

Redes sociales en línea: macro y microanálisis de páginas de Facebook en una ciudad mediana

Emiliano M. Gutiérrez¹; Juan M. C. Larrosa²

Introducción

El objetivo principal de este trabajo es identificar patrones de estructura en redes sociales en línea (OSN) a nivel de páginas de Facebook, así como la red completa que conforman todas las sub-redes (urbana). Asimismo queremos identificar las principales comunidades y actores. Facebook se ha impuesto durante mucho tiempo como una OSN en todo el mundo con más de mil quinientos millones de usuarios activos para 2018 (Facebook, 2019). Individuos, familias, empresas, causas, instituciones y sitios falsos de eventos conviven en la OSN. La gran popularidad de Facebook lo convirtió en un objetivo deseable para obtener información demográfica y comercial en términos de la composición de los actores que interactúan en línea.

El capital social se inserta en los sitios de Facebook (Brooks et al., 2014) y las personas usan sus sitios personales para interactuar con familiares, amigos, conocidos e incluso parejas desconocidas. La rica diversidad de interacciones hizo de este OSN un laboratorio para probar diversas

1 Instituto de Inv. Económicas y Sociales del Sur (IIESS). emiliano.gutierrez@uns.edu.ar

2 Departamento de Economía, Universidad Nacional del Sur (UNS). jlarrosa@uns.edu.ar

hipótesis. Pueden surgir nuevos mercados que conectan a los usuarios con las empresas de cada sector comercial.

Nuestro trabajo muestra diferentes páginas de seguidores de Facebook de empresas ubicadas geográficamente en un área urbana que comprenden diferentes sectores y organizaciones socioeconómicas. Modelamos cada sector como una red de actores conectados a través de la página de fans. Buscamos comprender cómo cambia la estructura de estas redes en todos los sectores.

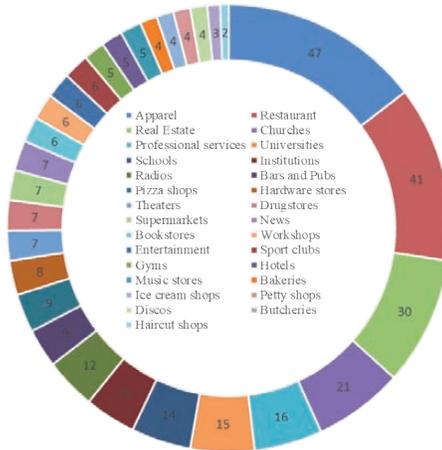
Datos y análisis

Obtuvimos datos de las fanpages de Facebook con casi todos los negocios con una tienda física en la ciudad de Bahía Blanca (Argentina). Descargamos 325 redes que comprenden 94,692 nodos con 10.638.433 enlaces (incluidos bucles). Las redes son páginas de fanáticos (fanpages) de Facebook capturadas a lo largo del año 2015 (del 1/1/2015 al 31/12/2015). Capturamos una parte significativa de diferentes digitales de una ciudad entera con una población total de 335,000 habitantes. Todos los datos son anteriores al escándalo de Cambridge Analytica.

Identificamos los siguientes sectores entre las redes de páginas de fans como impulsores socioeconómicos comunes: restaurantes, panaderías, heladerías, bares y pubs, pizzerías, escuelas, librerías, universidades, instituciones, teatros, gimnasios, hoteles, farmacias, indumentaria, talleres, entretenimiento, clubes deportivos, carnicerías, ferreterías, pequeñas tiendas, supermercados, servicios profesionales, tiendas de música, noticias, radios, inmobiliarias,

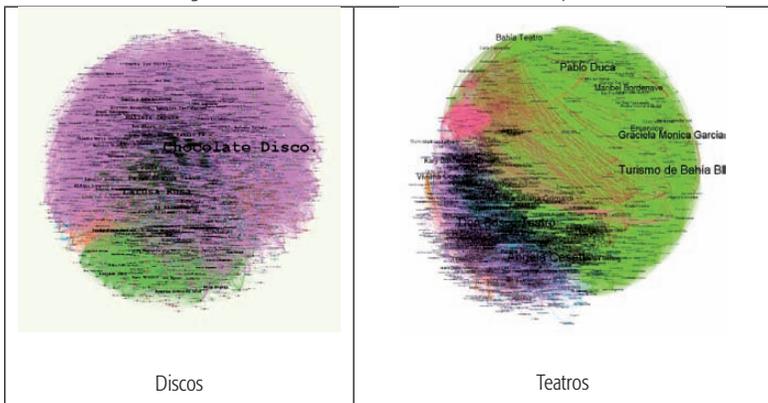
discos, peluquerías e iglesias. De esta manera, cada una de las redes detectadas pertenece a una de estas categorías. La Figura 1 muestra una preeminencia en tiendas de ropa y restaurantes.

