Detección y análisis de comunidades en redes sociales (#TodosSomosPolitécnico)

Erick López Ornelas

Universidad Autónoma Metropolitana – Cuajimalpa, Av. Vasco de Quiroga 4871, Cuajimalpa, Santa Fe, C.P. 05348, México D. F., Teléfono: 52 55 5814 6500

elopez@correo.cua.uam.mx

Rocío Abascal Mena

Universidad Autónoma Metropolitana – Cuajimalpa, Av. Vasco de Quiroga 4871, Cuajimalpa, Santa Fe, C.P. 05348, México D. F., Teléfono: 52 55 5814 6500

mabascal@correo.cua.uam.mx

Sergio Zepeda Hernández

Universidad Autónoma Metropolitana – Cuajimalpa, Av. Vasco de Quiroga 4871, Cuajimalpa, Santa Fe, C.P. 05348, México D. F., Teléfono: 5814-6502

jzepeda@correo.cua.uam.mx

Resumen

Este artículo plantea la utilización del método de mapas jerárquicos para la detección de comunidades en una red social. El corpus utilizado en este artículo está basado en la interacción realizada en Twitter por un conjunto de actores, en el movimiento estudiantil del 2014 *#TodosSomosPolitécnico*. Al aplicar los mapas jerárquicos es posible identificar un conjunto de comunidades muy bien definidas en torno al movimiento. Un análisis basado en la visualización de las comunidades es realizado para corroborar la pertinencia de la detección.

Palabra(s) Clave(s): comunidades, mapas jerárquicos, redes sociales, Twitter.

1. Introducción

El uso de plataformas de redes sociales en contextos políticos alrededor del mundo, no solo ha cobrado protagonismo para difundir movimientos sociales sino que se ha convertido en una importante modalidad de socialización y difusión de información. En el caso de México, la integración de las Tecnologías de la Información y la Comunicación (TICs) y el acceso a las redes sociales digitales ha permitido que los jóvenes tengan un medio alternativo de comunicación y poder, de esta manera, interactuar, organizarse y visibilizar sus causas.

Una de las principales redes sociales que se utiliza es Twitter, cuya característica principal es que los mensajes que se envían tienen una longitud máxima de 140 caracteres, llamados *tuits*, que pueden ser almacenados y categorizados en temas a partir del uso de etiquetas precedidas por el símbolo # (almohadilla), mismo que permiten seguir, buscar y encontrar conversaciones relacionadas con un tema en común. La generación de comunidades internas en la red social es un fenómeno natural que se gesta cuando existen muchos actores que generan información relacionada con algún tema en específico.

La identificación de comunidades dentro de una red social es un tema importante ya que, además de simplificar la gran cantidad de información contenida en la red social, permite detectar grupos o sectores importantes que interactúan entre sí.

En este artículo hacemos un estudio sobre el movimiento estudiantil del Instituto Politécnico Nacional (IPN). Dicho movimiento, gestado a finales del año 2014, buscaba detener los cambios en planes de estudio y el nuevo reglamento interno del Instituto, así como democratizar las decisiones escolares. Para el análisis en Twitter se seleccionó el hashtag *#TodosSomosPolitécnico* como representativo del movimiento. Entre los múltiples elementos de análisis que se pueden generar del movimiento *#TodosSomosPolitécnico*, decidimos realizar un análisis de las principales comunidades que se gestaron dentro de la red social y de este modo mostrar de manera concisa y

simplificada la gran cantidad de información generada en la red. A continuación mostramos en la Fig. 1 la red generada por las intervenciones alrededor del movimiento *#TodosSomosPolitécnico*.

