



# PROYECTO DE ANALISIS DE DATOS –

MÉTODOS AVANZADOS DE INVESTIGACIÓN

Profesor Andrés Lombana.

Maria Mercedes Corral Strassmann

Febrero 24 de 2020

## CONTENIDO

PROYECTO DE ANALISIS DE DATOS – .....	1
1 INTRODUCCION Y CONTEXTO .....	3
2 MÉTODOS Y HERRAMIENTAS .....	3
3 DATOS.....	5
4 ANALISIS .....	8
4.1 Análisis de Usuarios: .....	8
4.2 Análisis de Tiempos: .....	9
4.3 Análisis de Frecuencia de tweets:.....	10
4.4 Análisis de Frecuencia de “url’s”: .....	11
4.5 Análisis de tweets: frecuencia y conteo de palabras.....	12
4.6 Análisis de tweets, menciones y re-tweets: .....	13
4.7 Análisis de contenido de algunos tweets: .....	15
4.8 Análisis de red:.....	17
11 CONCLUSIONES.....	23
12 REFERENCIAS .....	25
12.1 Otras Referencias .....	25

## 1 INTRODUCCION Y CONTEXTO

En Colombia, las (IES) - Instituciones de Educación Superior cumplen con varios objetivos misionales que trascienden la formación de profesionales, por ejemplo, la Javeriana, en su Planeación Universitaria 2016-2021, establece como objetivos misionales el ejercicio de la docencia, la investigación y el servicio con excelencia, como universidad integrada a un país de regiones, con perspectiva global e interdisciplinar.

En este sentido hay contribuciones teóricas que han resaltado los cambios en las maneras tradicionales de producir conocimientos y la forma en que estos pueden impactar la sociedad, particularmente a través de la innovación. El argentino Jorge Alberto Sábato propone en la década de los 60, una estructura triangular entre la industria, la infraestructura científica y tecnológica y el estado como modelo para el desarrollo de la política de ciencia y tecnología en América Latina (Sábato, Botana 1968). El **sistema de innovación** fue propuesto como un modelo funcional que motiva la innovación como el resultado de interacciones entre distintos tipos de actores (Lundvall, 1993), por su lado, el **modelo de la triple hélice** (Leydesdorff y Etzkowitz, 1993) resalta el papel de las organizaciones formales, y las relaciones entre los actores para que el conocimiento y la información incidan en el desarrollo económico y social de los países.

Las anteriores contribuciones ayudan a modelar la relación de la universidad con su entorno desde sus funciones de formación y de investigación y producción de conocimiento, pero que sucede con la tercera misión de la IES, donde ésta se enfoca en el servicio y en la contribución social al país.

El objetivo principal de mi investigación se focaliza en esa tercera misión de las IES, y en particular de la Pontificia Universidad Javeriana, con el fin de formular una propuesta de modelo que emerja de las diversas formas en que las universidades colombianas se relacionan con su entorno, que permita valorar las actividades en la heterogeneidad de sus impactos, beneficiarios y características.

Esto permitirá reconocer las distintas actividades que se realizan desde la universidad e impactan la sociedad, además de aquellas que ya se han popularizado y socializado.

En este proyecto se pretende realizar una pequeña investigación preliminar para consecución de datos utilizando herramientas avanzadas para este fin. Los datos se obtienen de Media Cloud o de la red social Twitter; para el caso de mi investigación relacionada con universidades, sus funciones, misiones, y relaciones con su entorno, la idea es buscar información a través de Twitter que permita realizar algún análisis relacionado con el propósito de la investigación y poder generar algunas conclusiones.

## 2 MÉTODOS Y HERRAMIENTAS

El éxito de la investigación estará siempre en los datos, en la naturaleza de estos, en los esquemas de recolección y finalmente en los métodos de análisis que se escojan. Walliman

(2011), en su libro nos presenta las formas en que encontramos los datos, estos se encuentran en dos formas, datos primarios, y datos secundarios. Los primeros son directos de la observación y los segundos llevan interpretación. Existen cuatro tipos de datos primarios de acuerdo con su esquema de recolección, por medida, por observación, por interrogación y por participación. Los datos primarios son los más confiables. Los datos secundarios, dependen de las fuentes para su confiabilidad.

Los datos base para este proyecto son datos primarios tomados a partir de participación en la red twitter con base en la creación de un .bin. A partir de estos primarios va una interpretación para generar un análisis lo cual los convierte en datos secundarios. En este caso se toman datos de una semana y mas adelante explico con cuál herramienta se recolectan los datos.

Adicionalmente está el método de análisis de los datos, Hernández Sampieri (2010), plantea los métodos de análisis de los datos, métodos cuantitativos y métodos cualitativos. Según Sampieri los fenómenos actuales de las ciencias son tan complejos que requieren métodos mixtos, cualitativos y cuantitativos.

Para este proyecto se utiliza inicialmente método cualitativo por el tipo de herramientas y de muestra que se obtiene. Los datos provienen de redes sociales, son datos no estructurados, necesitan interpretación. Posteriormente el análisis puede llegar a dar unos elementos cuantitativos producto de este, estos datos se refieren a conteo de palabras, de mensajes, usuarios que claramente son cuantitativos.

Como herramientas para este proyecto se utiliza TCAT (Twitter Capture and Analysis Toolset), que es un conjunto de herramientas para recuperar y recoger tweets de Twitter y analizarlos de varias formas. TCAT fueron desarrolladas por investigadores University of Amsterdam con el propósito de apoyar el avance de los métodos digitales de investigación. (Guía de TCAT).

Con la versión que cuenta la Javeriana DMI-TCAT instalada en servidores de Caoba y el laboratorio de Big Data (facultad de Comunicación + Ingeniería), para uso académico de estudiantes y profesores, se creó la base para generar los datos para este proyecto. (Guía de TCAT).

Para el análisis se tomó desde Admin, creando un “bin” y algunas palabras seleccionadas, con los datos iniciales se realizan algunos “queries” y exclusión de algunas palabras para refinar un poco las búsquedas.

Los datos recolectados se descargan en archivos **CSV** para su posterior análisis. Se utilizaron otras herramientas externas a TCAT para el análisis y las estadísticas, Open Refine, Excel y para el análisis de redes se hizo con Gephi.

Inicialmente se generan un número de datos considerable que pueden ser la muestra, pero con un análisis rápido se ven datos que no aportarían al objetivo; por lo tanto, se realizan unos filtros sobre los tweets que me permitan tener un data set más afinado.

Se realizan filtros por país, incluyendo solo Colombia y posteriormente excluyendo Cuba para llegar a datos que estén más de acuerdo con el objetivo.

### 3 DATOS

Se crea inicialmente un “bin”, desde Admin, utilizando TCAT de la Universidad Javeriana, y este “bin” se crea con las siguientes condiciones:

- Fecha inicio: 10 02 2020
- Fecha fin: 17 02 2020
- El “bin” se crea con algunas palabras claves que ayuden a recibir tweets que sean de utilidad para el proyecto de investigación. Estas palabras fueron:
  - Universidades
  - Misión
  - Extensión.

Para estas palabras claves se presentaron 71.000 tweets, como base de la recolección planteada.

Una primera revisión general de los datos muestra datos de universidades en el mundo por lo cual se hace un filtro solamente para Colombia. Este filtro deja un conjunto de 2.220 datos. Iniciando algún tipo de análisis se observan muchos datos de Cuba con lo cual se refina la búsqueda a excluir a Cuba.

