



Lección 1. Indicadores generales a nivel de nodo para analizar Redes Sociales



Análisis de Redes Sociales

Indicadores generales a nivel de nodos para analizar Redes Sociales

UNIVERSIDAD DE CÓRDOBA

El Análisis de Redes Sociales (ARS) estudia la interacción entre los actores en redes sociales. Es un área de investigación interdisciplinar en la que están implícitas la Sociología, la Psicología, la Antropología, la Física, las Matemáticas y la Informática, entre otras.

Los orígenes del ARS, para el desarrollo de conceptos sociológicos, puede situarse a comienzos de la década de 1930, donde el enfoque sociométrico se transforma como una forma de conceptualizar la estructura de las relaciones sociales establecidas entre pequeños grupos de individuos.

En este tema veremos algunos de los indicadores a nivel de nodo que nos permiten llevar a cabo un ARS.

1. Introducción

El ARS implica tener en cuenta múltiples niveles de análisis (Figura 1). Las diferencias entre los actores se pueden interpretar en base a las limitaciones y oportunidades que surgen de la forma en que éstos están inmersos en las redes; la estructura y el comportamiento de las redes están basadas en las interacciones que se producen entre sus actores.

En esta lección examinaremos los indicadores a nivel de nodo, que nos permiten realizar estudios de las relaciones locales que se encuentran en la red. A partir de las conexiones entre los nodos podemos encontrar nodos que tienen menos restricciones que otros y nodos que tienen más oportunidades que otros debido a la posición que tienen en la red. Tener una posición favorable significa que un nodo puede conseguir mejores ofertas en los intercambios y que será un foco para la deferencia y atención por parte de aquéllos en posiciones menos favorables. Para definir las posiciones favorables, las oportunidades y las restricciones contamos con diferentes formas de analizarlo. El ARS nos proporciona definiciones y mediciones que nos permiten hacer este análisis y son las veremos en este tema.

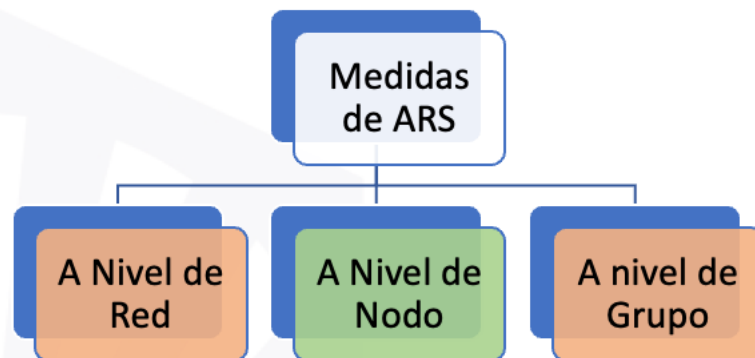


Figura 1. Diferentes niveles de análisis de redes sociales.

Recordamos que en el ARS podemos valorar diferentes medidas a distintos niveles. La semana anterior vimos los indicadores globales, y esta semana nos centraremos en las medidas locales y a nivel de grupo. Concretamente, en esta lección veremos las medidas a nivel de nodo [1]:

- **Medidas globales (a nivel de red):** proporcionan información más compacta que permite evaluar la estructura global de la red, proporcionando información sobre propiedades importantes de los fenómenos sociales subyacentes.
- **Medidas locales (a nivel de nodos):** proporcionan información más específica evaluando la estructura y relaciones de cada nodo en la red. Están basadas en el concepto general de centralidad (redes no dirigidas) o prestigio (redes dirigidas), una medida general de la posición de un nodo en la estructura global de la red social. Se usan para identificar los nodos claves de la red. Muestran como las relaciones se concentran en unos pocos individuos, dando una idea de su poder social.
- **Medidas a nivel de grupo:** nos permiten analizar las diferentes agrupaciones que se pueden realizar entre los nodos y cómo de cohesionados están esos grupos con respecto al resto.