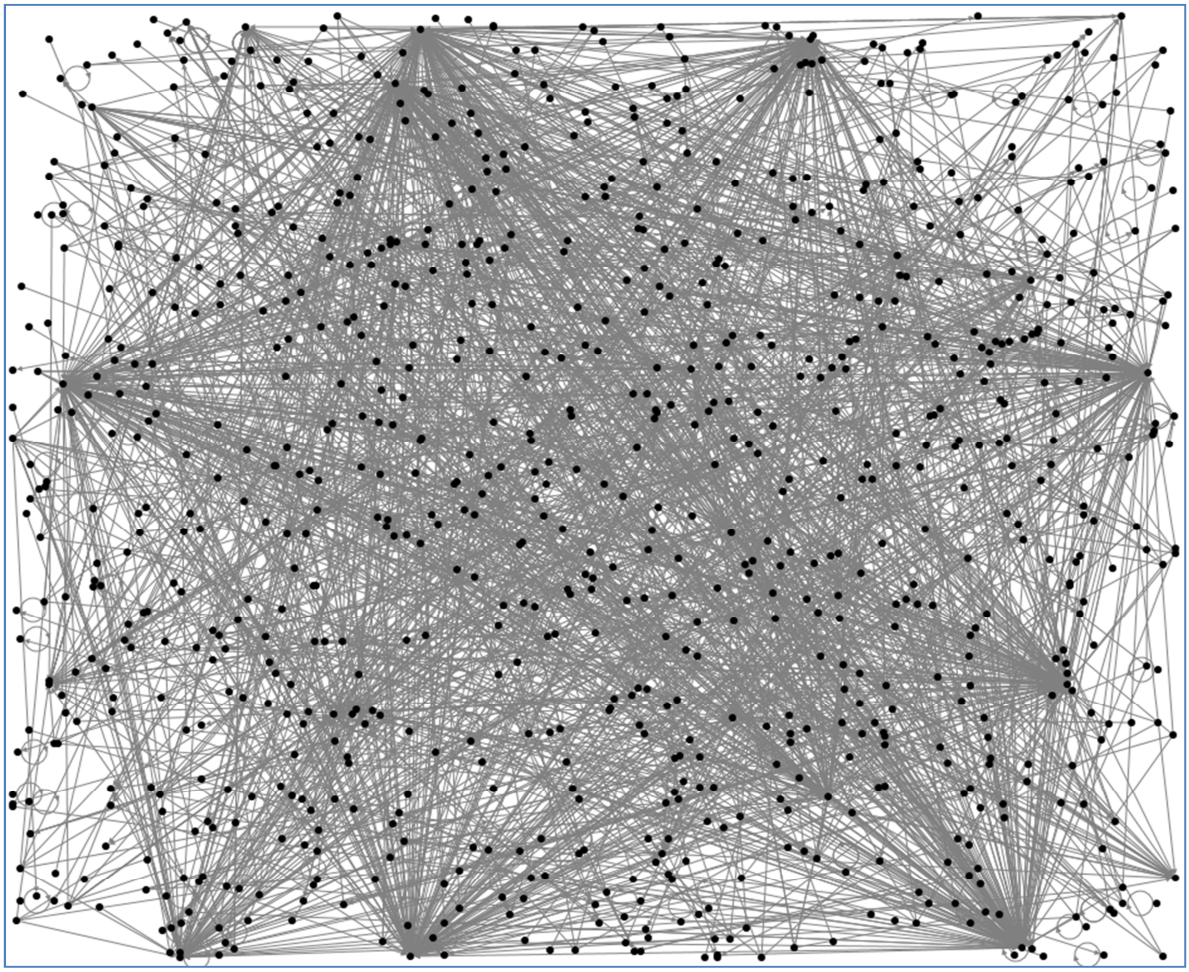

Fig. 1. Red de opiniones alrededor del movimiento #TodosSomosPolitécnico.

La información concentrada en la Fig. 1 corresponde a la extracción realizada el 24 de noviembre del año 2014. En la red están representados 857 tuiteros y 1884 interacciones entre ellos. Esta red de interacción es una red compleja por lo que es importante poder detectar comunidades para poder simplificar y analizar mejor las interacciones que en ella existen.

2. Detección de comunidades en una red social

La estructuración de comunidades, es una propiedad de las redes sociales actuales [1]. Una comunidad puede ser definida como un conjunto de nodos que están más densamente conectados entre ellos que con el resto de la red o que existe una mayor comunicación entre ellos. La importancia de este planteamiento radica en que se espera que los nodos que están contenidos dentro de una misma comunidad compartan atributos, características comunes o relaciones funcionales [2]. Sin embargo, no existe una definición exacta de lo que es, o cómo debería ser particionada la red en una comunidad.

Una partición es la división de una red en comunidades o *clusters*, de modo que todo nodo pertenece a algún *cluster*. Además, estas comunidades pueden estar jerárquicamente estructuradas, es decir, dos o más comunidades al fusionarse pueden formar una comunidad de un nivel superior. Este tipo de estructuras pueden ser representadas mediante un árbol o dendrograma [2]. Por otro lado, en el caso de que un nodo sea asignado a más de una comunidad hablamos de particiones empalmadas u *overlapping*. Es obvio que conforme crece el número de nodos, dificulta de manera extrema la selección de la mejor partición del grafo.