Por lo tanto, la muestra para el análisis se determina como:

- Palabras claves: universidades/misión/extensión
- Query: Colombia
- Excluir: Cuba
- Total de tweets: 1.354

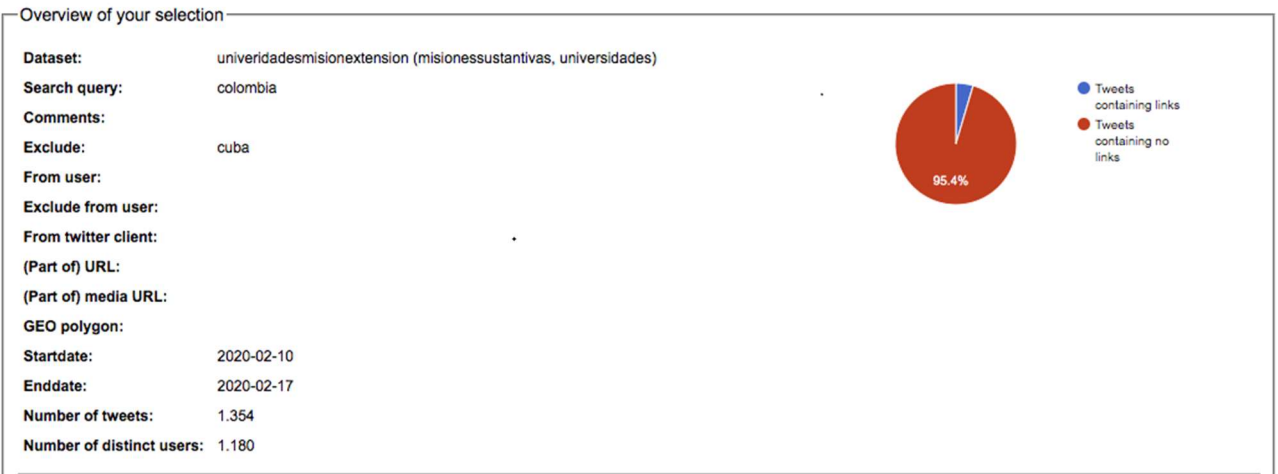
Las siguientes Gráfica 1a y Grafica 1b muestran resultados de las búsquedas ya filtradas.

**Select the dataset:**  
universidadesmisionextension --- 115.129 tweets from 2020-02-11 00:02:00 to 2020-02-21 16:43:17 21.822.217 tweets archived so far (and counting)

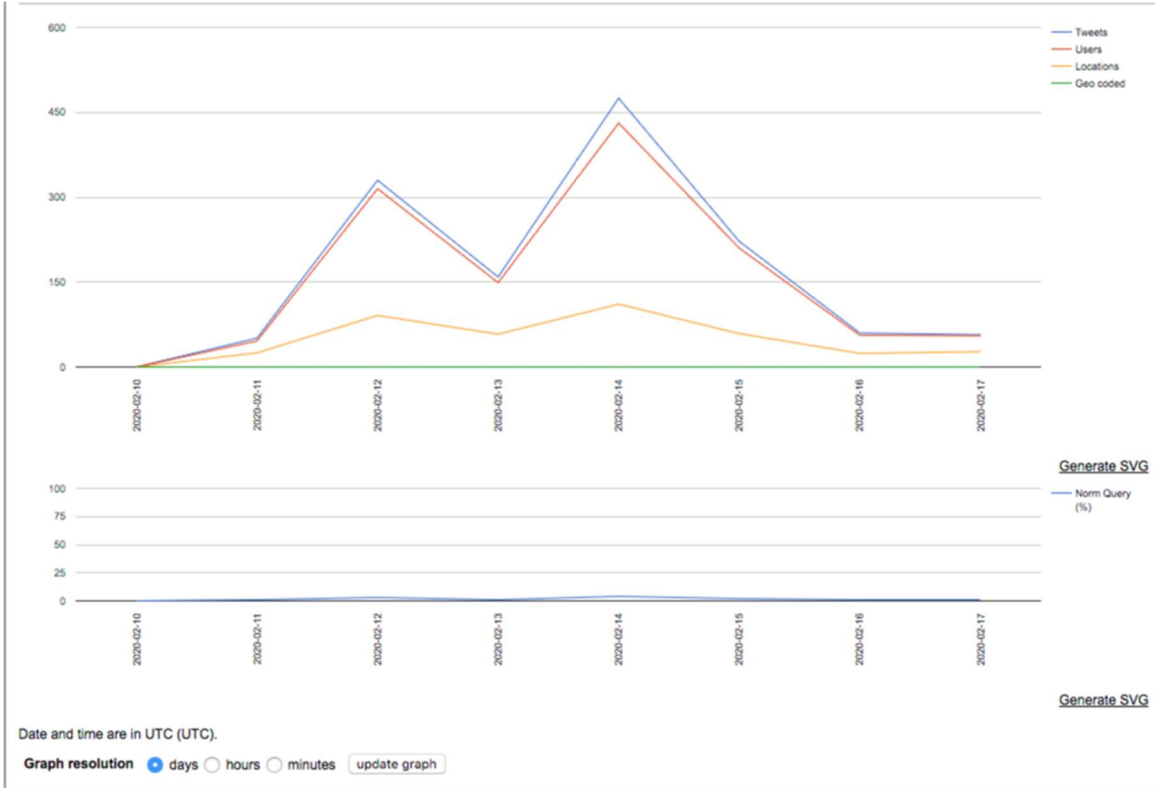
**Select parameters:**

Query:	<input type="text" value="colombia"/>	(empty: containing any text*)
Exclude:	<input type="text" value="cuba"/>	(empty: exclude nothing*)
From user:	<input type="text"/>	(empty: from any user*)
Exclude user:	<input type="text"/>	(empty: exclude no users*)
User bio:	<input type="text"/>	(empty: anything in biography*)
User language:	<input type="text"/>	(empty: any language*)
Tweet language:	<input type="text"/>	(empty: any language*)
Twitter client URL/descr:	<input type="text"/>	(empty: from any client*)
(Part of) URL:	<input type="text"/>	(empty: any or all URLs*)
(Part of) media URL:	<input type="text"/>	(empty: any or all media URLs*)
GEO bounding polygon:	<input type="text"/>	(POLYGON in <a href="#">WKT</a> format.)
Startdate (UTC):	<input type="text" value="2020-02-10"/>	(YYYY-MM-DD or YYYY-MM-DD HH:MM:SS)
Enddate (UTC):	<input type="text" value="2020-02-17"/>	(YYYY-MM-DD or YYYY-MM-DD HH:MM:SS)

\* You can also do AND or OR queries, although you cannot mix AND and OR in the same query.



Grafica 1a - resultados de TCAT



Grafica 1b - resultados de TCAT

### 3.1 Estadísticas:

El conjunto de estadísticas de datos obtenidos se muestra en la siguiente gráfica:

A	B	C	D	E	F	G	H
Date	Number of tweets	Number of tweets with links	Number of tweets with hashtags	Number of tweets with mentions	Number of tweets with media uploads	Number of retweets	Number of replies
2020-02-11	51	5	17	38	7	33	6
2020-02-12	330	17	54	313	11	183	111
2020-02-13	159	10	24	143	8	100	31
2020-02-14	475	21	8	452	9	406	29
2020-02-15	222	2	12	204	6	173	25
2020-02-16	60	3	2	54	2	39	13
2020-02-17	57	4	8	45	5	29	11

Grafica 2 - estadísticas

Las estadísticas muestran, lo siguiente:

Los 1354 tweets se concentran en la mitad del período, adicionalmente el número de re-tweets es de 963 que equivale al 71%.

El número de tweets con links, hashtags, o menciones no es considerable en la muestra y no llega al 1%.

El número de replies es de 226 que corresponde al 16% de la muestra.

Por el contrario, el número de tweets con menciones es de 1249 que corresponde al 92%.

Se extraen como datos para análisis:

- Export de tweets
- Hashtags
- Id de usuario
- Número de twtets
- Export de la red para ser llevada a Gephi

Con esta base se obtiene lo siguiente datos, que se ven reflejados en las gráficas de BATABASE.IO, y las tablas correspondientes.

## 4 ANALISIS

De acuerdo con el planteamiento de los datos recolectados se puede generar algún análisis.

Realizando alguna mirada sobre los tweets producto de esta búsqueda, se observa que la mayoría se refieren a situaciones de orden público que se presentó durante la semana de la investigación. Por lo tanto, los resultados son de alguna manera inesperados frente a lo que se pensaba.