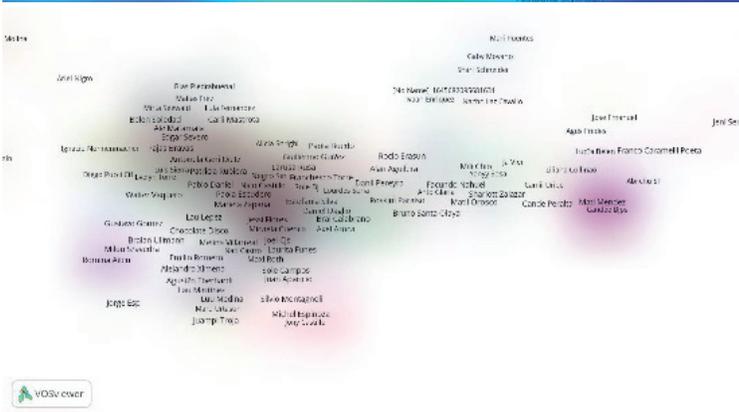
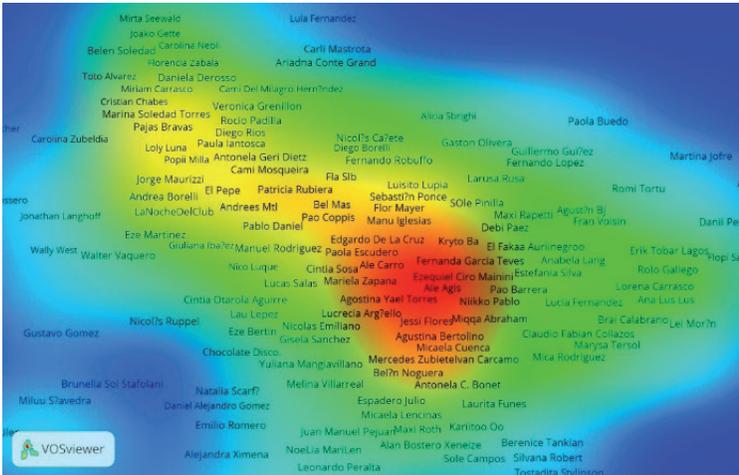
Figura 1. Composición por sector



Una vez procesados, los datos se transformaron en subredes particulares del conjunto. Por ejemplo, la Figura 2 muestra cuatro subredes donde la complejidad surge de inmediato y dificulta la identificación de actores y estructuras.

Figure 2. Visualizaciones de sub-redes de la red princ





Mapa de calor de discos

Si bien es difícil comprender a los actores y las comunidades en particular, podemos dividir el análisis de la siguiente manera.

Macroanálisis y microanálisis

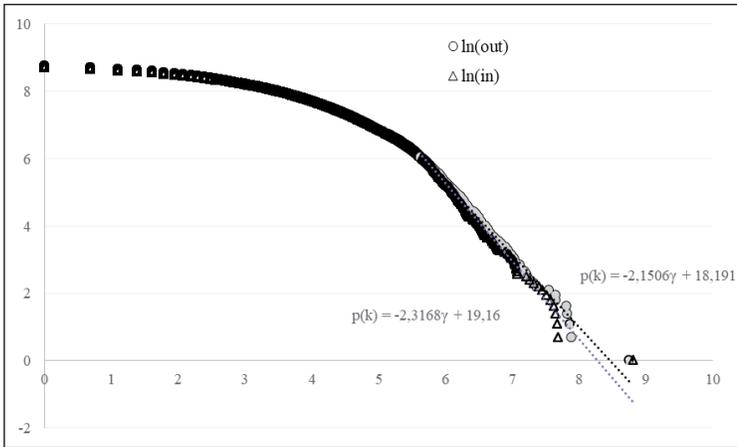
Siguiendo el aporte de Borgatti y Halgin (2011) quienes remarcan dos dominios principales de la teoría del análisis de redes sociales (ARS): “teoría de la red”: se refiere a los mecanismos y procesos que interactúan con las estructuras de la red para producir ciertos resultados para individuos y grupos, y “teoría de redes”: se refiere a procesos que determinan por qué las redes tienen la estructura que tienen.