Cómo encontrar la partición óptima es, probablemente, el problema abierto más importante de la investigación en estructura de comunidades. Una gran variedad de métodos y algoritmos, cada uno de ellos con su propia definición intrínseca de comunidad, han sido desarrollados para intentar extraer la partición óptima de una red. Algunos de ellos tratan de optimizar un índice global de calidad de la partición, como puede ser su Modularity [3] o Surprise [4]. Otros, sin embargo, utilizan la matriz de adyacencia para extraer información del grafo, aplicando, por ejemplo, métodos espectrales [5]. Además, estimaciones de máxima verosimilitud [6], o elementos extraídos de la Teoría de la Información [7], son solo unos pocos ejemplos de métodos que han sido aplicados con relativo éxito a la búsqueda de comunidades.

En la literatura existen trabajos interesantes en torno a la detección de comunidades, por ejemplo [8] propuso un algoritmo basado en dos principios: (i) naturaleza intrínseca de las comunidades, y (ii) detección longitudinal, de igual manera en [9], desarrollaron un algoritmo de detección de comunidades solapadas basado en la idea de "amistad" entre los miembros de una comunidad, donde algunos de estos miembros se comportan como líderes de grupo. En [10] se propone un algoritmo capaz de indicar la influencia de los vértices del grafo contando el número de triángulos que cada vértice comparte con sus vértices adyacentes.

3. Objetivo y metodología

Los movimientos sociales actuales han empezado a utilizar elementos nuevos para su difusión, expresión y su expansión. Las redes sociales son un ejemplo muy claro de estos elementos donde su utilización ha comenzado a tomar fuerza sobre todo en movimientos sociales organizados por los jóvenes. El poder analizar lo que sucede en estos medios sociales digitales puede ser de gran ayuda para poder identificar las diferentes etapas por las que atraviesa un movimiento social y la fuerza que este movimiento puede llegar a alcanzar. Gente que esté relacionado con el movimiento, como políticos, asociaciones, empresas, podrían estar muy interesados en el poder analizar y entender el cómo se está desarrollando el movimiento y sobre todo, cuáles son las fuerzas que están interactuando en el mismo. Para realizar esta detección y análisis, se realizó un conjunto de extracciones de tuits generados el día del movimiento y utilizando el hashtag #TodosSomosPolitécnico. Se utilizó la herramienta NodeXL para realizar una primera extracción de los tuits. La visualización mostrada en la figura 1 fue realizada utilizando esta herramienta. Posteriormente, se utilizó el algoritmo basado en los Mapas Jerárquicos, el cual se detalla en la sección siguiente, para poder realizar el conjunto de agrupaciones y simplificaciones sobre el grado inicial correspondiente al movimiento social. Finalmente un análisis e interpretación es realizada a partir de la detección de comunidades obtenida.

4. Mapas jerárquicos

La definición más general de la comunidad es la de que en la red exista un grupo de nodos que están densamente interconectados. Mientras tanto, desde el punto de vista de la propagación de la información, una comunidad es un grupo de nodos en los que es más probable que se conserve la información en lugar de extenderse. Teniendo en cuenta que el modelo de propagación fundamental de la información es el “random walk” [7], entonces la estructura de una comunidad puede ser identificada mediante una búsqueda local en la estructura. Algunos estudios recientes [11,12] han demostrado que la modularidad [13], la cual es una función de calidad, es utilizada para encontrar comunidades donde existen grupos de nodos densamente conectados.

En este artículo, utilizamos los “mapas jerárquicos” para describir la dinámica de los enlaces y nodos dirigidos, así como de las redes ponderadas para identificar las interacciones locales dentro de la red. Estas interacciones locales permiten calcular el flujo de la información que se puede trasmitir por el nodo, en otras palabras el grado de interconexión que existe entre dos nodos [7, 14]. En consecuencia, es importante entender el flujo completo de la información en la red. Un grupo de nodos donde la información fluye de manera rápida y sencilla puede ser agregado y definido como un módulo bien conectado. Los enlaces entre los módulos y las veces que se comunicaron, permite identificar el grado de conexión entre los módulos, lo que permite realizar una simplificación del grupo generando un módulo y conservando al nodo principal de este módulo o comunidad [Fig. 2].

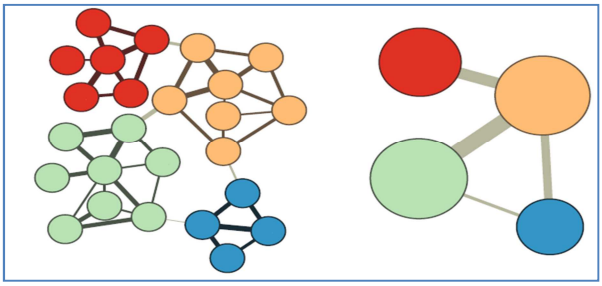

Fig. 2. Simplificación y agrupación de un grupo de nodos.