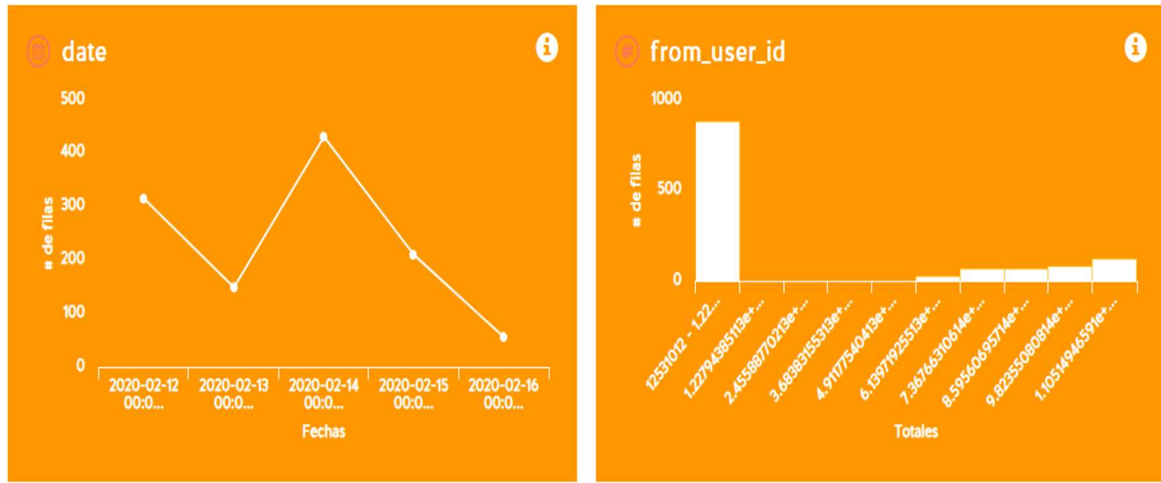
### 4.1 Análisis de Usuarios:

En la Gráfica 3a – Fechas de recolección – feb 10 – feb 16, User-id – usuarios que realizaron tweets

1248	2020-02-11	168228785	migo130680			11940	1053	4993	4		0	3	3	3	0	27	3									
1249	2020-02-11	9773436640022	godagelato			1061	38	1103	1		0	3	0	0	3	60	3									
1250	2020-02-12	9773436640022	godagelato			1064	38	1102	1		0	3	0	0	1	60	3									
1251	2020-02-13	1613497159	FAguerrero1976			2032	50	166	1		0	3	0	4	4	0	0									
1252	2020-02-14	283318070	sergioaraujoc			36984	23496	5715	101		0	3	1	5	248	0	0									
1253	2020-02-14	623185438	cb_inmobiliaria			56125	1168	3521	23		0	3	3	4	0	1	1									
1254	2020-02-14	1944186218	florisanto			28213	791	1229	4		0	3	3	3	0	0	0									
1255	2020-02-14	7965226066692	CceeuAntioquia			16406	974	2996	4		0	3	3	3	0	0	0									
1256	2020-02-14	8912677438650	CarlosC77161692			13715	41	85	0		0	3	3	3	0	0	0									
1257	2020-02-14	9761099349775	JuanMan57493258			55766	645	634	0		0	3	3	6	0	0	0									
1258	2020-02-15	9773436640022	godagelato			1071	38	1103	1		0	3	0	0	1	60	3									
1259	2020-02-16	189165807	Contagioradio1			196754	34661	1726	357		0	3	0	0	17	0	0									
1260	2020-02-17	9773436640022	godagelato			1077	40	1102	1		0	3	0	0	1	60	3									
1261	2020-02-14	189165807	Contagioradio1			196648	34634	1725	357		0	4	0	0	64	0	0									
1262	2020-02-13	4852078305	jolopa310			238	6	50	0		0	5	0	9	0	0	0									
1263	date	from_user_id	from_user_name	from_user_lang	from_user_tweet	from_user_follow	from_user_friend	from_user_listed	from_user_retweet	from_user_verified	in_data	se	retweets	by	user	mentioning	mentioned	total	nr	of	hashta	nr	of	tweets	with	hashtags
1264																										

Gráfica 3a -usuarios y tweets





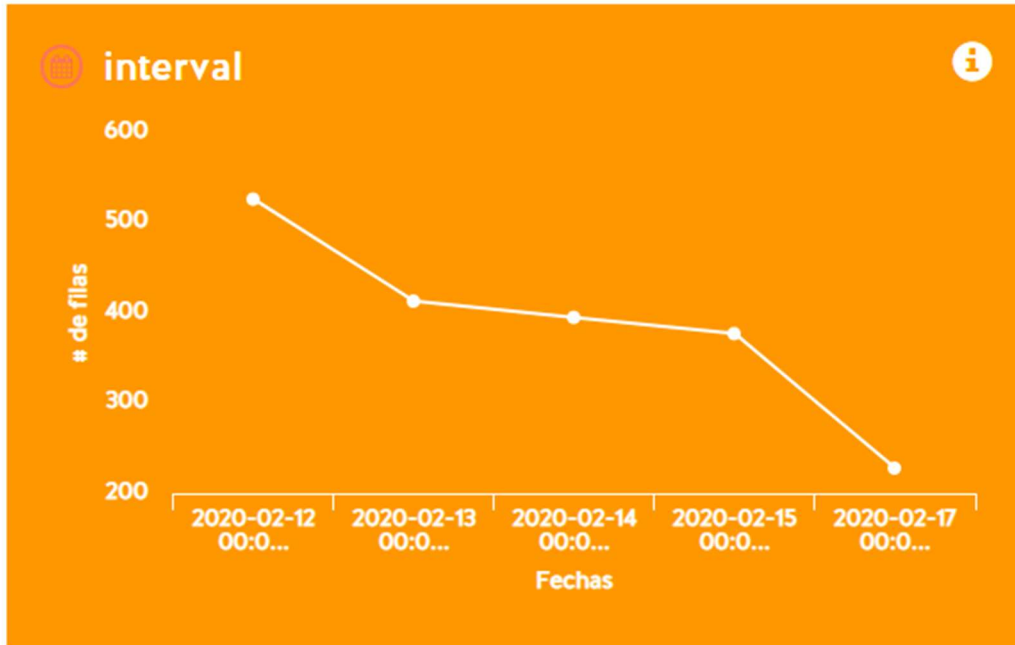
Gráfica 3b – Fechas de recolección – feb 10 – feb 16, User-id – usuarios que realizaron tweets

Únicamente 13 usuarios realizan 3 tweets, un usuario 4 tweets y un usuario 5 tweets. No se muestra en la tabla, pero 59 usuarios realizan 2 tweets y el resto uno o ninguno. El número de re-tweets de estos usuarios no es representativo, llama la atención el número de menciones del usuario “sergioaraujoc” que llega a 248. Este usuario con 3 tweets reúne una gran cantidad de menciones que se observan posteriormente en los diagramas de red. Dentro del análisis de usuarios cabe resaltar este usuario “sergioaraujoc” pues presenta en el data set,

from-user-tweet count:	36.984
from-user-followers:	23.496
From-user-friend-count:	5.715
From-user-listed:	101
Mentioned	248

#### 4.2 Análisis de Tiempos:

En la Gráfica 4 – análisis de intervalo de tiempo, se muestra el comportamiento del número de registros por fechas de la muestra, y se ve una clara disminución a partir del 13 de febrero que puede deberse a la misma situación de orden, ya mencionada. Esta misma situación se aprecia en la gráfica de estadísticas de este data set.



Gráfica 4 – análisis del intervalo de tiempo

### 4.3 Análisis de Frecuencia de tweets:

En la Gráfica 5, se muestra que el número de tweets como ya se mencionó, el máximo es 5, teniendo una tendencia el 1 y el 0, sobresale el usuario “sergioaraujoc” quien con 3 tweets reúne una gran cantidad de menciones que se observan posteriormente en los diagramas de red. Los usuarios que se pueden identificar de universidades, en este data set tienen pocos tweets y adicionalmente no aparecen todas.

1262 rows

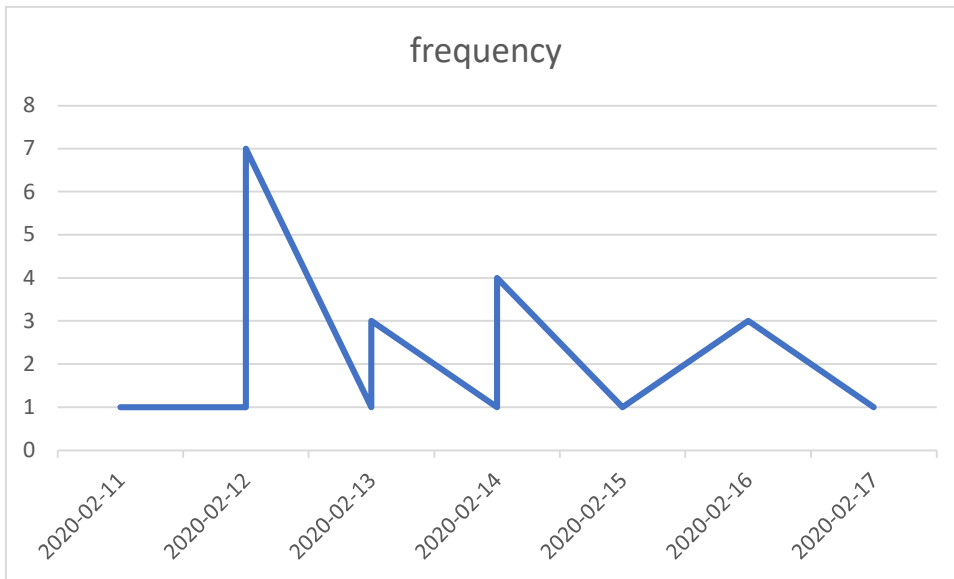
Show as: rows records Show: 5 10 25 50 rows Sort

