, sirvieron para que el trafico se pusiera de locos!!!

75361421671661568 -> Ya es imposible circular en la ciudad, trafico por todos lados...

75361587900317696 -> Con este tráfico ya estuvo que no llegue a entrenar...

75361662110142464 -> Tráfico.... :( Tengo hambre... (@ Trafico de Marina Nacional y Laguna de Mairan) <http://4sq.com/1CXTm2>

75361658482065408 -> @YazminAlmazan Tráfico te AMO!"// no mames!!! :|// todo lo q amas hace daño porq no amar lo q te daña jajaja

75361840154165248 -> Atorado en el trafico wtf

75362000338817025 -> @monst5p por toda la ciudad esta pesadisimo el trafico  
 75361987462299649 -> @LAUcubanita Ah! No me consta, pero lo decía por ellos. He visto trafico en lluvia y literal la mayonesa tiene mas iniciativa al volante.  
 75368129588494336 -> Que pinchi trafico carajo  
 75368103994859522 -> Es una mentada de madre el trafico!!!  
 75368295804571648 -> Es increíble lo que voy a escribir, pero sí, en este semáforo si hace falta un Poli que "dirija" el tráfico.  
 75368465669685250 -> Atrapado en el trafico y cantando high and dry.  
 75368636264628224 -> En serio como amo a mi bicicleta en momentos de trafico...  
 75369063802617857 -> Una bomba ahuyentará al tráfico #espregunta  
 75369132593397760 -> He cumplido mis 2 horas en el trafico! #NPPVM  
 75369105154244608 -> PoRQuE. Ya QuieRo LlegARts. A. Mi cASiTa. PinKy TraFiCO Me DeSEspEroOoo !!  
 75369487616061440 -> Calor + tráfico = #fail  
 75369646198497280 -> El dia que alguien entienda la logica detras del trafico se va a ganar el premio nobel de la paz y de economia... Algo inedito  
 75370371662086144 -> @MarcePaquini jaja pobree amiga!! Yo tmb apenas voy a mi ksa y si el trafico esta del ascoo!!!  
 75370370038890497 -> "@dadomdadom: El tráfico está "inmamable" / @el\_gabe. Aguas!  
 75370364875714560 -> El tráfico raya en lo obsceno...  
 75370464372985857 -> Este trafico es una burla .....  
 75376755002707968 -> @Atteconqueso ya vez por que no tienes novio!! Tienes que dejar de quejaste y disfrutar de todo (eso incluye tráfico, lluvia y demás)  
 75377158528307200 -> Quiero llegar a mi casaaa. El trafico #memata.  
 75377840027213824 -> Pinshi trafico!!!a las 9pm??#noesdedios  
 75379034711789568 -> Atorada n el trafico yo y mi camiobetita, escuchando Al buen @digitaldiego. Bss pa ti  
 75379859613954048 -> El trafico al mil #noesdedios  
 75380625435148288 -> Maldito trafico de insurgentes... Ya no?  
 75380956273455104 -> Atorado en el tráfico (@ Pueblo de San Lorenzo Tezonco) <http://4sq.com/mExSWh>  
 75381884116418561 -> El trafico de esta noche #EsEmputante :@  
 75382713187708928 -> @dmimica qué lindo lugar para el trafico ☹☹  
 75383148313186305 -> Llegué y 792 trafifantes se columpiaron, pinche tráfico.  
 75383616389120000 -> @RafagaMix solo trafico  
 75385476692967424 -> @monerohernandez ¿tráfico?¿inundación?... bueno , ya llegaste.  
 75387392944635905 -> Estoy hasta el huevo del puto trafico!!!! -exclamo el siempre cortes doctor-  
 75391415168942081 -> Pff que tráfico,y eso qué la lluvia estuvo bien leve , ya van cómo siete accidentes los que veo  
 75391531653152768 -> @mpcl78 si! Es una delicia, ya imagino el trafico, pero no importa.  
 75394188375625728 -> llegando a casita (@ Trafico de churubusco :s) <http://4sq.com/m9yaKx>  
 75397313203417088 -> Finalmente no me tocó lluvia. Por el contrario el tráfico para bajar de Interlomas estuvo de lo más light. #WIN  
 75397400327499776 -> de regreso a casa.....mucho calor aun pero...nada de trafico wiiiiiiii  
 75399975550468097 -> @ralvarezbanzo al rancho en chihuahua a pasar unos días y descansar del tráfico de esta cd jaaa  
 75400250914914304 -> Vamos por Luky... (@ Tráfico 2o Piso) <http://4sq.com/m7whVp>  
 75400584609538048 -> RT @renobartilotti: Alguien olvidó decirle al tráfico que se quedara en su fucking casa!!! Y eso que vengo en (cont) <http://tl.gd/apum10>  
 75404611804073985 -> @dannie\_prado me estresaaa el trafico y la gente pendeja para manejar ajajaja  
 75405097915518976 -> Que mala vibra de la gente de trafico que no me dio chance de irme,... :(... Mugres envidiosos!!!  
 75407215502176256 -> RT: @Am\_der #frasesdenovioceloso....¿Por qué tan tarde? // Respuesta: es que había muuucho tráfico, si ajá. ☹☹  
 75408553350926336 -> I'm at El trafico horrible (Revolucion, Mexico City) <http://4sq.com/jcSvs5>  
 75409214436151297 -> Voy a cantar un corrido escuchen muy bien ms compás para la #reinadelsur traficante muy famosa nacida allá en Sinaloa la tía Teresa Mendoza  
 75412161198948352 -> Pobres de los q se chutan diario el tráfico de la tarde en Revolución. Malditos microbuseros dispararon mis nervios y @0oBeKeRo0 también ☹☹