Esta agrupación radica en la dualidad entre la búsqueda de la estructura comunitaria en las redes y el problema de codificación: encontrar un código eficiente, buscamos entonces, una partición *M* de nodos *n* dentro de los módulos *m* para reducir al mínimo la distancia del llamado "random walk". Utilizando el módulo de partición M, la distancia promedio para que la información pase de un nodo a otro está dado por la ecuación 1[14], la cual está formada por dos términos, el primero es la entropía del movimiento entre los módulos, y el segundo es la entropía del movimiento en los módulos.

$$L(\mathbf{M}) = q_{\curvearrowright} H(\mathcal{Q}) + \sum_{i=1}^{m} p^{i}_{\circlearrowright} H(\mathcal{P}^{i}). \qquad [1]$$

Los resultados de este método son mostrados utilizando la red generada en el movimiento #*TodosSomosPolitécnico*.

5. Análisis de comunidades

La detección de comunidades en una red social que refleja movimientos sociales, resulta de gran interés para poder captar la fuerza del movimiento, de esta forma no solo podremos extraer actores importantes, sino grandes comunidades. Estas permitirán vislumbrar hacia dónde se está desplazando el movimiento, en qué etapa se encuentra y cuáles son las grandes fuerzas de interacción en el mismo.

En esta sección aplicamos el algoritmo de mapas jerárquicos descrito en la sección 3 para simplificar la red social inicial, lo cual provoca una disminución considerable en el número de nodos visibles. De los 857 actores en el movimiento (tuiteros), fueron agrupados y categorizados 15 grandes comunidades. De estas 15 comunidades identificadas, el actor más representativo fue seleccionado para abanderar la comunidad. De este modo las 15 comunidades se muestran en la Tabla 1, donde tenemos el módulo de agrupación, el nombre de la comunidad, la cantidad de información que comparten entre ellos, a lo que le llamamos el "Flujo de la información"

y el número de nodos que se encuentran agrupados o que son aglutinados dentro de esta gran comunidad.

Módulo	**Comunidad**	**Flujo de la información**	**Nodos en la comunidad**
1	donjosmansergio	27%	182
2	manceramiguelmx	13%	260
3	revistazocalo	9.1%	110
4	ipn_mx	3.3%	51
5	israellorenzana	3.3%	20
6	sep_mx	2.9%	58
7	rlescas	2.4%	42
8	arambula_rc1	1.9%	42
9	lenagarfe	1.4%	38
10	calacuayomx	0.97%	20
11	caluninno	0.48%	10
12	buendia_tunel	0.44%	7
13	jenarovillamil	0.43%	12
14	lamandujano	0.25%	3
15	alex_pahe	0.15%	2

Tabla 1. Principales comunidades identificadas en el movimiento #TodosSomosPolitécnico.

-

Al aplicar el algoritmo existe una simplificación visual muy importante, reduciendo la carga cognitiva para la identificación de las interacciones de cada comunidad. En la Fig. 3 mostramos un ejemplo de la selección del usuario @donjosmansergio y las interacciones que tiene con el resto en la red original. En este escenario resulta imposible identificar a los usuarios con los que interactúa y mucho menos identificar importancias en las relaciones. En la Fig. 4 se muestra una simplificación utilizado los mapas jerárquicos donde únicamente muestra las comunidades con las que interactúa. Esta nueva representación resulta de gran ayuda para identificar con qué comunidades se relaciona el usuario y la importancia que existe en esta relación.