2. Indicadores a nivel de nodo

Los indicadores a nivel de nodo también son conocidos como medidas locales [2]. Estos indicadores permiten evaluar la posición que cada uno de los actores alcanza en la estructura general.

En un primer análisis, el ARS se centra en estudiar el vecindario inmediato: las díadas y las tríadas en las cuales cada nodo está inmerso, y luego se va ampliando para estudiar los grupos concretos que se pueden crear dentro de la red. En la figura 2 se muestra la red social de amistades entre 34 integrantes de un club de kárate en una universidad estadounidense en la década de 1970. Tal y como se estudió la semana anterior, es una red pequeña, con solamente 34 nodos y 2 grupos diferenciados. Los dos grupos eran representados por los dos colores de los nodos y enlaces. Además, el tamaño de los nodos estaba en función del grado de cada actor, esto es, del número de conexiones que tienen. Así, en cada grupo nos encontramos actores más representativos por el hecho de tener un mayor número de conexiones. Esta red será la que se use de ejemplo para cada una de las medidas que se estudien en este tema.

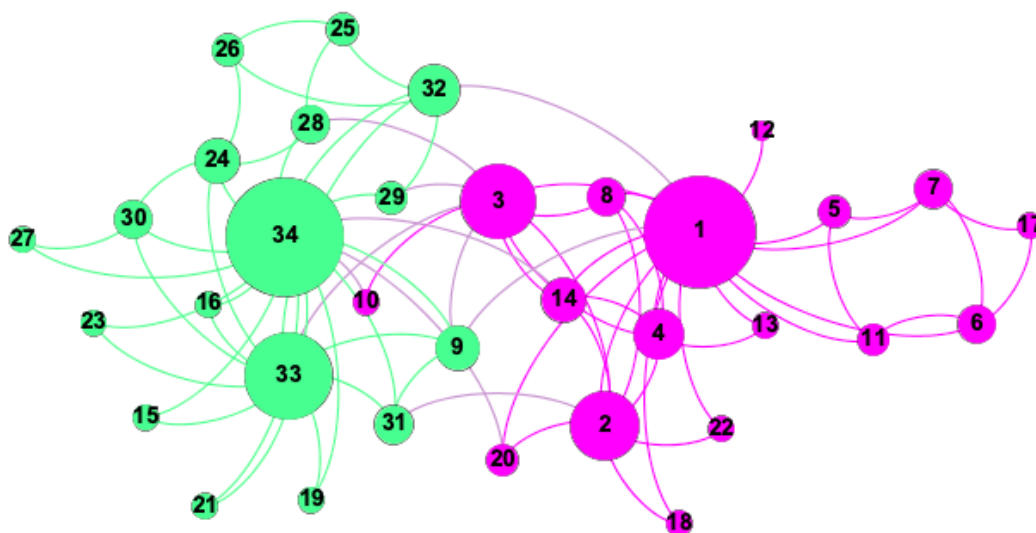


Figura 2. Red social de amistades de un club

El *grado de un actor* es el número de enlaces que lo conectan con otros nodos. Esta medida nos explica hasta qué punto un actor puede estar limitado o puede limitar a otros. Su cálculo es diferente según se trata de una red dirigida o no dirigida.

En una red no dirigida (Figura 3), el grado es el número de enlaces que sale de cada nodo. En el ejemplo, el grado del nodo A sería 2 ($\text{grado}(A) = 2$) y el grado del nodo C sería 3 ($\text{grado}(B) = 3$).