//////////////////// Numero de gorjeos: 12 //////////////////////

75304148110819328 -> Zona Poniente: Obras viales sobre la Carretera Federal México-Toluca desde Carlos Echanove hasta Puente Ancona. Continua Alternativas  
 75314583706931201 -> Insurgentes y A. Obregón / Semáforo peatonal en verde, cruzo la calle y el imbécil de "Apoyo Vial" le da el paso a los coches.  
 75318153177071616 -> RT @DavidhijodeZeus Acabo de reportar un congestionamiento. Mi ánimo: acelerado, <http://bit.ly/khWjaR> con el GPS social @waze\_es  
 75318150576603137 -> Acabo de reportar un congestionamiento. Mi ánimo: acelerado, <http://bit.ly/khWjaR> con el GPS social @waze\_es  
 75320190337941504 -> Usarían el #fondometropolitano para obra vial para coches de 1,700 mdp en Monterrey <http://bit.ly/mwj9X5> ... y MTY se irá llenando de autos.  
 75326726258962432 -> Gracias por a tu preferencia @ApoyoVial continua trabajando para Tí.... Recuerde que puedes marcarnos al 5905 VIAL (8524).

75327168477011969 -> 17:25 Congestionamiento vial sobre la Av. Benito Juárez de Sebastián Lerdo de Tejada a Instituto Literario.  
 75327340787412992 -> RT @ApoyoVial: Gracias por a tu preferencia @ApoyoVial continua trabajando para Tí... Recuerde que puedes marcarnos al 5905 VIAL (8524).  
 75327841474052096 -> 17:28 Congestionamiento vial sobre Nicolás Bravo de Gómez Farías a Sebastián Lerdo de Tejada.  
 75331574421393408 -> Identifique los señalamientos informativos y respete las indicaciones de los agentes de tránsito y bandereros de seguridad vial  
 75338297567756288 -> RT @xtrevi: #fondometropolitano a obra para coches de 1,700 mdp en Monterrey <http://bit.ly/mwj9X5> // Más #autos y + congestionamiento  
 75381795171991552 -> Identifique los señalamientos informativos y respete las indicaciones de los agentes de tránsito y bandereros de seguridad vial

//////////////////// Numero de gorjeos: 3 //////////////////////

75299708641357824 -> ZN Ligera carga vehicular en Av. 608 del Circuito Interior hasta Av. 412.  
 75300218299629568 -> Ligera carga vehicular en Av. 608 del Circuito Interior hasta Av. 412.  
 75303080278765568 -> Ligera carga vehicular en Av. 608 del Circuito Interior hasta Av. 412.