Llamaremos macroanálisis y microanálisis de red a estos dos dominios, respectivamente. El macroanálisis se refiere a la taxonomía y la estructura que se asemeja al análisis estructural, y el microanálisis se refiere a los actores y grupos que se centran en las centralidades y el análisis de la comunidad, respectivamente.

Macroanálisis

Comenzamos por estimar tantas métricas de red como pudimos. Primero nos enfocamos en identificar patrones en la estructura de la red. Contamos con 5.116.920 enlaces (después de la depuración) y 94.62 nodos. La densidad es 0,00011488 y el grado promedio es 5,426, la longitud promedio de sendero es de 2.905363. El número de componentes débilmente conectados alcanza los 87.617. Como se muestra en la Figura 3, el grado de entrada y salida sigue aparentemente una distribución de la ley de potencia en toda la red, donde ambas pendientes se encuentran en el rango de valores entre 2 y 3.

Figura 3. Distribución de frecuencia de grados de entrada y salida



Dada la información obtenida en la muestra, estimamos diversas métricas estructurales para que tengan una imagen más precisa de su estructura.

Variables

En la Tabla 1 describimos las variables estimadas con datos de red. Confiamos principalmente en el software Pajek (Batajel y Mvar, 2003) para dicha tarea.

Una de estas variables altamente interesantes es el género y es particularmente interesante saber que la participación femenina promedio es de 0,67, mientras que la participación masculina es solo de 0,25. El resto de los actores representan 0,08 comprendiendo instituciones, eventos y negocios, entre otros. La Figura 4 muestra la participación de género por sector.

Tabla 1. Código y descripción de variables estructurales

Code	Metric and variable	Description
N	Vertices	Range of the graph
g	Unique Edges	Non-repeated edges
g'	Edges With Duplicates	Repeated interaction edges
TE	Total Edges	Sum of two previous edge classification
SL	Self-Loops	Edges connecting vertices with themselves.
RVPR	Reciprocated Vertex Pair Ratio	The number of vertex pairs that have edges in both directions divided by the number of vertex pairs that are connected by any edge.
RER	Reciprocated Edge Ratio	The number of edges that are reciprocated divided by the total number of edges.
CC	Connected Components	The number of connected components in the graph
SVCC	Single-Vertex Connected Components	The number of connected components that have only one vertex.
MVCC	Maximum Vertices in a Connected Component	The number of vertices in the connected component that has the most vertices.
MECC	Maximum Edges in a Connected Component	The number of edges in the connected component that has the most edges.
MGD	Maximum Geodesic Distance (Diameter)	Largest path between more distant pair of vertices
AGD	Average Geodesic Distance	Average path between every pair of vertices
Dens	Graph Density	This is a ratio that compares the number of edges in the graph with the maximum number of edges the graph would have if all the vertices were connected to each other.
IIDA	Input-Input Degree Assortativity	Indicates the correlation between node in-degree to its connections in-degree.
IODA	Input-Output Degree Assortativity	Indicates the correlation between node in-degree to its connections out-degree.
OIDA	Output-Input Degree Assortativity	Indicates the correlation between node out-degree to its connections in-degree.
OODA	Output-Output Degree Assortativity	Indicates the correlation between node out-degree to its connections out-degree.
nl _{nn}	$n \ln n =$	A measure related to the small world effect
M	Total adjacency index: $m=$	
M1	The Zagreb group index 1: $M1=$	Molecular complexity
M2	The Zagreb group index 2: $M2=$	Molecular connectivity

Xr	The Randic connectivity index: $Xr=$	Connectivity index analogous to M2
F	The Platt index: $F=$	Molecular complexity index
AD	Average degree	Average degree of nodes
UnPair	Number of unreachable pairs:	Pair of nodes unreachable by each other
ADARP	Average distance among reachable pairs	A measure of average path among connected nodes
Bet	Network betweenness centralization	Average betweenness among nodes
Mod	Modularity	Graph modularity
Clust	Clusters	Based on modularity, the number of identified clusters
$M (pM)$	Male (Percentage of Male)	Number (percentage) of males
$F (pF)$	Female (Percentage of Female)	Number (percentage) of females
$B (pB)$	Business (Percentage of Business)	Number (percentage) of business
$I (pI)$	Institution (Percentage of Institution)	Number (percentage) of institutions
$O (pO)$	Other (Percentage of Other)	Number (percentage) of other nodes

