

Proyecto final de Redes complejas con Aplicaciones a Sistemas Biológicos

Redes sociales de difusión de la Noche de los Museos

Alan Givré
Lucía Pedraza
Yuditsabet Burgos

Introducción

El análisis interdisciplinario que permite a través del análisis y procesamiento de información estudiar y comprender distintos fenómenos sociales es fundamental para poder comprenderlos y al mismo tiempo tener herramientas para el desarrollo de políticas públicas.

“Redes sociales, antropología y arte tecnológico en Buenos Aires”, es un proyecto de investigación iniciado por la Universidad del Museo Social en el año 2017. Uno de las interrogantes principales de este proyecto ha sido cómo influyen los museos en el alcance social de la Noche de los Museos a través de las redes sociales; en especial Twitter. Los resultados alcanzados en este mismo año arrojaron poco interés por parte de la población hacia la Noche de los Museos y que los grandes centros culturales de administración estatal, así como algunas instituciones oficiales brindan gran apoyo en la difusión del evento.

A partir de esos resultados se decidió extender el trabajo para el evento “La Noche de los Museos” desarrollado en 2018, teniendo en cuenta la posibilidad de tener un mayor número de participantes, así como una masificación del evento que permite observar mayor interacción en las redes sociales en relación con el evento

Queremos investigar concretamente cómo se relacionaban las personas, las instituciones y los museos con la realización del evento mediante el flujo de información en Twitter, siendo que esta red social permite obtener cierto volumen de información durante determinado tiempo. A partir de los datos obtenidos y la utilización de redes complejas se pudo observar la estructura de la información que circuló durante el evento y los días cercano, para poder aplicar técnicas de detección de comunidades y centralidad para responder nuestra interrogante.

Por último, al trabajar en conjunto con profesionales de otras disciplinas, queda abierta la posibilidad de utilizar nuestros resultados obtenidos a través del flujo en las redes sociales con información obtenida en encuestas realizadas durante el evento, e información oficial sobre la participación en cada museo.

Hipótesis

Por un lado una pregunta inicial es cómo se iniciaba y propagaba la información del evento. Una de las hipótesis que utilizamos es que al ser un evento con actividades en distintas sedes, cada una con una cuenta oficial, estas cuentas jugarían un rol central.

Por otro lado nos interesaba conocer el tipo de comunidades que se formaban en la red. En este caso la hipótesis que sostenemos es que podrían verse expresados clusters según los intereses de los asistentes en los distintos museos, pudiendo confluir varios museos con actividades similares (artísticos, históricos, etc). Ambas hipótesis resultaron equivocadas.

Objetivos

Nuestro interés en realizar este estudio a través de la red de twitter se basó principalmente en la escalabilidad. La intención fue ver si era posible encontrar a través del espacio virtual algunas propiedades de la Noche de los Museos, como por ejemplo, cómo se difunde esta información, y si los interesados en el evento se agrupaban en comunidades por alguna afinidad en particular.

Datos y Metodología

Qué datos y por qué esos datos?

Queremos comprender cómo circula la información con respecto a un evento cultural, y en función de eso cómo se relaciona con la asistencia. Para eso utilizamos información obtenida de la Red Social Twitter. En esta red conviven tanto usuarios que corresponden a instituciones, personalidades de la cultura o la política con usuarios comunes. Por otro lado prácticamente todo el contenido de dicha red es público por lo que podemos acceder a toda la información que circuló con respecto a esta temática.

Para identificar aquellos tweets que nos interesan utilizamos los hashtags “NocheDeLosMuseos” (el oficial) y “LaNocheDeLosMuseos” (que si bien no era el oficial fue utilizado por muchos usuarios) y relevamos los tweets entre las fechas 4/11 y 13/11. Cabe aclarar que es posible que haya habido más tweets referidos a estos temas que no utilizaron ninguno de estos hashtags. Sin embargo, por la dificultad para identificarlos hemos decidido no relevarlos.

Obtuvimos un total de 4299 tweets con el hashtag “NocheDeLosMuseos” y 9606 con “LaNocheDeLosMuseos”. De cada tweet obtenemos la siguiente información (aunque no toda fue posteriormente utilizada): Nombre e Id del autor del tweets, fecha en la que fue publicado, texto y cantidad de retweets.