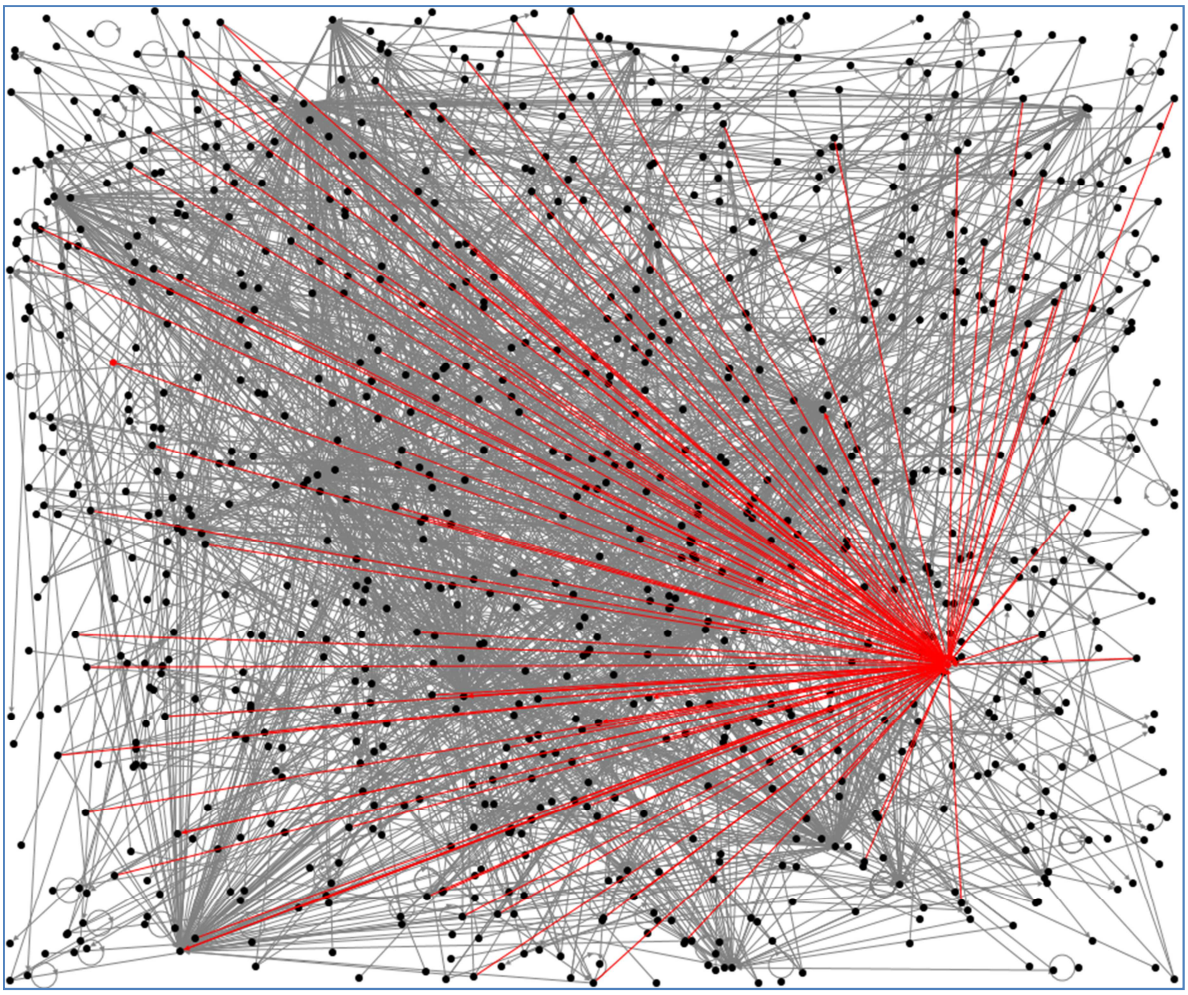

Fig. 3. Enlaces del usuario @donjosmansergio en la red original.

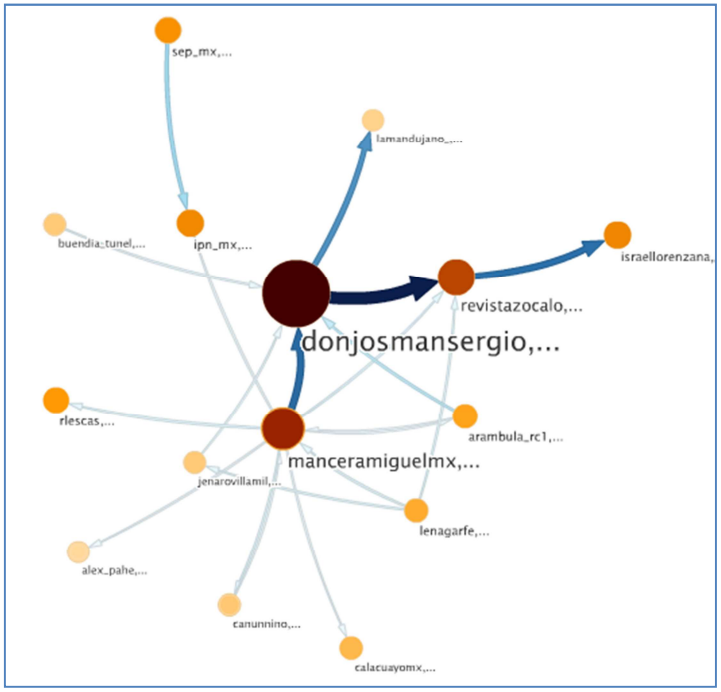

Fig. 4. Conjunto de comunidades identificadas en la red #TodosSomosPolitécnico.