//////////////////// Numero de gorjeos: 2 //////////////////////

75317154643329024 -> I'm at Trafico de Zaragoza (Calzada Ignacio Zaragoza, Mexico City) <http://4sq.com/k24Nk4>  
 75408553350926336 -> I'm at El trafico horrible (Revolucion, Mexico City) <http://4sq.com/jcSvs5>

Las características que destacan sobre los resultados son:

- Los grupos de gorjeos más grandes contienen información general relacionada para cada uno de los temas. En el ejemplo anterior de *Tráfico* podemos recuperar las palabras más importantes: *trafic, calor, maldit, normal, ator, lleg y lluvi*, donde *trafic* es el término común en todos los gorjeos pertenecientes al grupo siendo la razón por la que están unificados.
- Los grupos más pequeños se componen por eventos o noticias muy específicas y generalmente son una mezcla entre retweets y comentarios sobre esos temas en específico, también es muy común encontrar retweets que son extendidos en su contenido por otros usuarios para informar más a los lectores. Cuando algo muy importante emerge éste tipo de grupos son los más grandes y se pueden observar otros grupos con los mismo retweets pero con términos que pueden extender la misma noticia en direcciones diferentes.

Ahora bien, los resultados cualitativos son buenos y alentadores, pero no proporcionan información de qué tan bueno es el proceso que se está realizando o en qué índice de fusión se debe hacer un corte al dendograma para obtener los mejores resultados.

Se realizó un análisis cuantitativo el cual consistió en la evaluación de las características de los grupos generados en función del índice de fusión y de la evaluación del algoritmo de agrupamiento.

Para la evaluación de las características de los grupos se utilizó la medida de recuerdo, medida muy utilizada en el área de recuperación de la información, la cual define la completitud<sup>26</sup> del algoritmo que en el contexto del presente trabajo indica qué tan rápido se unifican todos los gorjeos en un solo grupo al variar el índice de fusión. No se implementó una medida de precisión<sup>27</sup> al no haber una certeza para definir cuando un gorjeo no pertenece a un subtema. Y es que imponer los resultados de un conjunto de datos de prueba no reflejarían el verdadero comportamiento debido a que siempre llegan datos diferentes al sistema que pueden o no estar mal formados o que pueden hablar del mismo tema pero de aspectos totalmente diferentes.

Se definió el recuerdo  $r$  como la proporción de gorjeos integrados en un grupo con por lo menos dos elementos. El recuerdo es una función que devuelve un valor de 0 cuando todos los gorjeos se han unido en un solo grupo y 1 cuando no se ha generado ningún grupo con por lo menos dos gorjeos.

$$r = \frac{(total\_de\_gorjeos) - (numero\_de\_gorjeos\_asignados\_a\_un\_grupo)}{total\_de\_gorjeos}$$

Por otro lado, se realizó la evaluación de la cantidad de grupos generados en función del índice de fusión que junto con el recuerdo da una idea del índice de fusión más adecuado para realizar el corte del dendograma.

En la figura 15 se presentan los resultados de número de grupos (en función del índice de fusión) y recuerdo (en función de índice de fusión) para los temas Tráfico, Homicidios y Asaltos. Tómese en cuenta que los dendogramas para cada uno de los temas fueron normalizados con el objetivo de observar las tendencias de los agrupamientos en un índice de fusión de 0 a 1.

---

<sup>26</sup> Proporción de material relevante recuperado.

<sup>27</sup> Proporción de material recuperado realmente relevante.