Para familiarizarnos con la distribución de los datos en el periodo estudiado graficamos la serie temporal de tweets. En la figura 1 se muestra la cantidad de tweets publicados con los hashtags analizados. En promedio se publicaron alrededor de 6 tweets por día con máximos locales concentrados entre las 16 y 20 horas. El máximo global se encontró en el horario en el que se desarrolló el evento (10/21 horas-11/5 horas). Puede observarse los máximos locales más grandes un día antes y un día después de haberse desarrollado la Noche de los Museos.



Figura 1. Evolución temporal de la cantidad de tweets en el periodo analizado

Por otro lado el proyecto original consistía en hacer algunas encuestas en algunos museos elegidos, a usuarios, preguntando sobre la asistencia a ese y otros museos esa noche, y algunas cualidades del asistente como su edad, formación, intereses y canal de información sobre el evento. A su vez, se les preguntaría a cada entrevistado su usuario de twitter para de esta manera poder linkear ambos datos y obtener atributos muy completos de alguno de los nodos del grafo.

Finalmente dichas encuestas no pudieron ser realizadas, por lo que solo utilizamos la información obtenida de Twitter.

API y TWEOPY

Para obtener la información de Twitter utilizamos la Interfaz de Programación de Aplicaciones (API) de dicha red. Accedimos a ella a través de la librería de python "Tweepy". Recopilamos los tweets a partir del día 13/11 en retrospectiva por 10 días (siendo la noche de los museos el 10/11), ya que la API solo permite recopilar información de unos 10 días. Una de las mayores dificultades resultó que la API de Twitter permite descargar sólo cierta cantidad de información cada 15 minutos. De esta manera, el tiempo limitante en los cálculos realizados resultó no ser el tiempo computacional, sino este factor de regulación por parte de Twitter. Los datos descargados fueron guardados en un archivo con formato json ya que networkx es muy compatible con este formato, y permite armar los grafos muy rápidamente.

Construcción de la red

Con la información de retweets relevados construimos una red. En dicha red cada usuario interactuando representa un nodo, y cada retweet es una arista (dirigida hacia la persona que es retweeteada). Esta red tenía una Componente Gigante donde se concentraba el 87% de los nodos y donde además se ubicaron el 95% de los museos en cuestión (figura 2).