Una característica importante al realizar la detección de comunidades, es que podemos analizar sub redes de información, esto es el poder analizar qué es lo que está pasando dentro de la comunidad, cuáles son las interacciones importantes, quiénes son los actores importantes en la sub red y qué fuerzas interactúan dentro de la comunidad. Este análisis es de gran ayuda para determinar la importancia de esta comunidad y sobre todo el poder identificar y clasificar a la comunidad.

Por ejemplo en la Fig. 5 exploramos la sub red “donjosmansergio” y podemos apreciar como existe una interconexión muy fuerte entre 4 actores (@donjosmansergio, @epigmenioibarra, @tapiafernanda y @luisghernan), los cuales son periodistas críticos de la información política actual. El usuario @donjosmansergio es a su vez un interlocutor importante con otros periodistas como (@virgiliocaballe o @julioastillero). Esta gran comunidad concentra a un gran conjunto de periodistas que estuvieron siguiendo e interactuando activamente en el movimiento.

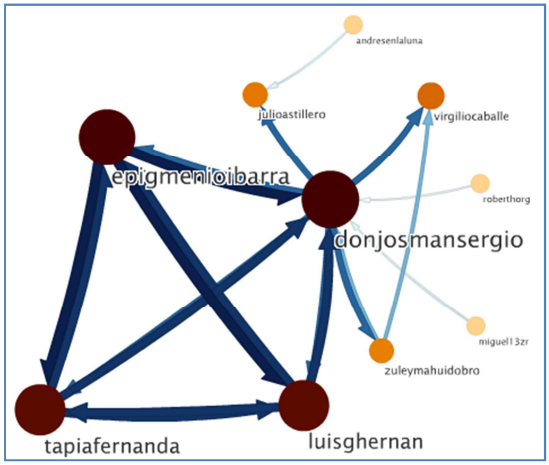

Fig. 5. Exploración interna de la sub red @donjosmansergio.

Otra comunidad importante a analizar es la sub red abanderada por el Jefe de Gobierno del Distrito Federal, el Dr. Miguel Ángel Mancera. En esta comunidad podemos observar como la @coordinadora1dmx, que en realidad es la coordinación de organizaciones sociales, colectivos, estudiantiles, sirve de enlace no solo con @manceramiguelmx, sino también muestra una interacción importante con @epn y el movimiento @yosoy_132cancun. Esta es una comunidad completamente de carácter político, donde interactúan básicamente políticos y organizaciones activistas. [Fig. 6]

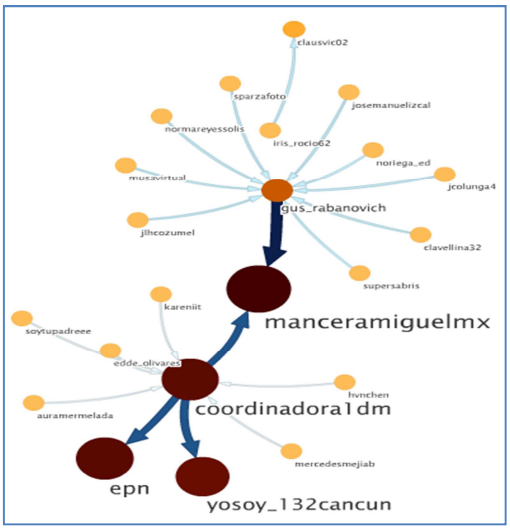

Fig. 6. Exploración interna de la sub red @manceramiguelmx.

La comunidad abanderada por “ipn_mx” muestra una agrupación de actores que están directamente relacionados con la educación superior en el país. Está abanderada por la figura institucional del Instituto Politécnico Nacional, pero dentro de esta comunidad contamos también con la presencia de @echuayffet (Secretario de Educación Pública), además de un conjunto de observatorios que buscaban determinar el rumbo que tomaría el movimiento. En la Fig. 7 se muestran estas relaciones.