	id	date	from_user_id	from_user_name	from_user_lang	from_user_tweet	from_user_follow	from_user_friend	from_user_listed	from_user_retweeted	from_user_verified	tweets in data set	retweet
441	2020-02-13	1613497159	FAguerrero1976		2032	50	166	1		0	3	0	
616	2020-02-14	283318070	sergioaraujoc		36984	23496	5715	101		0	3	1	
691	2020-02-14	623185438	cb_inmobiliaria		56125	1166	3521	23		0	3	3	
753	2020-02-14	1944186218	florianlo		28213	791	1229	4		0	3	3	
843	2020-02-14	796522606689209604	CceeuAntioquia		16406	974	2996	4		0	3	3	
861	2020-02-14	891267743865032704	CarlosC77161692		13715	41	85	0		0	3	3	
871	2020-02-14	976109934977921026	JuanMan57493258		55786	645	634	0		0	3	3	
1121	2020-02-15	977343964002236416	goclagelato		1071	38	1103	1		0	3	0	
1165	2020-02-16	189165807	Contagioradio1		196754	34661	1726	357		0	3	0	
1257	2020-02-17	977343964002236416	goclagelato		1077	40	1102	1		0	3	0	
575	2020-02-14	189165807	Contagioradio1		196648	34634	1725	357		0	4	0	
465	2020-02-13	4852078305	jolopa310		238	6	50	0		0	5	0	

Gráfica 5 – análisis de numero de tweets

#### 4.4 Análisis de Frecuencia de “url’s”:

En la Gráfica 6a, se presenta la frecuencia de “url”, la cual oscila entre 1 y 7 y estos se refieren básicamente a los “url’s” de la Gráfica 6b, los cuáles muestran claramente que se refieren a noticias de orden público, incluso uno de estos url se refiere a un tema de universidades latinoamericanas pero referente a un tema de denuncia judicial. El “url” de mayor frecuencia – 7, se refiere a laguna situación del esmad, que ayuda a concluir la situación de orden.



Gráfica 6a – frecuencia de url’s

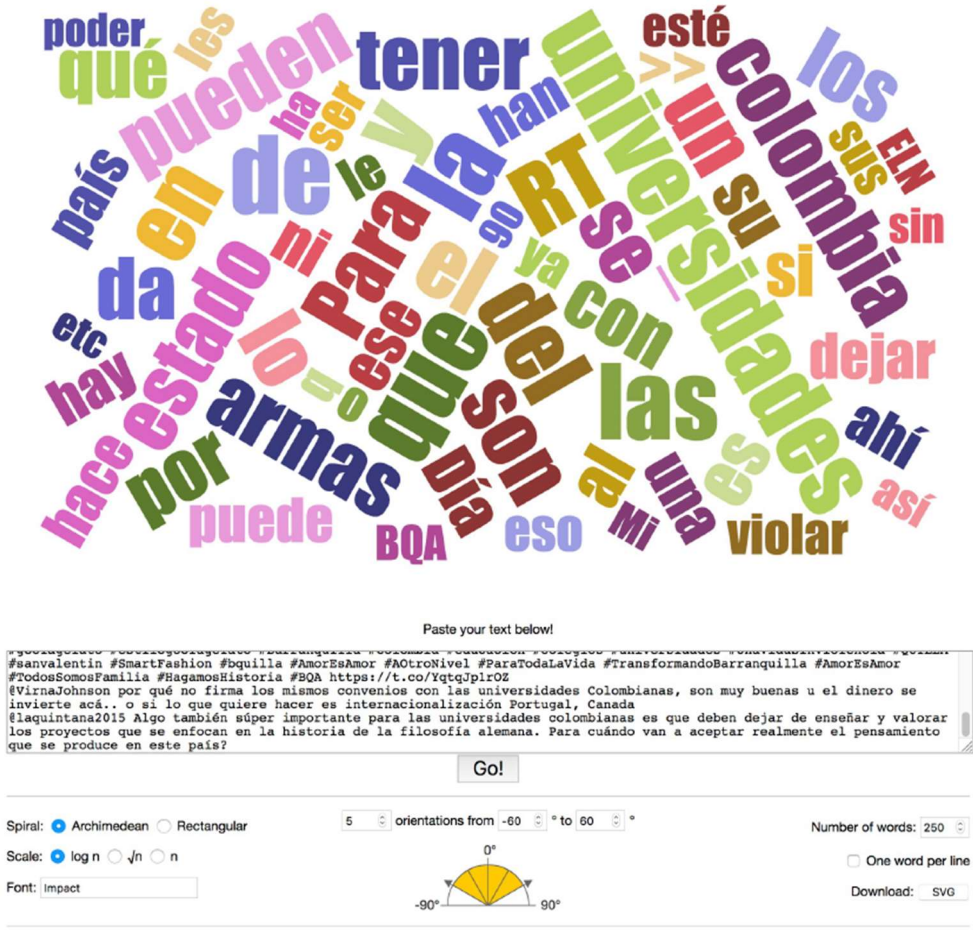
2020-02-13	3	<a href="https://www.elcolombiano.com/antioquia/seguridad/la-alcaldia-de-medellin-estreno-protocolo-de-reaccion-ante-explosivos-EC12455357">https://www.elcolombiano.com/antioquia/seguridad/la-alcaldia-de-medellin-estreno-protocolo-de-reaccion-ante-explosivos-EC12455357</a>
2020-02-14	3	<a href="http://www.resumenlatinoamericano.org/2020/02/13/colombia-docentes-de-universidades-publicas-denuncian-montaje-judicial/">http://www.resumenlatinoamericano.org/2020/02/13/colombia-docentes-de-universidades-publicas-denuncian-montaje-judicial/</a>
2020-02-16	3	<a href="http://contagioradio.com/docentes-de-universidades-publicas-denuncian-montaje-judicial/">http://contagioradio.com/docentes-de-universidades-publicas-denuncian-montaje-judicial/</a>
2020-02-12	4	<a href="https://twitter.com/Col_Informa/status/1227267136256696322">https://twitter.com/Col_Informa/status/1227267136256696322</a>
2020-02-14	4	<a href="http://contagioradio.com/docentes-de-universidades-publicas-denuncian-montaje-judicial/">http://contagioradio.com/docentes-de-universidades-publicas-denuncian-montaje-judicial/</a>
2020-02-12	7	<a href="https://www.elcolombiano.com/antioquia/seguridad/policia-y-esmad-a-las-universidades-polemica-propuesta-de-alcald-de-daniel-quintero-r">https://www.elcolombiano.com/antioquia/seguridad/policia-y-esmad-a-las-universidades-polemica-propuesta-de-alcald-de-daniel-quintero-r</a>
date	frequency	tweetedurl
URL frequency for unividadesmisionextension with search colombia from 2020-02-10 to 2020-02-17		

Gráfica 6b – url’s más mencionados

4.5 Análisis de tweets: frecuencia y conteo de palabras

Para el análisis de los tweets se tiene una vista de frecuencia de palabras donde se puede analizar algo el tipo de lenguaje utilizado

En la Grafica 7a, podemos observar la frecuencia de palabras en el data set base de la recolección, en su representación gráfica por tamaño de ocurrencia de palabras. En una mirada general resaltan palabras como “universidades”, “Colombia”, “estado”, “pueden”, y algunos conectores como “para” “que” y verbos como “son” “pueden” “dejar”, que tendrían algún valor para mi objetivo, pero sus complementos son palabras como “armas” “violar” “eln” lo cual nos lleva a temas de orden público.



Gráfica 7a – conteo palabras y su representación

En la Gráfica 7b de DataBasic, se aprecia el correspondiente conteo de palabras como “universidades” – 1428, “Colombia” – 450, “policía Colombia” – 439, “estado” – 434, los bigrams y trigrams de esta gráfica nos muestran resultados como “la autonomía

universitaria”, “envuelven la policía Colombia”, “la policía Colombia”, que reflejan situaciones de orden público.