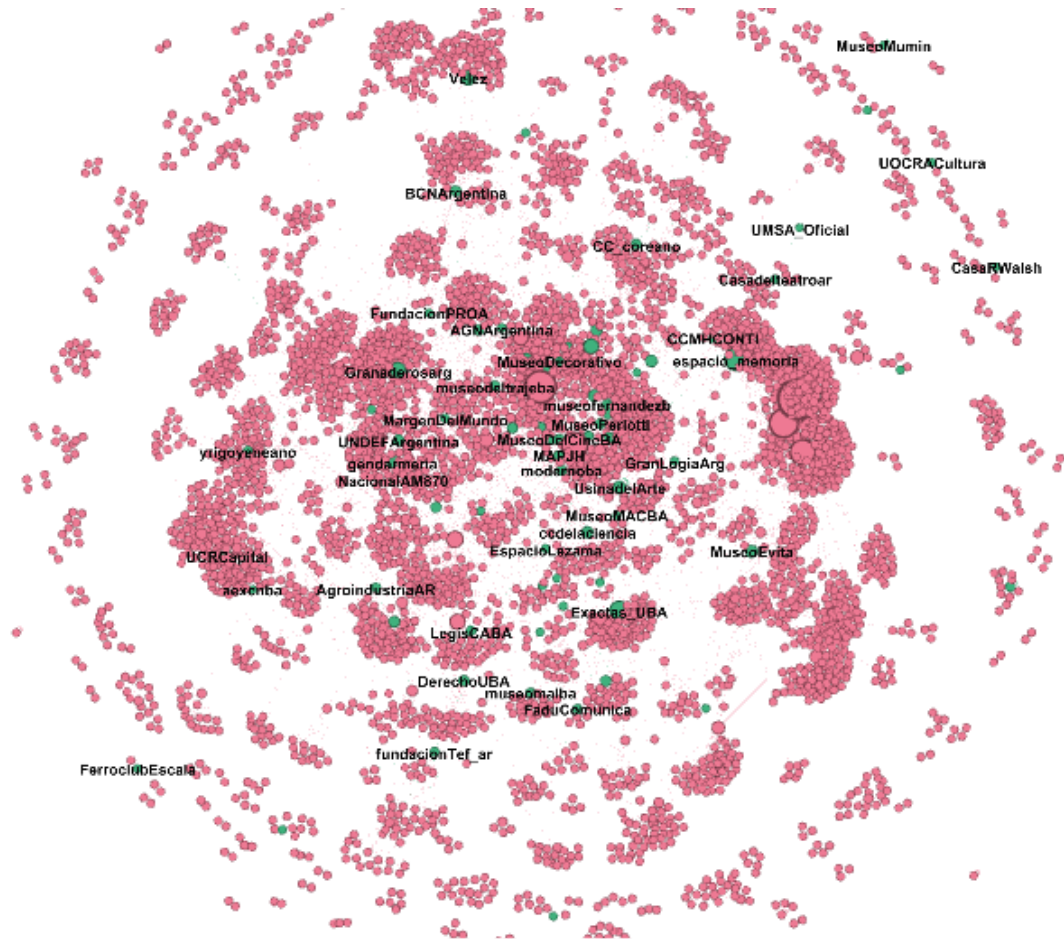


Figura 2. Red de retweets. Círculos rosados son usuarios, círculos verdes son museos.

Resultados

Características de la Red

En la tabla 1 se pueden ver algunas de las principales características que recuperamos de la red dirigida. Según estas características la red resultó ser poco conectada y con una probabilidad baja de encontrar triángulos dentro de la misma. El grado máximo de entrada superó notablemente al de salida, cosa que era esperable para una red dirigida.

La red original tiene 4503 nodos, mientras que la componente gigante posee 3907 nodos.

Tabla 1. Características de la red dirigida

ID	clustering global	densidad	enlaces	kmax out	kmax in	kmean	kmin out	kmin in	nodos
LaNoche delos Museos	0.003699	0.000305	6187	31	210	1.3739	0	0	4503

Analizamos la distribución de grados de entrada (figura 3a) y de salida (figura 3b) de la red, así como en la red general (figura 3c). Se puede observar que todas las distribuciones parecen aproximarse a una ley de potencias, esto indica que la red es libre de escalas y por tanto sigue una autoorganización.

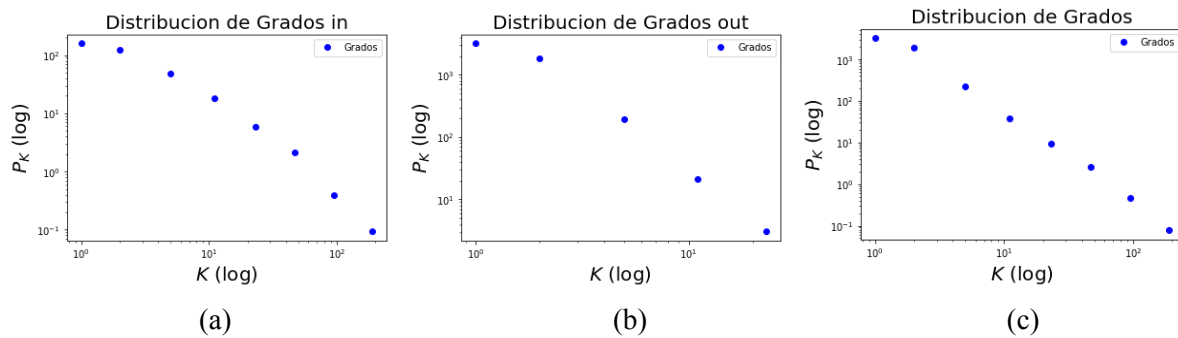
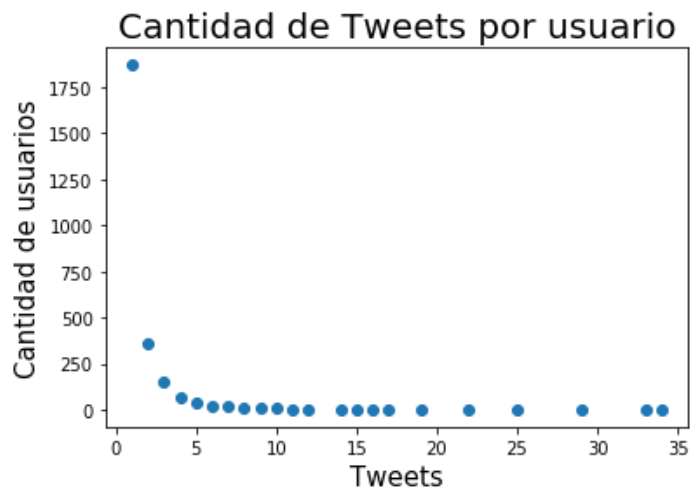