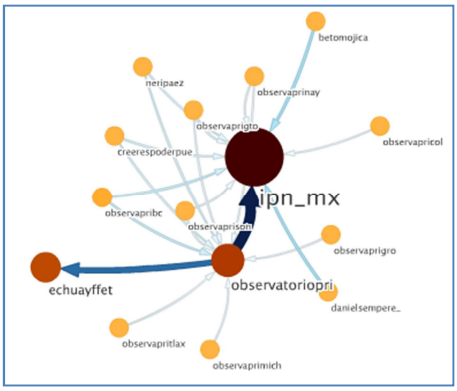

Fig. 7. Exploración interna de la sub red @ipn_mx.

La comunidad estudiantil del movimiento, también se encuentra muy bien identificada y agrupada. Al analizar la comunidad “rlescas” nos damos cuenta que la mayoría de los actores son estudiantes, los cuales intervinieron en el movimiento. @rlescas, @pacosanz9, @morganblut, @patricia_aries, todos se definen como estudiantes universitarios activos y críticos de lo que estaba pasando con el movimiento. [Fig. 8]

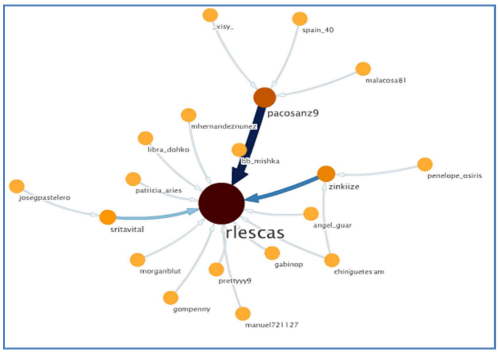

Fig. 8. Exploración interna de la sub red @rlescas.

Finalmente, la comunidad denominada “arambula_rc1” concentró actores que intervinieron en el movimiento, pero más como consumidores de la información y no como activistas del movimiento. Toda la estructura de esta comunidad converge en el usuario @arambula_rc1, quien es un usuario al que todos hacen mención. Esta característica resulta muy importante en las redes sociales ya que se convierte en un actor importante en esta sub red [Fig. 9]. Si quisiéramos diseminar un mensaje, sabríamos perfectamente al actor al que deberíamos dirigirnos.

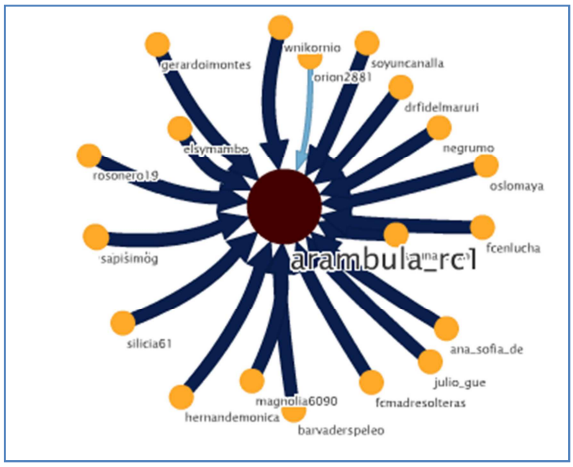

Fig. 9. Exploración interna de la sub red @arambula_rc1.

6. Discusión

El desarrollo de métodos que puedan detectar estructuras de comunidades en redes sociales es un elemento muy importante para el análisis que se pueda realizar sobre una red social, además de que puede desvelar relaciones subyacentes entre los elementos de una red.

En este artículo, hemos realizado un análisis de una red social generada a partir de un movimiento social estudiantil, el cual tomó gran importancia a nivel nacional a finales del año 2014. La detección de comunidades dentro de la red social permitió identificar un conjunto de características importantes:

1. El número de comunidades que se gestan internamente.
2. La importancia de cada una de las comunidades determinada por el tamaño de las interacciones existentes en la misma.
3. Identificación de actores importantes dentro de la comunidad y la interacción interna en cada comunidad, esto analizando la topología interna de cada comunidad.
4. Dirección de la red al analizar el tamaño de las interacciones que se están generando en cada comunidad.

7. Conclusiones

El Análisis de las Redes Sociales (ARS) resulta ser una herramienta muy poderosa para identificar comportamientos, identificar actores, hacer diagnósticos, descubrir relaciones y también para detectar comunidades. Además, este análisis es susceptible de ser aplicado en áreas diversas como la biología, la política, la computación, la sociología, etc.

La identificación de comunidades presenta un desafío interesante en la precisión con la que se categorizan los múltiples actores o nodos de una red. Sin embargo, los algoritmos actualmente implementados muestran que se está avanzando por el buen camino.