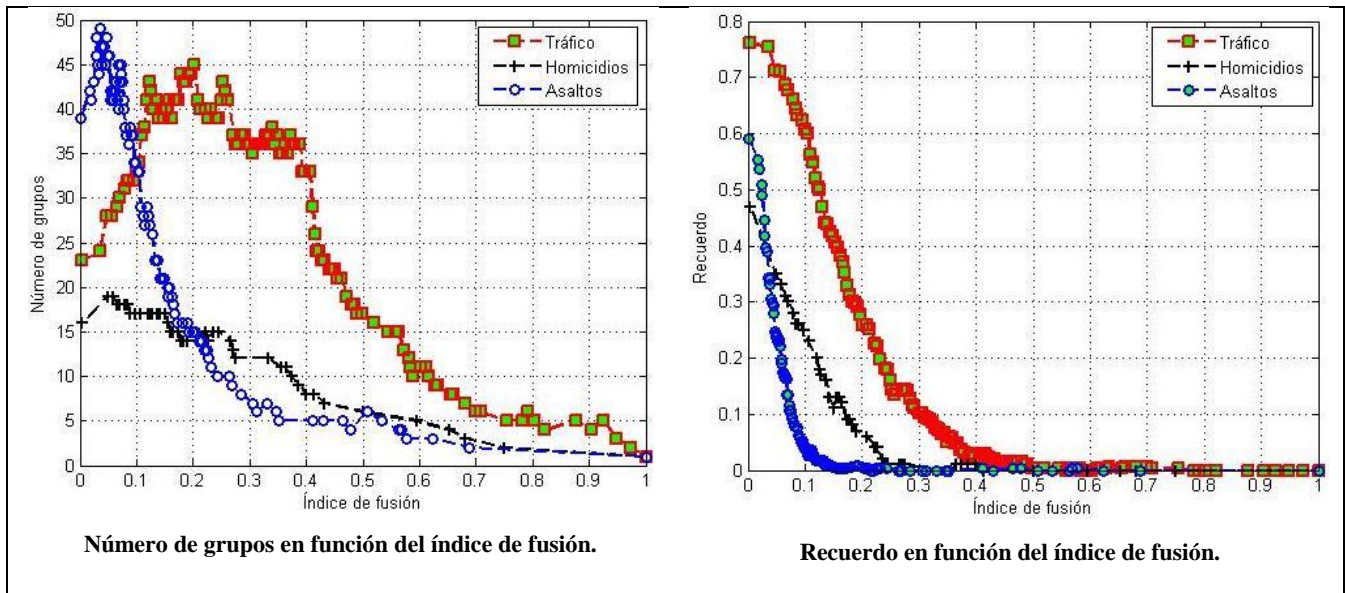


Figura 15 – Resultados de los agrupamientos para los temas Tráfico, Homicidios y Asaltos

De la figura anterior se puede observar que se inicia con un número muy alto de grupos y que el recuerdo es inferior al 20% (lo que significa que se han agrupado más del 80% de los gorjeos). Y es que realmente son pocos los términos que logran conformar el diccionario de términos, dando como resultado que un gorjeo termine siendo representado por pocos rasgos contenidos en él y que, en el momento de calcular la matriz de distancia energética, tiene una distancia de 0 con otros gorjeos, cosa que ocurre con los retweets donde las palabras que agregan los usuarios a nuevos retweets no afectan al cálculo de la distancia con un mismo retweet porque puede que los términos nuevos no aparezcan en más gorjeos. Por lo tanto, se tienen al inicio del agrupamiento muchos grupos formados por gorjeos semánticamente iguales.

Aclarado lo anterior se puede observar que, independientemente del tema analizado, el comportamiento del dendograma se puede dividir en tres zonas según su valor de fusión para hacer el corte posteriormente en el dendograma:

- Valor de corte entre 0.0 y 0.2: En esta zona el recuerdo es muy bajo y se encuentran la mayor cantidad de grupos conformados casi en su totalidad por gorjeos semánticamente muy idénticos y algunos muy cercanos a los ya mencionados.
  - Los grupos de gorjeos con mayor cantidad de éstos son semánticamente parecidos y dan información en general sobre la temática.
  - También se encuentran varios grupos formados por retweets y retweet semánticamente idénticos.