Figura 3. Distribución de grado(k) de la red dirigida en escala logarítmica.

Además, quisimos conocer cantidades aproximadas de tweets por usuarios y por museos. Pudimos observar que en general pocos usuarios realizaron más de 5 tweets y que el promedio solo se hacían entre 3 y 4 tweets. Por otra parte, los museos no tuvieron mucho protagonismo en la realización de tweets, solo un 90% realizó entre 1 y 5 tweets, mientras que unos 3 museos tweetearon más de 10 veces. Figura 4



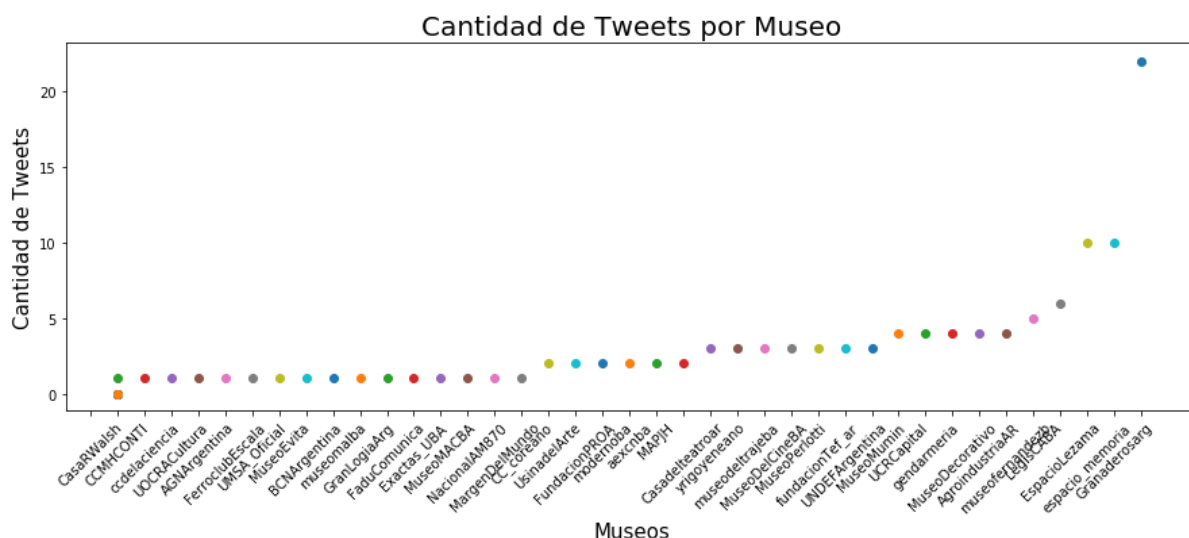


Figura 4. Cantidad de Tweets por museo.

Hubs

En toda red es importante conocer los nodos más conectados, en la nuestra los nodos de mayor grado no fueron precisamente los museos como realmente esperábamos. En general, hubo nodos muy conectados de diferentes naturaleza; algunas cuentas de organismos culturales y de derechos humanos como Espacio, Memoria y DDHH, y Secretaría de Cultura Nación. Otros como el Museo de Gendarmería y el gobierno de la ciudad de Buenos Aires, y otras dos cuentas de personas retweteadoras que tomaron los hashtag que fueron tendencia en ese momento y retwetearon cosas sin contenido relacionado con esos hashtag.