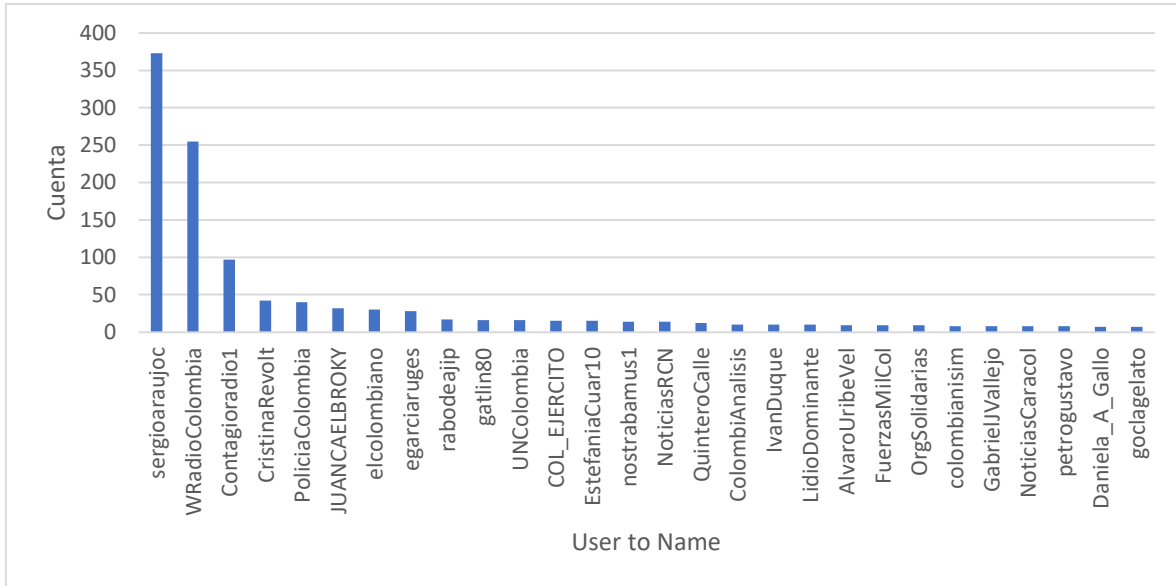
Gráfica 7b – conteo palabras y su representación

#### 4.6 Análisis de tweets, menciones y re-tweets:

Utilizando la estadística de menciones, se tienen 590 menciones y 969 re-tweets. En esta vista se tienen 373 te-tweets del usuario “sergioaraujoc”, el usuario “contagioradio 1” tiene 98 re-tweets. El usuario “WRadioColombia con 255 retweets.

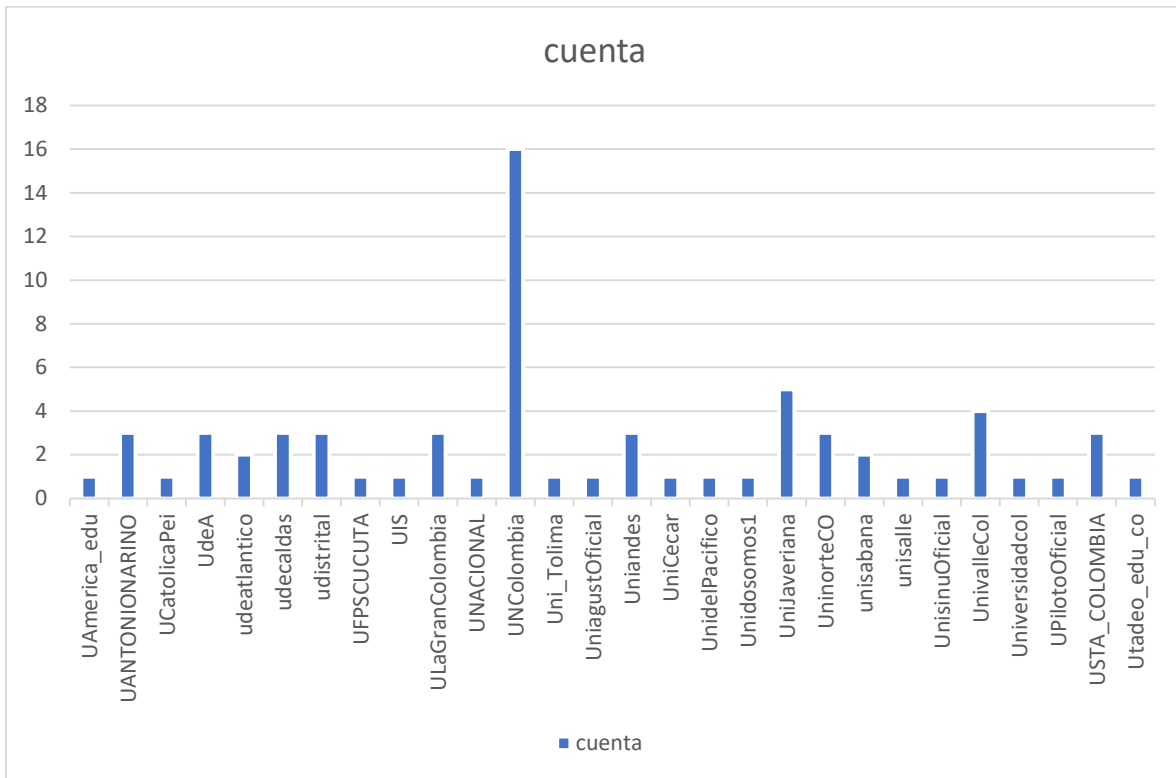
Algunas menciones de 40 para “policía nacional”, “el colombiano” con 30 y aparece un usuario “egarciarujes” con 28 re-tweets y el usuario “CristinaRevolt” con 42. Estos son los valores más representativos lo cual lleva a una dispersión aún mayor de la red.

En la Gráfica 8a , se pueden apreciar los usuarios más representativos del data set, por número de menciones y/o re-tweets. Se puede observar la concentración en unos pocos usuarios.



Gráfica 8a – usuarios más representativos

Los usuarios de las universidades también tienen su presencia en este data set, 68 menciones.



Gráfica 8b – usuarios de universidades

Sin embargo, su representación es muy baja frente a otros usuarios, teniendo en esta gráfica anterior, Gráfica 8b, todos los usuarios de universidades. Sin embargo, no hay tweets de estos usuarios

#### 4.7 Análisis de contenido de algunos tweets:

122829251 681604813 4	15816 82731	2020- 02-14 12:18:5 1	sergioa raujoc	El cuento de que la autonomía universitaria da para que universidades sean un bastión y guarida del terrorismo, que me lo envuelvan. La @PoliciaColombia y las @FuerzasMilCol no pueden tener territorios vedados. Son las armas de la democracia. La capacidad coercitiva del estado
-----------------------------	----------------	--------------------------------	-------------------	---

122843145 239368090 1	15817 15856	2020- 02-14 21:30:5 6	sergioa raujoc	Ese cuento de que la autonomía universitaria da para que las universidades sean bastión y guarida del terrorismo, que me lo envuelvan. La @PoliciaColombia y las @FuerzasMilCol no pueden tener territorios vedados. Son las armas de la democracia. La capacidad coercitiva del estado
-----------------------------	----------------	--------------------------------	-------------------	---

Estos dos tweets del usuario “sergioaraujoc”, son los que concentran las RT- 373, indicadas anteriormente y por lo tanto concentran la red. Claramente se refiere a algún comentario que genera situaciones de respuesta.

122920611 987426918 6	15819 00551	2020- 02-17 0:49:11	RivasR egalOn e	Que buena Noticia, las mayoría de Universidades Colombianas inaugurarán sus Regionales en Florencia, en Puerto Carreño, en Puerto Inírida, en Mocoa, en Mitú, en Yopal y en San José del Guaviare. Bien.
-----------------------------	----------------	---------------------------	-----------------------	--

122796720 689143398 4	15816 05171	2020- 02-13 14:46:1 1	Rankia Colomb ia	Mejores Universidades Virtuales en Colombia <a href="https://t.co/jnO0h5BeMz">https://t.co/jnO0h5BeMz</a>
-----------------------------	----------------	--------------------------------	------------------------	--

122802732 807847116 9	15816 19505	2020- 02-13 18:45:0 5	foris	Experiencias de Universidades en Colombia con respecto al éxito estudiantil. <a href="https://t.co/QHkQuLs8qu">https://t.co/QHkQuLs8qu</a> #Colombia #StudentSuccess #Retención #Foris
-----------------------------	----------------	--------------------------------	-------	--

Estos tres tweets mencionan situaciones importantes de las universidades en Colombia, pero las menciones y seguidores o RT son mínimos, y no se ven en las redes.