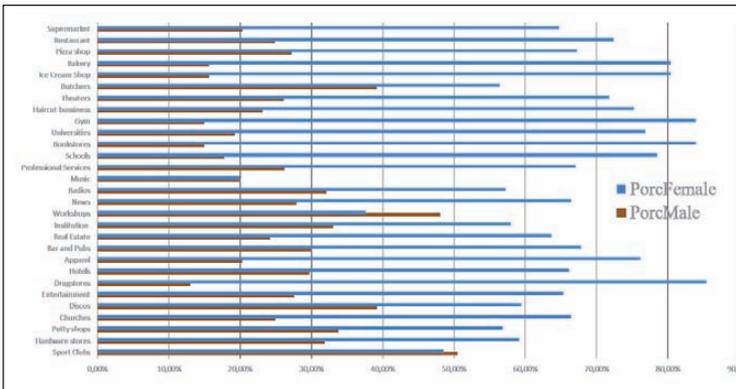


Figura 4. Participación por género por sector

Acciones de los actores

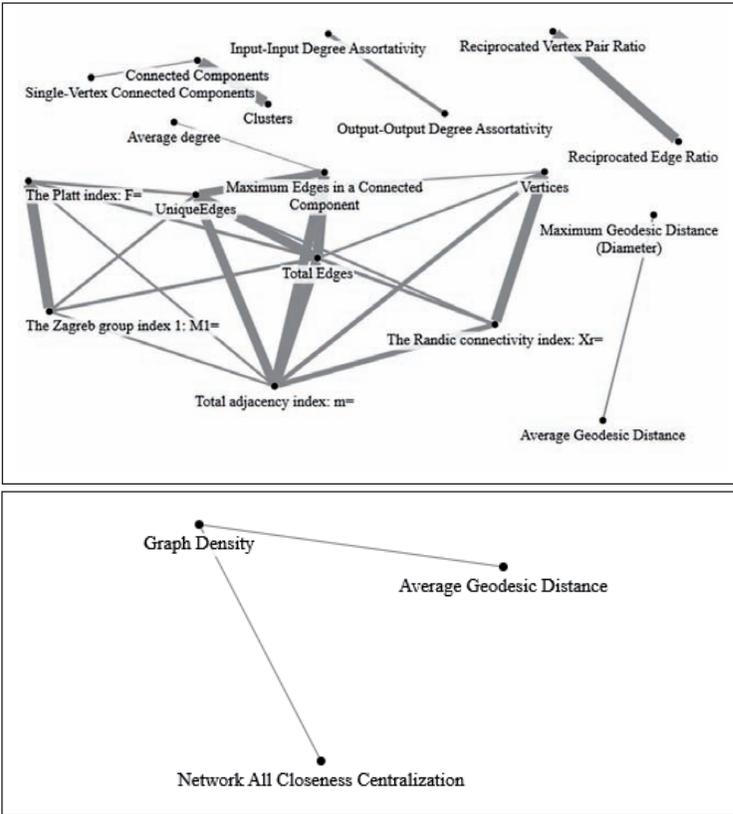
Las redes se descargaron con el paquete NodeXL (Smith et al., 2010) donde pudimos identificar diversas acciones de actores en el entorno de la página de fans de Facebook. Dado el formato del OSN Facebook, solo se contabilizan las siguientes acciones descritas en la tabla siguiente.

Tabla 2. Relaciones y tipos de enlaces detectados

Relationship	Type
Co-Liker	User Liked Same Post
Liked Comment	User Liked Comment User Liked Post
Reacted to Post	User Reacted Post
User Tagged in Post	User Tagged
User Tagged in Comment	
Commented Post	User Commented Post
Shared Post	User Shared Post
Co-Sharer	User Shared Same Post
Co-Reactor	User Reacted Same Post
Commented Comment	User Commented Comment
Consecutive Commenters	Consecutive Commenter
Co-Commenter	User Commented Same Post

Estimamos la correlación entre las métricas estructurales para identificar patrones. La matriz de correlación es 33x33 y eso presenta un resultado desordenado para una fácil interpretación. De esta manera, nos enfocaremos en correlaciones positivas superiores a 0,8 y correlaciones negativas superiores a -0,7. Esta información se resume en la Figura 5.