Centralidad de los nodos

Como mencionamos antes los museos en general no fueron los más conectados. Aun así consideramos importante determinar qué tan relevantes resultaban ser en la difusión de información dentro de la red; ya que como hemos visto hasta aquí algunos fueron hubs y retwetearon más de 5 veces. Además, los museos suelen seguirse entre ellos. Para esto aplicamos dos medidas de centralidad que a nuestro criterio se ajustaban a lo que queríamos modelar. Primeramente determinamos intermediariedad, es decir si la información se intercambia por geodésicas. Para esta medida las cuentas con mayor centralidad resultaron ser las mismas que los hubs. Luego, determinamos centralidad por autovector y los resultados fueron muy similares. Por último, nos dimos cuenta que necesitábamos encontrar una medida de centralidad global que privilegiara a los nodos que pertenecieran a zonas centrales de la red, más que a nodos que localmente fueran importantes, pero que globalmente estuvieran marginalizados. Con este criterio elegimos la centralidad PageRank por sobre el Eigenvector, para eliminar situaciones patológicas. Por este método los nodos más centrales no resultaron ser los museos (aunque estos tienen cierta centralidad), sino cuentas institucionales estatales o no estatales que promocionan distintos museos o la Noche de los Museos en general. También había algunos políticos y personas famosas involucradas en los retweets. Las tablas 2-4 muestran detalladamente los nodos más centrales.

Por tanto, todas las medidas de centralidad aplicadas demostraron que 6 museos de 71 fueron claves en la difusión de la información, mientras que 8 cuentas institucionales culturales y sociales se encargaron de difundir más el evento que los propios museos. En la figura 5 se puede observar la posición de los nodos más centrales resaltados en tamaño y color verde.