122880307 680453427 2	15818 04458	2020- 02-15 22:07:3 8	Daniela _A_Gall o	El caso de TODAS las universidades de Colombia, si quienes estudiamos somos la minoría y no tenemos acceso tan evidente y sencillo a atención mental, ¿cómo será para el resto de los jóvenes?
-----------------------------	----------------	--------------------------------	-------------------------	--

Este tweet puede ser una solicitud a las universidades por una necesidad de atención mental, lo mismo son tweets que no se ven representados en la red.

122820943 394927001 8	15816 62922	2020- 02-14 6:48:42	rhonori c	Colombia. Docentes de Universidades Públicas denuncian montaje judicial - Resumen Latinoamericano <a href="https://t.co/Sgu1JHwixf">https://t.co/Sgu1JHwixf</a>
-----------------------------	----------------	---------------------------	--------------	--

122808707 908467916 9	15816 33751	2020- 02-13 22:42:3 1	Contagi oradio1	Docentes de universidades públicas denunciaron un montaje judicial que se estaría creando en su contra por promover pensamiento crítico en Colombia <a href="https://t.co/eNTWpb0WTo">https://t.co/eNTWpb0WTo</a> <a href="https://t.co/vOgHF6J1m">https://t.co/vOgHF6J1m</a>
-----------------------------	----------------	--------------------------------	--------------------	---

122830413 125471027 2	15816 85500	2020- 02-14 13:05:0 0	Contagi oradio1	"No vamos a dejar que acallen al pensamiento crítico en Colombia" docentes de universidades públicas denunciaron que serían víctimas de un montaje judicial <a href="https://t.co/eNTWpbixKW">https://t.co/eNTWpbixKW</a>
-----------------------------	----------------	--------------------------------	--------------------	--

122756588 645167104 0	15815 09489	2020- 02-12 12:11:2 9	carriq ue618 1	¿Por qué es tan polémica la propuesta de enviar fuerza pública a las universidades? <a href="https://t.co/hfLjxrYzBK">https://t.co/hfLjxrYzBK</a> vía @elcolombiano
-----------------------------	----------------	--------------------------------	----------------------	--

122768607 820059033 8	15815 38145	2020- 02-12 20:09:0 5	Daniel_ Vasque zT	En Colombia las universidades públicas se volvieron un santuario para los criminales. La Policía no puede actuar. En Medellín casi impactan un bus con un explosivo <a href="https://t.co/fZvp5QzE9N">https://t.co/fZvp5QzE9N</a>
-----------------------------	----------------	--------------------------------	-------------------------	--



122795261 934112768 0	15816 01693	2020- 02-13 13:48:1 3	JESUS LADEU TH	Quintero estrenó protocolo de reacción ante explosivos en las universidades via @elcolombiano móvil <a href="https://t.co/nd547OLdbO">https://t.co/nd547OLdbO</a>
-----------------------------	----------------	--------------------------------	----------------------	--

Claramente estos tweets, son situaciones de orden público y de quejas hacia las universidades públicas. De estos tweets anteriores, el usuario “contagioradio 1” generó 98 RT con este twwet, algo que representa un poco más en la gráfica.

#### 4.8 Análisis de red:

Para centralidad de la red, e identificación de comunidades/clusters, se utiliza la modularidad como medida de la estructura de la red. Sirve para mirar el agrupamiento de los nodos. Se utilizó una Resolución de 5.0 con una modularidad resultante de 0.769. Gráfica 9.

Sobre los algoritmos utilizados en el análisis de red, Page rank, para la medición de la importancia de cada nodo en la red, se asignó una probabilidad de 1.0 y un Epsilon de 0.001. Gráfica 10.

Sobre el data set se utilizó algoritmo Page-rank, para centralidad de la red, así mismo y para la identificación de comunidades/clusters, se afinó la Modularidad. Para la distribución de la red se utilizó el algoritmo Force atlas 2. Gráfica 13, muestra los parámetros de este algoritmo.

Para el posicionamiento de los nodos con respecto a los otros se utilizó el algoritmo Force atlas 2, se aplica una reducción de la dispersión y una gravedad para acercarlos.

Parámetros utilizados:

En las siguientes graficas se muestran los grafos y los respectivos parámetros asignados en cada unos de los algoritmos descritos anteriormente.