Figura 5. Correlación más alta que 0,8 (arriba) y menor que -0,7 (debajo)



Por otro lado, las métricas de distancia están asociadas negativamente. Por ejemplo, la centralización de la proximidad de la red, la distancia geodésica promedio y la densidad tienen una alta asociación negativa, lo que significa que cuanto mayor es la densidad, menor es la distancia promedio y el valor más bajo para la cercanía de la red. Todo muy comprensible en el contexto de las redes en línea.

Repetimos el ejercicio de explorar patrones en los datos a través de PCA pero ahora incluimos todas las variables,

el comportamiento y la estructura juntos. Esto se presenta en la Tabla 6. A medida que se introducen muchas más variables en el análisis, el porcentaje de varianza explicado disminuye. Requerimos en los últimos 10 componentes principales para alcanzar el 80% de explicabilidad. Solo analizaremos tres componentes principales que alcanzan el 54,61% de la varianza explicada. Esto se resume en la Tabla 7. Los valores propios positivos tienen una celda con fondo verde y los negativos usan un fondo rojo. En las dos primeras PC hay un patrón claro en el que participar en publicar y dar me gusta o responder a una publicación es negativo asociado a algún otro comportamiento del usuario (como comentar y reaccionar a una publicación), pero también a características estructurales como la cantidad de interacciones (medidas por la cantidad de aristas) y la participación femenina. Por lo tanto, las subredes altamente comprometidas están asociadas a la participación no femenina (hombres y organizaciones) y al pequeño tamaño y reacciones menores y comentarios posteriores.

Los componentes 3 y 4, por otro lado, señalan que la reciprocidad surge al comentar conjuntamente una publicación y que se opone a etiquetar y compartir publicaciones, y también a variables estructurales como las relacionadas con el rango (vértices, bordes, bucles automáticos, entre otros) y una vez más la participación femenina y la variedad de resultados. Ahora tenemos que las subredes no femeninas muestran más reciprocidad y comentarios compartidos al tiempo que presentan un rango más pequeño, una variedad de salida-salida y también un comportamiento de intercambio y etiquetado.

Tabla 6. ACP entre todas las variables

Variable	PC1	PC2	PC3	PC4	Component	Eigen value	Variance	Variance
Co.Liker...2		0,254734			1	11,19990121	25,45%	25,45%
Liked.Comment...3					2	6,182837678	14,05%	39,51%
Reacted.to.Post...4					3	3,337121674	7,58%	47,09%
User.Tagged.in.Post...5					4	3,310621247	7,52%	54,61%
User.Tagged.in.Comment...6			-0,20616		5	2,364362915	5,37%	59,99%
Commented.Post...7		0,208234			6	2,23337513	5,08%	65,06%
Shared.Post...8			-0,24803		7	1,827400427	4,15%	69,22%
Co.Sharer...9	0,212239		-0,24295		8	1,392377329	3,16%	72,38%
Co.Reactor...10		-0,27211			9	1,112738266	2,53%	74,91%
Commented.Comment...11	0,029527	0,233712			10	1,10713091	2,52%	77,43%
Consecutive.Commenters...12	-0,20365	0,21676			11	1,016643641	2,31%	79,74%
Co.Commenter...13			0,257838		12	0,997919721	2,27%	82,01%
User.Liked.Same.Post...14		0,243945			13	0,941025187	2,14%	84,14%
User.Liked.Comment...15					14	0,911640003	2,07%	86,22%
User.Liked.Post...16		0,219673			15	0,882896632	2,01%	88,22%
User.Reacted.Post...17			-0,3251		16	0,804528862	1,83%	90,05%
User.Tagged...18			-0,18358		17	0,657698774	1,49%	91,55%
User.Commented.Post...19	-0,20798				18	0,5907364	1,34%	92,89%
User.Shared.Post...20			-0,21774		19	0,531560423	1,21%	94,10%
User.Shared.Same.Post...21	0,213034		-0,21963		20	0,398708291	0,91%	95,00%
User.Reacted.Same.Post...22			-0,28521		21	0,348811285	0,79%	95,80%
User.Commented.Comment...23			0,200212		22	0,296866834	0,67%	96,47%
Consecutive.Committer...24	-0,21827				23	0,250024836	0,57%	97,04%
Vertices		-0,28005			24	0,236331704	0,56%	97,60%
UniqueEdges	-0,2067	-0,24298			25	0,220997889	0,50%	98,10%
EdgesWithDuplicates		-0,24541			26	0,189933309	0,43%	98,53%
TotalEdges	-0,21648	-0,26113			27	0,141447922	0,32%	98,85%
Self-Loops		-0,26672			28	0,126860992	0,29%	99,14%
Reciprocated Vertex Pair Ratio			0,365306		29	0,088340609	0,20%	99,34%
Reciprocated Edge Ratio			0,378488		30	0,076749672	0,17%	99,52%
Connected Components					31	0,068016658	0,15%	99,67%
Single-Vertex Connected Components					32	0,058803114	0,13%	99,81%
Maximum Vertices in a Connected Component					33	0,038113709	0,09%	99,89%
Maximum Edges in a Connected Component					34	0,014582617	0,03%	99,93%
Maximum Geodesic Distance (Diameter)					35	0,008436811	0,02%	99,94%
Average Geodesic Distance					36	0,007814853	0,02%	99,96%
Graph Density					37	0,005257918	0,01%	99,97%
Input-Input Degree Assortativity					38	0,00505103	0,01%	99,99%
Input-Output Degree Assortativity					39	0,002137472	0,00%	99,99%
Output-Input Degree Assortativity					40	0,001813823	0,00%	99,99%
Output-Output Degree Assortativity				-0,40456	41	0,001357519	0,00%	100,00%
n ln n					42	0,000633005	0,00%	100,00%
Male					43	0,000288694	0,00%	100,00%
Female	-0,20148		-0,20707		44	6,02E-30	0,00%	100,00%