- Valor de corte entre 0.2 y 0.5: Se cumplen las características de grupos grandes para representar las generalidades del tema y se unifican grupos de retweets que son idénticos en grupos de retweets y gorjeos semánticamente parecidos donde se encuentran retweets y gorjeos que confirman los retweets existentes agregando más información.
- Valor de corte entre 0.5 y 1.0: Se forman pocos grupos donde destacan generalidades de diferentes subtemas. Se pierde la información muy específica para formar conjuntos de palabras alrededor de subtemas.

Retomando la idea de que se normalizó el dendograma para el análisis anterior y que cada dendograma tiene un índice de fusión máximo distinto para cada conjunto de datos, para presentar los datos, el índice de fusión máximo de cualquier dendograma se debe multiplicar por las proporciones de corte.

Ahora, una vez obtenido el dendograma, es importante evaluar su concordancia con la matriz de distancia para saber hasta qué punto el agrupamiento jerárquico es adecuado para los datos de la investigación. Para este fin, existen técnicas que permiten medir el ajuste, o lo que es lo mismo medir esa *distorsión*, ya que las relaciones de distancias son necesariamente distorsionadas al ser llevadas a una representación bidimensional [22].

Para medir esa distorsión se optó por calcular el grado de relación a través del coeficiente de la correlación producto-momento conocido como *coeficiente de correlación cofenético* (CPCC). Para su cálculo, primeramente se parte de la matriz de distancia y se forma el dendograma señalando en éste los valores del índice de fusión según su orden consecutivo de creación igual a como se mencionó en el capítulo 4. Con estos se confecciona una matriz de las mismas dimensiones que la matriz de distancias, la cual será la *matriz cofenética*  $C$  donde las relaciones entre cada elemento vendrán dadas por el valor de afinidad que le corresponda en la jerarquía del agrupamiento [22] [24].

Para el ejemplo del capítulo 4, los valores de  $C$  que relacionan a los elementos (1,2) y (3,4); se puede observar que forman grupos particulares y es claro que los valores correspondientes son 1 y 2 respectivamente. Para los valores correspondientes a elementos más alejados, por ejemplo 2 y 4, se asciende por el árbol y se toma el valor del índice de fusión que corresponde al nodo superior de unión de dichos elementos, en este caso 3. El mismo valor corresponderá a la relación (1,4), lo cual es típico de las matrices cofenéticas, donde siempre habrá varios valores repetidos.