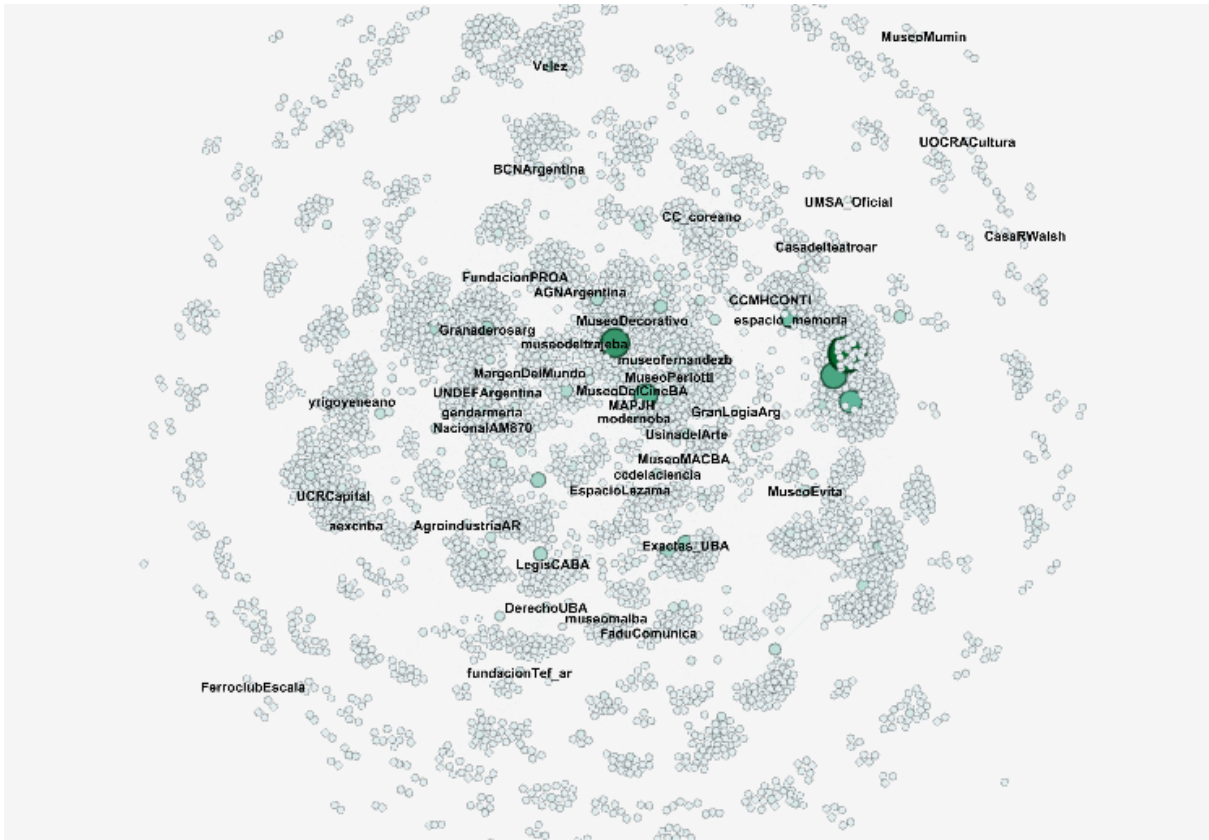


Figura 5. Nodos de la red, con los nodos de mayor centralidad PageRank resaltados en verde.

Tabla 2. Centralidad PageRank. (Verde: institucionales, Turquesa: museos, Amarillo: Famosos/Políticos con cargo/Funcionarios, Rojo: personas (incluye representantes gremiales), Magenta: sólo retwittean, posibles trolls)

Cuentas	PageRank Centrality	Tipo de Cuenta	Cuentas	PageRank Centrality	Tipo de Cuenta
hijos_capital	85	Institucional	sanmihanovich	6	Famoso/Político
La_Imposible	56	Institucional	FlavioBuccino	6	Persona
abuelasdifusion	48	Institucional	jesusucr	6	Famoso/Político
Granaderosarg	21	Museo	LegisCABA	5	Museo
charlypisoni	19	Persona	jrcapital	5	Institucional
anabelpomar	19	Persona	ezequielvera	5	Persona
prensadelsubte	19	Institucional	JorgeFavaloro	4	Persona
FelipePigna	12	Famoso/Político	GusFulco	4	Famoso/Político
Velez	11	Museo	vivicantoni	4	Famoso/Político
gendarmeria	8	Museo	Gracemazzola	4	Persona
cristianucr8	8	Persona	ElFortindeVelez	4	Institucional
DerechoUBA	7	Museo	AlejSanta	4	Famoso/Político

fsuarezlastra	7	Famoso/Político	smcassina	3	Persona
DiputadosAR	7	Institucional	ajolo	3	Persona
intaargentina	6	Institucional	pdpassarini	3	Persona

Tabla 3. Centralidad In Degree (Verde: institucionales, Turquesa: museos, Amarillo: Famosos/Políticos con cargo/Funcionarios, Rojo: personas (incluye representantes gremiales), Magenta: sólo retwittean, posibles trolls)

Cuentas	In Degree	Tipo de Cuenta	Cuentas	In Degree	Tipo de Cuenta
abuelasdifusion	194	Institucional	DiputadosAR	20	Institucional
hijos_capital	190	Institucional	jpcyf13	20	Institucional
Granaderosarg	124	Institucional	MGuouman	20	Famoso/Político
FelipePigna	65	Famoso/Político	smcassina	20	Persona
La_Imposible	63	Institucional	pdpassarini	20	Persona
Velez	61	Museo	eavogadro	19	Famoso/Político
gendarmeria	51	Museo	intaargentina	19	Institucional
fsuarezlastra	46	Famoso/Político	mpvischi	19	Famoso/Político
DerechoUBA	42	Museo	AgroindustriaAR	19	Institucional
jrcapital	33	Institucional	SDHArgentina	19	Institucional
LegisCABA	32	Museo	Comitelarralde	19	Institucional
JorgeFavaloro	26	Persona	Ejercito_Arg	19	Institucional
Gracemazzola	23	Persona	CrisAlvarezRod	18	Famoso/Político
ElFortindeVelez	23	Institucional	Comunicacionucr	18	Museo
jesusucr	20	Famoso/Político	DisfrutemosBA	18	Institucional

Tabla 4. Centralidad Out Degree (Verde: institucionales, Turquesa: museos, Amarillo: Famosos/Políticos con cargo/Funcionarios, Rojo: personas (incluye representantes gremiales), Magenta: sólo retwittean, posibles trolls)