El poder extraer hashtags (#) a partir de una red social como Twitter, permite explorar movimientos y tendencias que actualmente se generan, así como identificar las interacciones que existen dentro de la misma.

Como trabajo futuro sería interesante poder comparar la técnica utilizada en este artículo (mapas jerárquicos) con algunas otras para poder determinar la confiablidad y la precisión de la misma. Además de poder experimentar esta técnica con otros ejemplos de red en áreas diversas.

8. Referencias

[1] M. Girvan, M. E. J. Newman, “Community structure in social and biological Networks”. Proceedings of the National Academy of Sciences of the United Statesof America. Vol. 99. No. 12. 2002. 7821–7826 pp.

[2] S. Fortunato, “Community detection in graphs”. Physics Reports. Vol. 486. No. 3-5. 2010. 75–174 pp.

[3] M. E. J. Newman, M. Girvan, “Finding and evaluating community structure in Networks”. Physical Review E - Statistical, Nonlinear and Soft Matter Physics. Vol. 69(2 Pt 2). No. 16. 2004.

[4] R. Aldecoa, I. Marín, “Deciphering network community structure by surprise”. PLoS ONE. Vol. 6. No. 9-8. 2011.

[5] H.-W. Shen, X.-Q. Cheng, “Spectral methods for the detection of network community structure: a comparative analysis”. Journal of Statistical Mechanics: Theory and Experiment. Vol. 10. No. 13. 2010.

[6] M. E. J. Newman, E. A. Leicht, “Mixture models and exploratory data analysis in networks”. Proc Natl Acad Sci USA. Vol. 104. No. 23. 2007. 9564–9569 pp.

[7] M. Rosvall, C. Bergstrom, “Maps of random walks on complex networks reveal community structure”. Proceedings of the National Academy of Sciences of the United States of America. Vol. 105. No. 4. 2008. 1118–1123 pp.

[8] R, Cazabet, F. Amblard, C. Hanachi, “Detection of Overlapping Communities in Dynamical Social Networks”. Proceedings of the 2010 IEEE International Conference on Social Computing. 2010. 309-314 pp.

[9] C. Palazuelos, M. Zorrilla, “FRINGE: A new Approach to the Detection of Overlapping Communities in Graphs”. ICCSA, Lecture Notes. 2011. 638-653 pp.

[10] A. Stanoev, D. Smikov, L. Kocarev, “Identifying Communities by influence dynamics in social networks”. Physical Review. 2011.

[11] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, “Stability of graph communities across time scales”. Proc. Natl. Acad. Sci. USA 107. No. 12755. 2010.

[12] Y. Kim, S.-W. Son, H. Jeong, “Finding communities in directed networks”. Phys. Rev. E. No. 81, 016103. 2010.

[13] M. E. J. Newman, “Modularity and community structure in networks”. Proc. Natl. Acad. Sci. USA. Vol. 103. No. 8577. 2006.

[14] M. Rosvall, D. Axelsson, C. T. Bergstrom, “The map equation”. Eur Phys J Spec Top. Vol. 178. No. 13. 2009. 23 pp.

9. Autores

El Dr. Erick López-Ornelas es profesor-investigador en el Departamento de Tecnologías de la Información de la Universidad Autónoma Metropolitana - Cuajimalpa en la Ciudad de México. Recibió su doctorado en Informática en la Universidad Paul Sabatier de Toulouse Francia, en el 2005. Sus intereses de investigación incluyen: visualización de información, análisis de redes sociales y la Interacción Humano Computadora.

La Dra. Rocío Abascal-Mena es profesor-investigador en el Departamento de Tecnologías de la Información de la Universidad Autónoma Metropolitana - Cuajimalpa en la Ciudad de México. Realizó su doctorado en el Instituto Nacional de Ciencias Aplicadas (INSA) de Lyon, Francia, en el 2005. Sus intereses actuales de investigación incluyen: desarrollo y explotación de bibliotecas digitales, Web semántica, lingüística computacional, recuperación y análisis de información, Interacción Humano Computadora, análisis de redes sociales.

El Dr. Sergio Zepeda-Hernández realizó su doctorado en Ingeniería Eléctrica en la especialidad de Informática del Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (México) en 2009. Actualmente es profesor-investigador del Departamento de Tecnologías de la Información de la Universidad Autónoma Metropolitana - Cuajimalpa en la Ciudad de México. Sus intereses de investigación incluyen la ingeniería web, análisis de redes sociales, recuperación de información, Interacción Humano Computadora, usabilidad y diseño de interfaces.