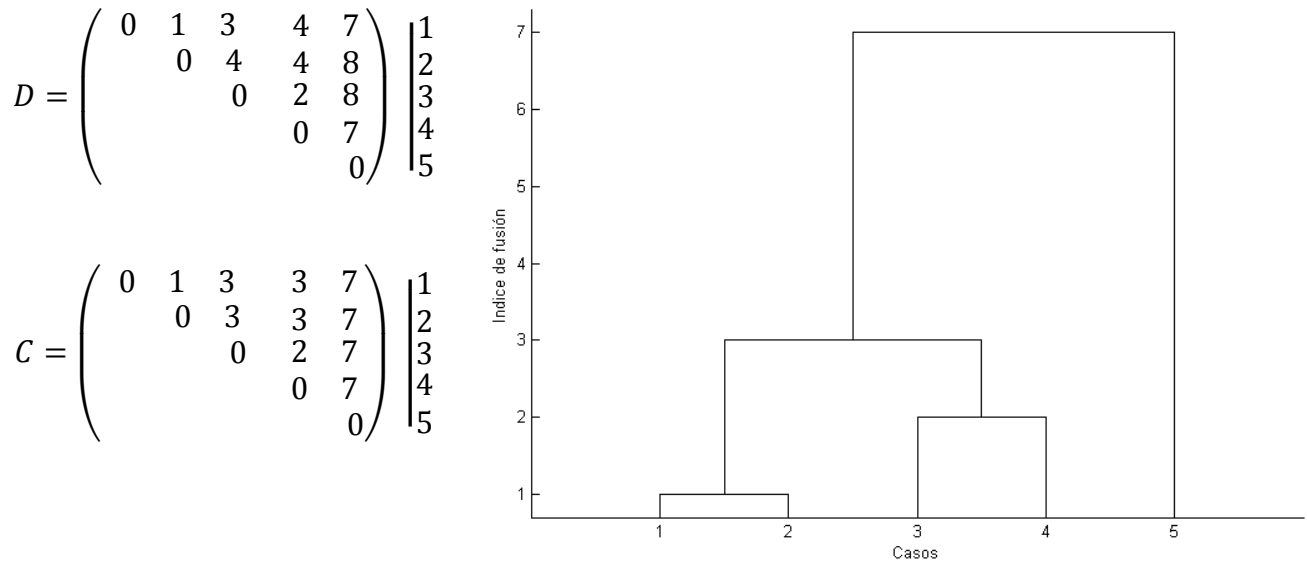


Figura 16 – Ejemplo del cálculo de una matriz cofenética.

Para calcular el CPCC son necesarias las dos matrices de  $n \times n$  elementos, pero basta tomar los  $M$  valores,  $M = (n - 1)/2$  de ambas matrices debido a que son simétricas.

Siguiendo la notación utilizada en el capítulo 4,  $d_D(i, j)$  representa la distancia que existe entre los elementos  $i$  y  $j$  y  $d_C(i, j)$  es la distancia cofenética o el nivel en el dendograma donde los elementos  $i$  y  $j$  son agrupados para un algoritmo de agrupamiento en particular. Por lo tanto, el CPCC para las matrices  $D$  y  $C$  se define como:

$$CPCC = \frac{\sum (d_D(i, j) - m_D)(d_C(i, j) - m_C)}{\sqrt{\sum (d_D(i, j) - m_D)^2 \sum (d_C(i, j) - m_C)^2}}$$

Donde

$$m_D = \left(\frac{1}{M}\right) \sum d_D(i, j)$$

$$m_C = \left(\frac{1}{M}\right) \sum d_C(i, j).$$

Una alta correlación entre matrices será señal de poca distorsión, por lo que escogiendo los resultados de la técnica que brinde el mayor valor de CPCC podemos considerar que las relaciones implícitas en la matriz de distancias están siendo reflejadas con la mayor fidelidad. Los valores de CPCC van de -1 a 1 y, según [22] [23] [24], los valores superiores a 0.8 indican una buena representación de la matriz de afinidad por parte del dendograma, aunque hay autores que consideran que valores superiores a 0.90 no garantizan necesariamente que el dendograma resumiera adecuadamente las relaciones.

Específicamente para el sistema desarrollado se utilizó la matriz de distancia energética y se implementó una pequeña aplicación que recorre el dendograma obtenido como resultado del agrupamiento de los gorjeos para formar la matriz cofenética y calcular así el CPCC.

En la tabla 7 se presentan los resultados asociados al coeficiente de correlación cofenético:

<b>Tema</b>	<b>coeficiente de correlación cofenético</b>
Tráfico	0.9618416733503183
Homicidios	0.9828575149307823
Asaltos	0.9912319664290344

**Tabla 7 – Resultados de coeficiente de correlación cofenético para los términos tráfico, homicidios y asaltos.**

Estos resultados al ser muy cercanos a 1 permiten un grado de confiabilidad muy alto creando muy baja distorsión, es decir, que el dendograma representa la matriz de distancia energética original de una manera confiable.

Con todo lo anterior se puede concluir que por un lado el algoritmo de agrupamiento implementado hace un trabajo excelente para agrupar los gorjeos y que el análisis cualitativo en conjunto con el análisis cuantitativo del dendograma permitió establecer la distancia de corte óptima de 0.3 del índice de fusión máximo.