Cuentas	Out Degree	Tipo de Cuenta	Cuentas	Out Degree	Tipo de Cuenta
Comunicacionucr	19	Institucional	MartaSinH83	5	Persona
marcelalarrosa8	12	Posible Troll	AdhilacInfo	4	Institucional
Natbaires	7	Posible Troll	ZulemaBeltrami	4	Persona
horacioPacific	7	Posible Troll	eneroman_nora	4	Persona
Samanta50280978	7	Posible Troll	museofernandezb	4	Museo
CabaStreaming	7	Institucional	pabloavelluto	4	Famoso/Político

mafolini	6	Famoso/Político	NoraKreimer	4	Persona
EvolucionCABA	6	Institucional	LorenaP12106475	4	Posible Troll
EVOLousteau	6	Famoso/Político	ucrsiempre	4	Institucional
Lore__Marquez	6	Persona	graciela_solo	4	Posible Troll
MarianaCaccia	5	Famoso/Político	BeatrizBotana2	4	Persona
eduardorflorio	5	Persona	perfumealibro	4	Posible Troll
martimarilyn	5	Posible Troll	MarceloSoldati	4	Posible Troll
RadioTrendTopic	5	Famoso/Político	hijos_capital	4	Institucional
EmiCagnacci	5	Persona	MarceloLorenzu1	4	Posible Troll

Por otro lado intentamos evaluar la centralidad de los museos rompiendo la red mediante varios métodos, primeramente eliminamos todos los museos (Zotenko et al., 2008), luego al azar y por último los nodos centrales obtenidos por los métodos PageRank y Grado. Obtuvimos que al quitar 71 cuentas al azar, la componente principal se reduce a aproximadamente 3790 nodos, como se puede ver en la figura 6, es decir, disminuye 713 nodos. Sin embargo, al quitar los 71 museos, la LCC se reduce a 3181 nodos, menor que la hipótesis nula con un p -valor < 0.01 . Si bien los museos resultaron tener importancia, al quitar los nodos de mayor PageRank, o los nodos de mayor grado, la red se rompe en forma mucho más extrema, quedando LCC de apenas 1517 y 1481 nodos respectivamente. Ver tabla 5

Dado estos resultado concluimos que las cuentas de los museos son levemente más centrales que una cuenta al azar pero no son los nodos más centrales de la red.

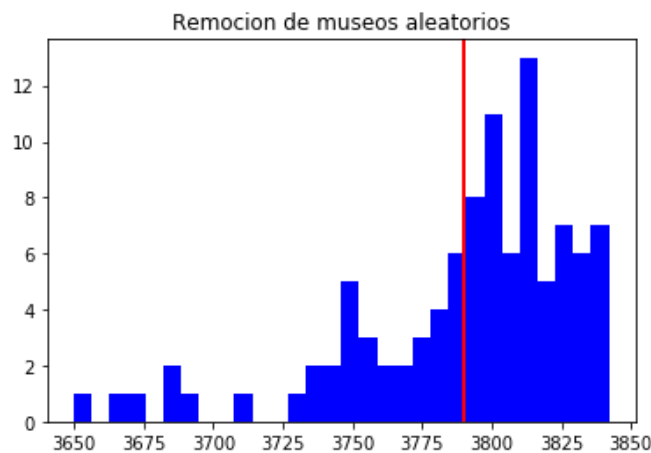


Figura 6. Histograma con 100 realizaciones, donde se muestra el tamaño de la componente mayor conexas al quitar 71 cuentas al azar.

Tabla 5. Número de nodos resultantes ante la remoción de nodos centrales y museos

Ordenado	Aleatorio	PageRank	Grado
3181	3790	1517	1481

Comunidades

Se usó el algoritmo de Louvain en el programa Gephi para distinguir las comunidades de la red, con una resolución de 3.5 que prioriza encontrar comunidades relativamente grandes. Como resultado se encontraron 7 comunidades en la LCC, como se muestra en la figura 7.