## 5 Modularity Report

---

### 6 Parameters:

Randomize: On

Use edge weights: On

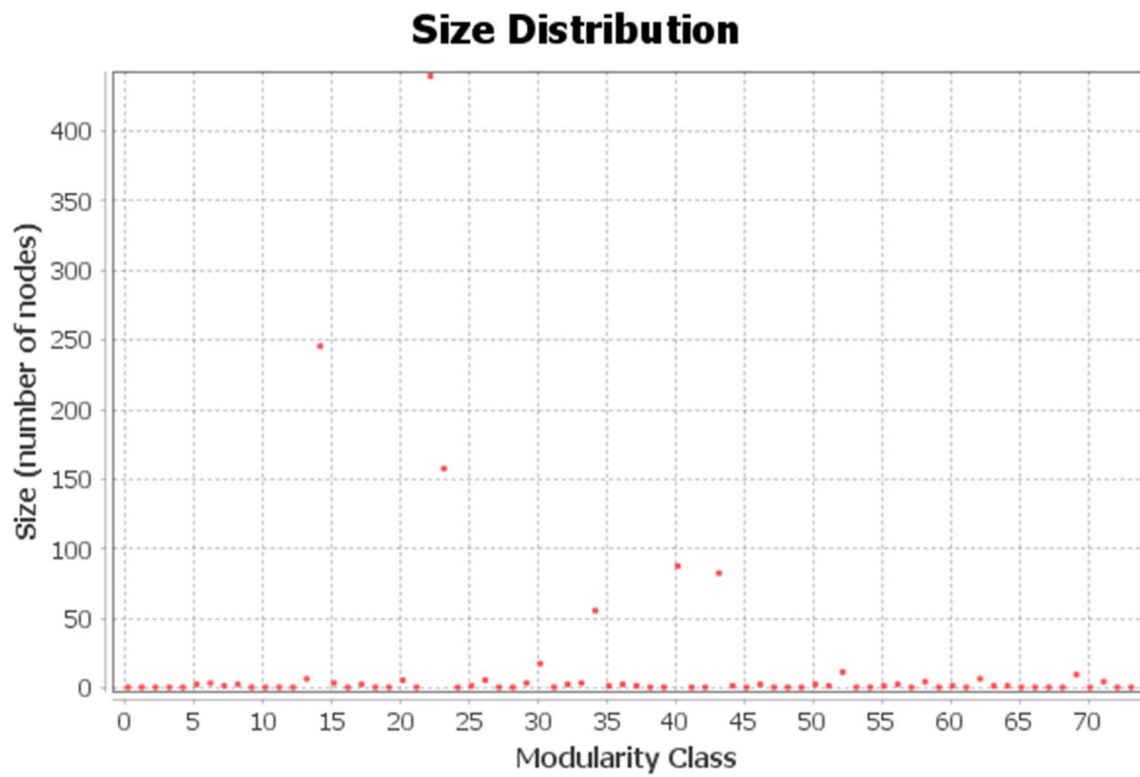
Resolution: 5.0

### 7 Results:

Modularity: 0.769

Modularity with resolution: 4.647

Number of Communities: 74



Gráfica 9 – Modularidad

## 8 PageRank Report

---

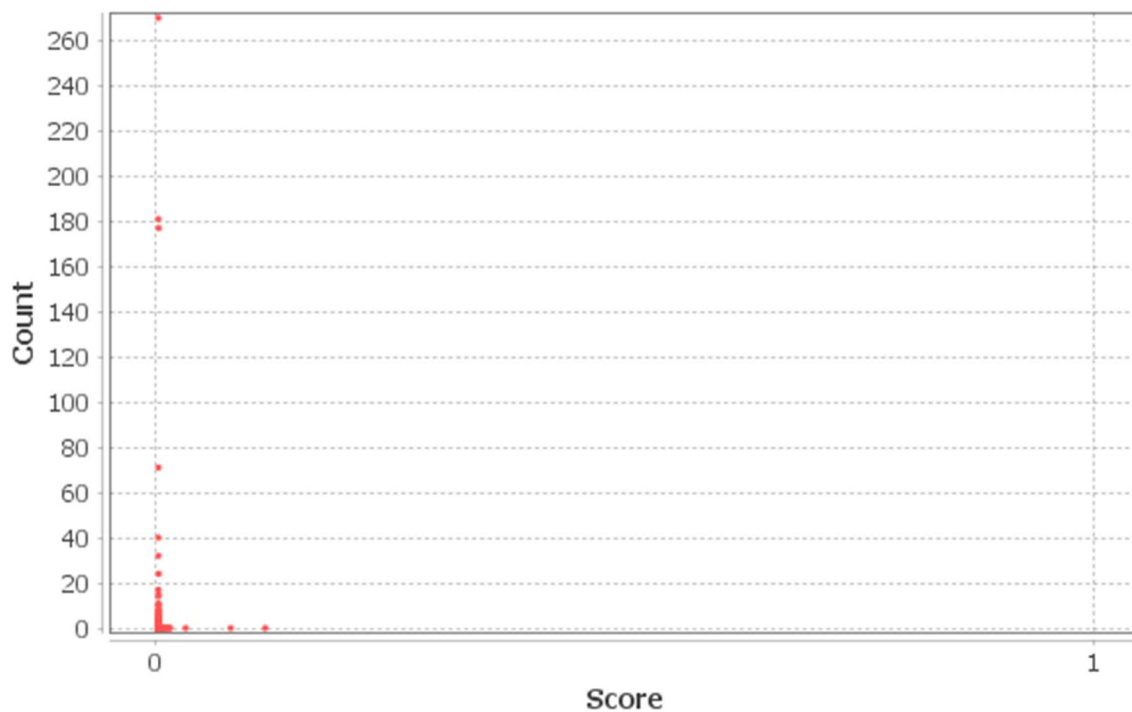
### 9 Parameters:

Epsilon = 0.001

Probability = 1.0

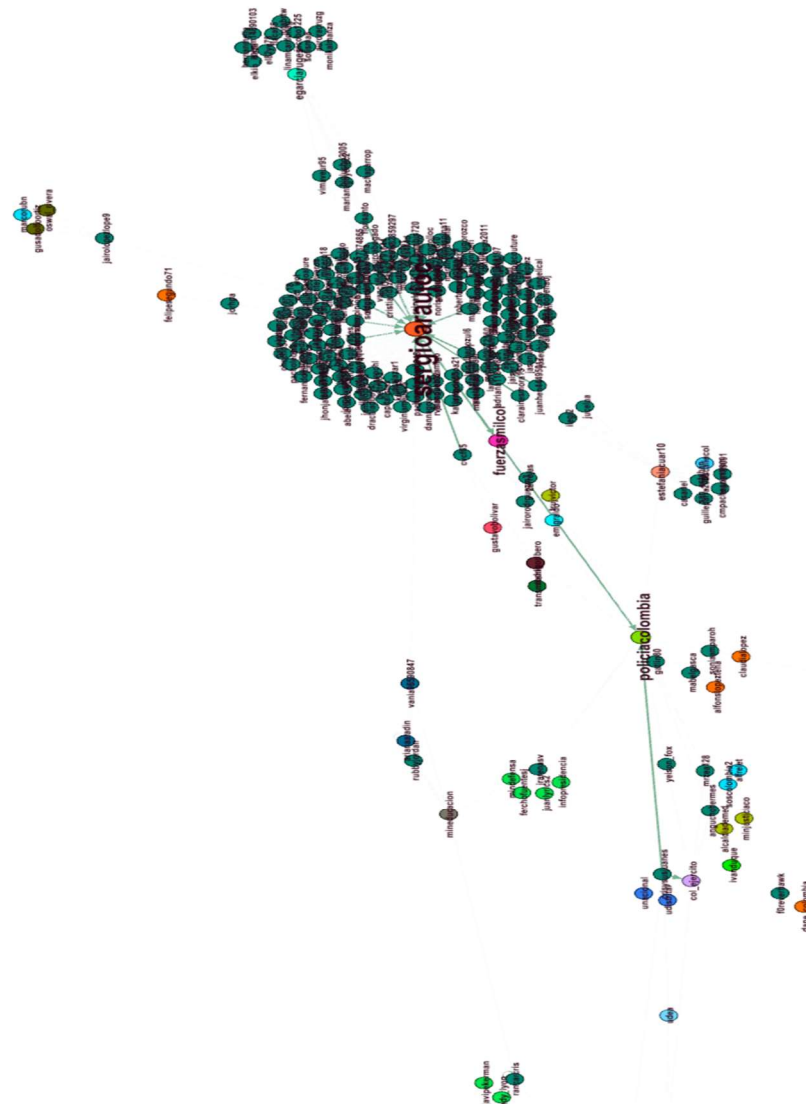
### 10 Results:

### PageRank Distribution



Gráfica 10 – Page rank

Con base en algunas gráficas de red en Gephi, que permiten generar análisis claramente la red es bastante dispersa, en su mayoría. Se presenta una concentración importante la cual se observa en la Gráfica 11, alrededor del usuario “sergioaraujoc”, quien tiene registrados 3 tweets y a partir de estos genera esta gran cantidad de menciones.

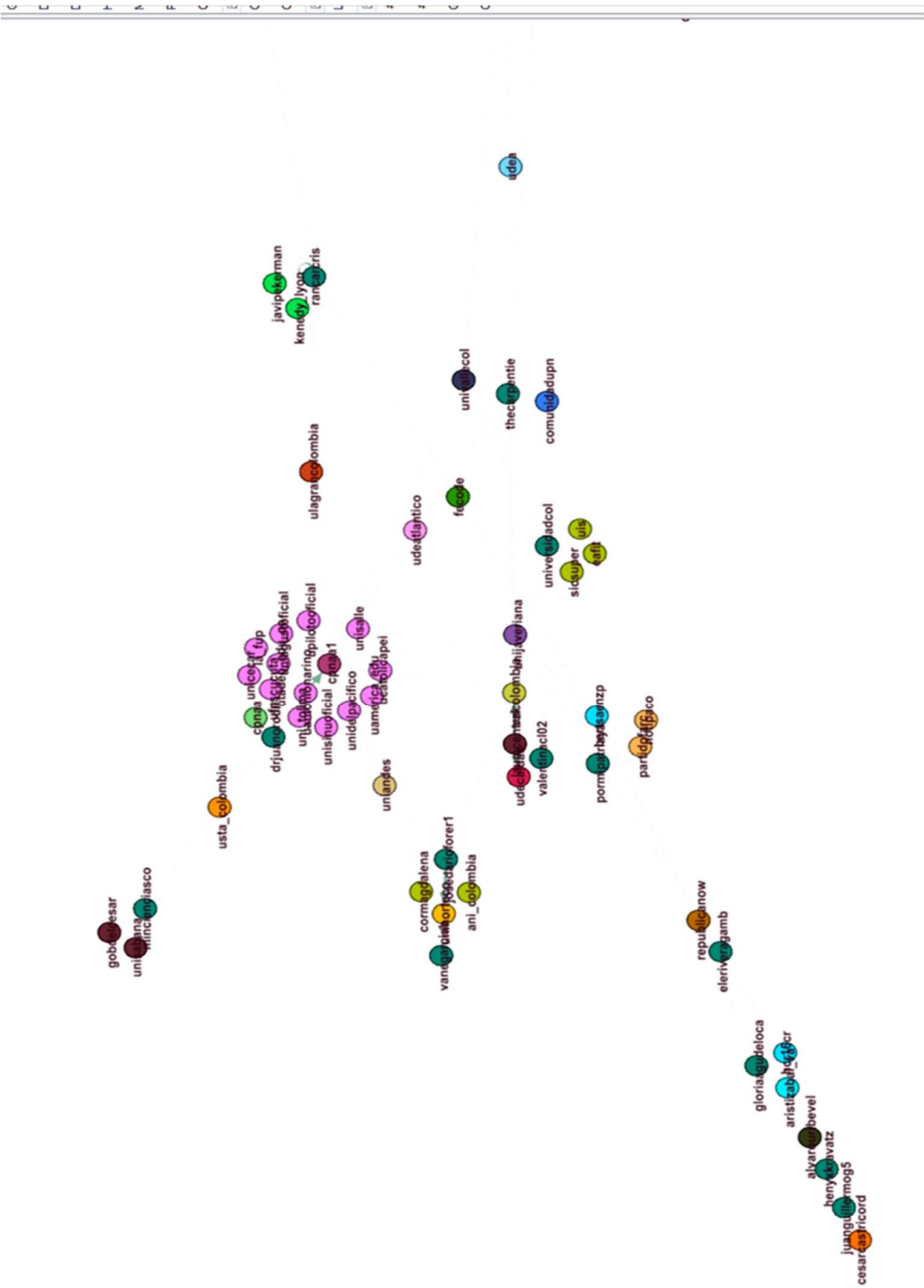


Gráfica 11 - Grafica de red teniendo en cuenta los 1354 resultados

En la Gráfica 12a y 12b,, se observa una pequeña concentración (tono fucsia) que se refiere a los usuarios de las universidades que aparecen en la muestra. Se concentran en un punto, pero sigue siendo disperso el resto de los usuarios.