Tabla 7. Los primeros cuatro PC

PC1	PC2	PC3	PC4
Co.Sharer...9	Co.Liker...2	Co.Commenter...13	Reciprocated Vertex Pair Ratio
Commented.Comment...11	Commented.Post...7		Reciprocated Edge Ratio
User.Shared.Same.Post...21	Commented.Comment...11	User.Tagged.in.Comment...6	
	Consecutive.Commenters...12	Shared.Post...8	Output-Output Degree Assortativity
	User.Liked.Same.Post...14	Co.Sharer...9	
Consecutive.Commenters...12	User.Liked.Post...16	User.Tagged...18	
User.Commented.Post...19	User.Commented.Comment...23	User.Shared.Post...20	
Consecutive.Committer...24		User.Shared.Same.Post...21	
UniqueEdges	Co.Reactor...10	Vertices	
TotalEdges	User.Reacted.Post...17	UniqueEdges	
Female	User.Reacted.Same.Post...22	EdgesWithDuplicates	
		TotalEdges	
		Self-Loops	
		Female	

Microanálisis

Estimamos centralidades y detectamos actores principales. Los grupos se detectan por clases de modularidad. Detectamos 8.900 comunidades. El componente gigante ocupa el 34% de todos los nodos (33 mil nodos). Ahora presentamos el Top 20 de centralidades en la Tabla 8. Estos actores son los principales difusores de información bajo cada definición relativa de centralidad. Una cosa está clara: es un entorno dominado por mujeres. Como tal, los pocos actores masculinos que se destacan se destacan con un fondo negro y una letra blanca en la celda. Sin embargo, surge otro resultado principal: los negocios (con las dos letras iniciales de codificación de Ne) están en todas las primeras posiciones, con algunos clubes (Cl) e instituciones (Inst) también presentes en las 10 primeras posiciones.

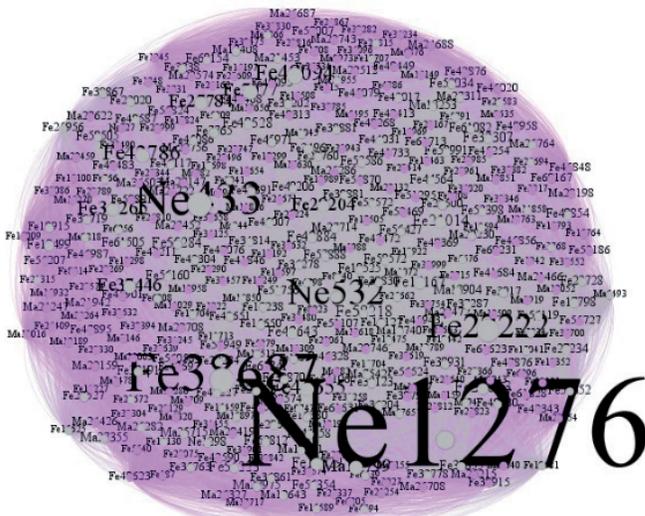
Tabla 8. Actores principales de acuerdo a la centralidad