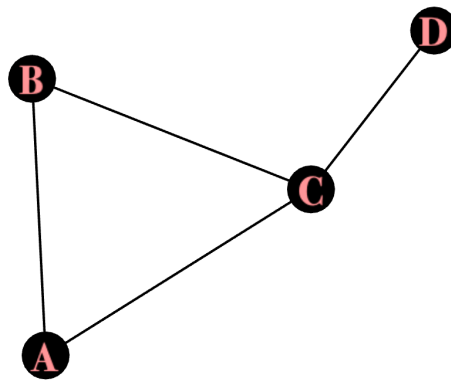


Figura 3. Ejemplo de red no dirigida

En una red dirigida (Figura 4) se debe diferenciar entre grado de entrada y de salida y el grado total sería la suma de ambas. El grado de entrada se corresponde con las conexiones que tienen como destino el nodo en cuestión y el grado de salida se corresponde con las conexiones que tienen como origen el nodo en cuestión. En el ejemplo el nodo C tendría como grado de entrada 2 y como grado de salida 1, su grado total sería, por tanto, 3. Así, $\text{grado}_{\text{in}}(C) = 2$, $\text{grado}_{\text{out}}(C) = 1$, el $\text{grado}_{\text{total}}(C) = 3$. Los nodos que tienen grado de entrada igual a 0 se llaman fuentes y los nodos que tienen grado de salida igual a 0 se llaman sumideros.

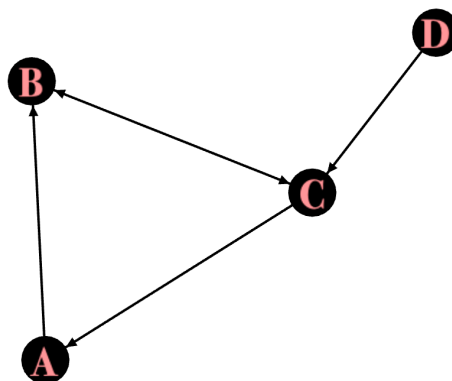


Figura 4. Ejemplo de red dirigida

Id	Etiqueta	Grado			
1	1	16	18	18	2
2	2	9	19	19	2
3	3	10	20	20	3
4	4	6	21	21	2
5	5	3	22	22	2
6	6	4	23	23	2
7	7	4	24	24	5
8	8	4	25	25	3
9	9	5	26	26	3
10	10	2	27	27	2
11	11	3	28	28	4
12	12	1	29	29	3
13	13	2	30	30	4
14	14	5	31	31	4
15	15	2	32	32	6
16	16	2	33	33	12
17	17	2	34	34	17

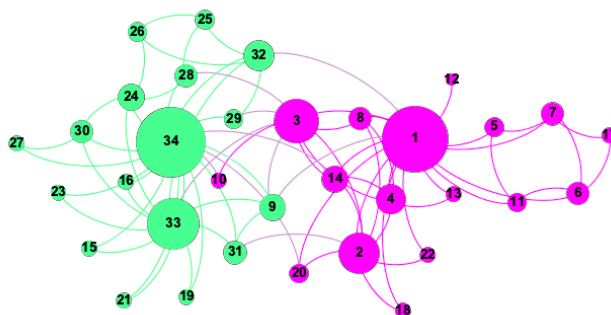


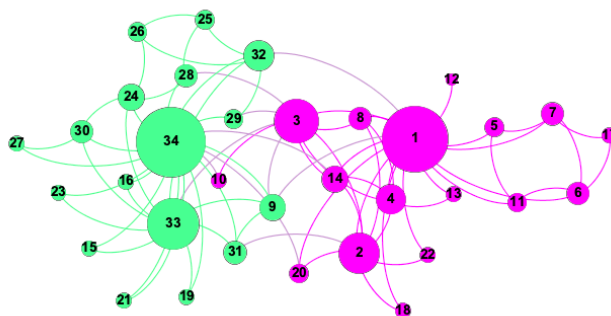
Tabla 1. Grado de los nodos de la red de la Figura 2