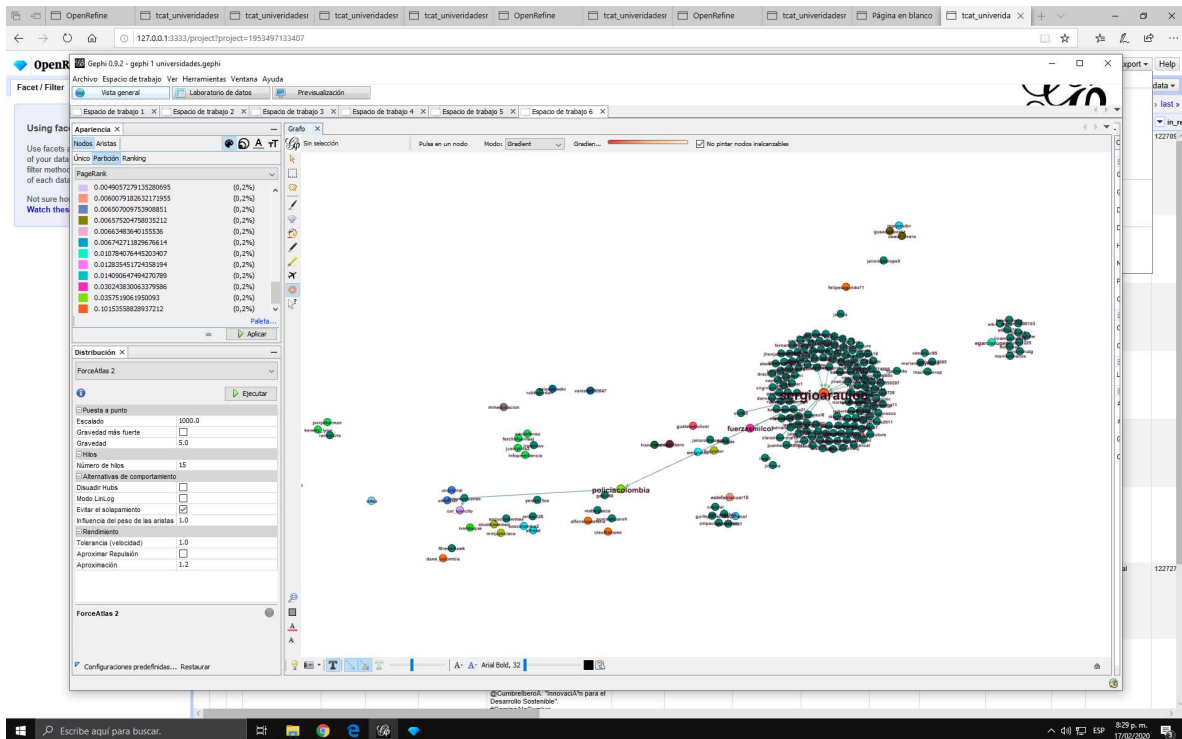


Grafica 12a – muestra de una concentración en fucsia que se refiere a las universidades



Grafica 12b – detalle de la concentración fucsia que son las universidades

La dispersión de la red es más clara en la Gráfica 13



Gráfica 13 – muestra la dispersión de los datos

El análisis general, realmente no aporta a la investigación, puesto que la intención de obtener tweets de las universidades o de lo que se pide a las universidades no se logró con esta muestra. Básicamente se concentró la muestra en algunos tweets generados por una situación de orden público que no era realmente el propósito. Pero se lograron algunos análisis sobre el data set generado.

## 11 CONCLUSIONES

De acuerdo con Rogers (2013), es importante tener claro cómo conseguir la data y como analizar los objetos digitales, hyperlinks, tags, search engine results, archived websites, social networking sites profiles, Wikipedia edits.

Por lo tanto, como conclusión al ejercicio, se deben tener en cuenta varios elementos precisos al momento de tomar las muestras de datos que se van a requerir. Para el caso de recolección de datos en redes sociales es importante tener conocimiento de las formas y esquemas para esta recolección, dado que es el insumo principal para el análisis.

Las investigaciones tienen algunas limitaciones con los datos tomados de redes sociales, puesto estos pueden ser inestables y no siempre ser permanentes; adicionalmente las herramientas pueden restringir los textos, y los datos se toman en determinados intervalos de tiempo que pueden también restringir la investigación. Rogers (2013).

Cómo conseguir la data y como analizarla, y de donde tomarla, son temas importantes a tener en cuenta para estas investigaciones. También hay que considerar que los datos de las redes son datos no estructurados y adicionalmente llevan la interpretación de quien está enviando el mensaje, lo cual significa que son datos cualitativos, e implica un mayor análisis de su contenido para llegar a mejores conclusiones.

Surgen algunas preguntas cuando nos enfrentamos a los medios sociales y a la recolección de datos a través de estas, ¿cómo abordar la investigación en internet?, ¿Cómo enfocar los medios?, ¿Que herramientas usar?, ¿Como definir de acuerdo con la investigación las preguntas anteriores?

De acuerdo con el objetivo planteado en este documento, no se logró recolectar datos que ayuden a revisar la relación de las universidades, con su entorno en todas las funciones que éstas desempeñan.

Para mí trabajo fue importante aprender a recolectar datos de redes sociales que me permitan posteriormente analizar y generar conclusiones.

Pienso que debo investigar más como las universidades se pronuncian a través de Twitter para lograr recolectar la información que realmente me pueda servir. Para esta investigación, sería importante contar con usuarios de las universidades, también a través de los portales y la forma en que las universidades se manifiestan en las redes sociales, en twitter principalmente.

Como un ejercicio siguiente me gustaría crear un “bin” con temas más precisos y búsquedas más alineadas con el propósito de mi investigación. Para este fin he pensado en algunas preguntas que me podrían servir para establecer resultados más afines a mi proyecto.

Las posibles preguntas que tengo serían:

- ¿Cuál ha sido la historia de las relaciones de la comunidad con otros actores a través de twitter? esta comunidad pueden ser las universidades o más precisamente la Javeriana.
- ¿Qué se le pide a la Javeriana a través de twitter?
- ¿Qué dice la Javeriana a través de twitter?
- ¿Con quién y cómo se relaciona la Javeriana a través de twitter?
- ¿Cómo valorar la relación universidad entorno a través de twitter?

Y para finalizar pienso que es importante continuar aprendiendo de estas herramientas robustas que pueden permitirme generar análisis interesantes e importantes para mi investigación.

Fue una excelente experiencia frente a otro tipo de datos y los resultados que se pueden llegar a obtener.



## 12 REFERENCIAS

Hernández Sampieri, R., Fernández Collado, C., Baptista Lucio, P. (2010). (5a Edición). "Capítulo 17: Los métodos mixtos."

Leydesdorff, L. y Etzkowitz, H. (1996). Emergence of a Triple Helix of University- Industry-Government Relations. *Science and Public Policy*, 23(5), 279-286.

Rogers, R. (2013). "Introduction: Situating Digital Methods" y "The End of the Virtual: Digital Methods." *Digital Methods*. Cambridge, MA: The MIT Press. pp 1-38.

Sábato, J. A. y Botana, N. (1968). La ciencia y la tecnología en el desarrollo futuro de América Latina. *Revista de la Integración*, (3), 11.

Walliman, N. (2011). "Ch. 6. The nature of data," "Ch.7. Collecting and analyzing secondary data", "Ch.8. Collecting primary data." *Research methods: The basics*. London: Routledge. pp. 65-127.

### 12.1 Otras Referencias

GUIAS de las herramientas compartidas para este Módulo. TCAT, Gephi, Database.IO