Label	Indegree	Label	Outdegree	Label	Betweeness	Label	Eigenvector
Ne1276	6710	Ne433	6315	Ne560	0,0065	Ne1276	1,0000
Ne433	2189	Ne532	2651	Cl38	0,0036	Ne532	0,3496
Fe53025	2166	Fe5175	2598	Ne433	0,0023	Ins286	0,3032
Ins286	2110	Ne560	2515	Fe38687	0,0023	Fe57167	0,2905
Fe38687	2046	Fe7077	2489	Fe22151	0,0021	Fe56135	0,2794
Fe57167	1960	Fe2496	2111	Fe22222	0,0020	Fe53025	0,2625
Fe56135	1839	Ne253	2110	Ne1276	0,0014	Fe60071	0,2388
Fe60071	1723	Fe22151	1912	Ne500	0,0013	Fe59739	0,1998
Ma27208	1642	Fe10093	1575	Fe5175	0,0012	Fe38687	0,1908
Ne1275	1529	Ne500	1515	Ma8972	0,0011	Fe60034	0,1602
Fe22151	1428	Fe1751	1433	Fe48094	0,0011	Fe54156	0,1548
Ma28613	1357	Fe22222	1375	Fe7077	0,0009	Fe53216	0,1494
Ne532	1183	Fe15607	1356	Fe42786	0,0008	Fe22151	0,1485
Fe53216	1179	Ma4026	1322	Ma14799	0,0008	Ne433	0,1413
Fe38473	1168	Fe10681	1268	Fe27194	0,0007	Fe39641	0,1402
Fe41971	1163	Fe10729	1247	Fe27784	0,0007	Fe41971	0,1383
Fe54156	1155	Fe4201	1229	Fe44655	0,0007	Fe55786	0,1320
Fe50071	1143	Fe7765	1162	Ma27208	0,0007	Fe50071	0,1283
Fe60034	1141	Fe7552	1148	Fe36885	0,0007	Fe58218	0,1233
Fe39641	1122	Fe38687	1138	Fe34265	0,0006	Fe42823	0,1201

Nota: Los actores están codificados con una codificación de 2 letras y números. Las dos primeras letras identifican Ne: Negocios, Fe: Mujeres, Ma: Hombres, Cl: Club, Ins: Institución.

Esto también se observa cuando reducimos la red completa a su mayor visualización de k-core. Esto se observa en la Figura 7. La más alta tiene un $k = 392$ que reduce la red completa a una subred de 324 nodos. Hay una preeminencia de actores empresariales (Ne) y femeninos (Fe) en esta muestra altamente comprometida.

Figura 7. K-núcleo de la red completa con $k=392$ y 324 nodos



Conclusiones y futuros senderos

Extrajimos patrones de la gran red de una ubicación específica al estimar los comportamientos de los actores y las métricas estructurales y asociarlos a través del análisis de componentes principales. Nos gustaría continuar con la estimación e identificación de grupos intersectoriales por modularidad. Lo haríamos

También usaríamos los comentarios para el análisis de

sentimientos dado que se han eliminado las interacciones completas.

En general, nos gustaría proporcionar una mejor comprensión de la dimensión digital de nuestra vida.

Referencias

- Batagelj, V. y A. Mrvar: "Pajek – Analysis and Visualization of Large Networks". in Jünger, M., Mutzel, P., (Eds.) Graph Drawing Software. Springer, Berlin 2003. p. 77–103.
- Borgatti, S.P. y D.S Halgin (2011), "On Network Theory", *Organization Science* 22(5), DOI: 10.1007/978-1-4419-5513-5_2
- Brooks, B., B. Hogan, N. Ellison, C. Lampe, y J. Vitak (2014), "Assessing structural correlates to social capital in Facebook ego networks", *Social Networks* 38: 1–15. <https://doi.org/10.1016/j.socnet.2014.01.002>
- Smith, M., Ceni A., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., Dunne, C., (2010). NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010/2013/2016, from the Social Media Research Foundation: <https://www.smrfoundation.org>