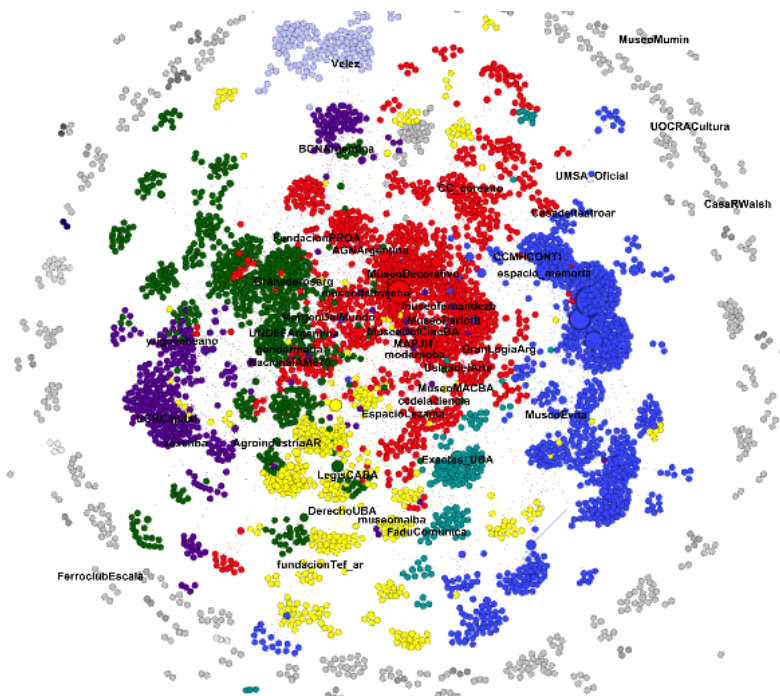


Figura 7. Comunidades encontradas en la Componente Principal resaltadas con distintos colores.

- Cluster central (25,1% de nodos)
- Comunidad kirchnerista y de DDHH (21,4% de nodos)
- Ejército (15,0% de nodos)
- Comunidad Radical (9,5% de nodos)
- Instituciones ligadas a CABA (8,7% de nodos)
- Exactas, FADU, etc (3,4% de nodos)
- Vélez (3,1% de nodos)

El cluster central tiene como cuentas centrales las cuentas institucionales de la Noche de los Museos y el Ministerio de Cultura, así como al secretario de estado de cultura Pablo Avelluto. Además, este cluster tiene una alta concentración de museos, 33 de los 71 que twetearon se encuentran en este cluster. Relacionada a esta comunidad, se encuentra una específicamente de las instituciones de la ciudad de Buenos Aires (como la Legislatura, el gcba, la Magistratura, etc). Otra institución estatal que resulta extremadamente relevante y que formó un cluster es el Ejército, incluyendo museos como el de Granaderos, y cuentas como Gendarmería.

Por demás, aparecen comunidades que se consideran preexistentes como partidos políticos. En particular, el kirchnerismo, que aparece ligada a las instituciones de DDHH-Hijos, Abuelas y a

FEDUBA. Otro cluster lo formó el radicalismo, que incluye ciertos museos radicales como el irigoyenista, y la sociedad de ex alumnos del CNBA y la biblioteca del Congreso. También aparece el equipo de fútbol de Vélez como una comunidad independiente, con muchos hinchas y cuentas ligadas. Por último, como una comunidad chica resultó Exactas y FADU, que incluye el museo informático.

Conclusiones

Las cuentas de los museos juegan un rol importante en la circulación de la información durante la noche de los museos aunque no el central, que lo juegan cuentas institucionales estatales o no estatales.

Al separar en clusters se observa que la red forma un cluster central con la mayoría de museos y las cuentas institucionales, mientras que el resto de los clusters parecería separarse en comunidades preexistentes a las interacciones durante esa noche.

Queda pendiente un análisis de las relaciones entre los usuarios para comprobar esa hipótesis.

Observaciones

Cosas que probamos y fallamos

- Realizar una comparación con encuestas realizadas en la Noche de los Museos.
- Realizar una red estructural de seguidores de los nodos obtenidos.
- Comparar la asistencia a los distintos museos.

Bibliografía

1. Lodi M. E., Díaz D., Dimare G., Arza A., Montini A., Scanio P. (2017): Los espacios conectados: Una indagación sobre las redes, la antropología y el arte tecnológico en Buenos Aires. Proyecto: “Análisis de redes sociales para la investigación en artes y curaduría”. Universidad del Museo Social Argentino.
2. Zotenko E., Mestre J., Przytycka T. (2008): Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality.