La red cuenta con un total de 34 nodos que representan a los miembros del club y 78 aristas que representan los vínculos antes descritos. En la Tabla 1, se muestra el grado medio de cada uno de los nodos de la red del club de kárate que se puso como ejemplo en la sección 2. Si observamos la figura 2 (se ha incluido junto a la tabla la representación de dicha red) donde los nodos estaban representados por su grado, se puede ver que los nodos etiquetados con 1 y con 34 son los dos nodos que tienen mayor grado, 16 y 17 conexiones, respectivamente. Luego, el nodo 33, 3 y 2, tienen 12, 10 y 9 conexiones, respectivamente. El resto de nodos oscila entre 5 y 2 conexiones. De este modo, se puede ver que los nodos 1 y 34 al tener mayor grado, serían más representativos usando esta medida. Parece claro que el hecho de tener mayor número de conexiones, te hace tener una posición más relevante en la red, tienes conexión directa con muchos otros actores, con lo que podrías influir en ellos más fácilmente. En la red que se muestra, esos nodos corresponden con el presendiente y el instructor de la clase de kárate. Dos actores bastante representativos en la red de estudio.

2.2 Accesibilidad

Un actor es accesible por otro si existe un conjunto de conexiones que permiten trazar un camino, desde el origen hasta el destino, sin tener en cuenta cuántos otros nodos puedan estar entre ellos. Si los datos son asimétricos o dirigidos, es posible que el actor A pueda alcanzar al actor B, pero que el actor B no pueda alcanzar al actor A. Con lazos simétricos o recíprocos cada par de actores es accesible si uno de ellos está conectado con el otro. Si algunos actores en una red no pueden alcanzar a otros, entonces existe una división potencial de la red. Podría indicar que la población que estamos estudiando está compuesta de subpoblaciones separadas.

En los datos de intercambio de información de la red de kárate se demuestra que todos los actores son accesibles por todos. Esto es algo que se puede verificar a simple vista. Si observamos la Figura 2, podemos ver que es posible encontrar un camino entre cualquier par de actores en la red. Al ser una red dirigida, la única opción es que tuviese diferentes componentes conexas, caso que no ocurre con la red de la Figura 2 que solamente tiene una única componente conexas.



(Recordatorio del grafo de la Figura 2 de la red que se está analizando)

2.3 Reciprocidad y transitividad

Con esta propiedad se busca clasificar las relaciones diádicas de cada actor. Así, algunos de los actores son fundamentalmente *fuentes* (es decir, tienen tendencia a enviar más que a recibir). Otros actores pueden ser fundamentalmente *sumideros* (es decir, tienen tendencia a recibir más que a enviar). Otros actores pueden enviar y recibir al mismo nivel.

En las redes no dirigidas todas las conexiones son recíprocas, pero en el caso de redes no dirigidas esto no es así, y resulta muy interesante su estudio. Si una gran proporción de conexiones son recíprocas, es una estructura en la cual se pueden producir una gran cantidad de comunicaciones inmediatas en los dos sentidos. Esto puede sugerir una estructura de las relaciones poco jerárquicas y con una gran cantidad de relaciones locales y particulares en ambas direcciones en lugar de una estructura monolítica.

Es también útil examinar las conexiones en términos de relaciones triádicas entre actores. El número de triadas que existen con 34 actores es muy grande; y, cuando se trabaja con datos dirigidos todavía son más. Comúnmente, el interés en la triadas se ha centrado en estudiar la propiedad de transitividad. Recordamos, como vimos la semana pasada que el principio de transitividad significa que si A está enlazada a B y B está enlazada a C, entonces A tendría que estar enlazada a C. Una forma especial de esta idea se conoce como la teoría del equilibrio. La teoría de equilibrio se centra en las relaciones transitivos tanto de efecto positivo como negativo. Se supone que, si a A le gusta B y a B le gusta C, a A le debería gustar C (equilibrio positivo). O si a A le gusta B y a B no le gusta C, entonces a A no le debería gustar C (equilibrio negativo). En estos estudios, lo relevante suele implicar si las relaciones de las triadas tienden hacia la transitividad o el equilibrio.

Para la red del club de kárate, el número de triángulos existentes en el grafo es de 45. Esto indica una alta cohesión existente entre los nodos que forman el grafo, e indica que existe una alta probabilidad que dado dos nodos enlazados, un tercer amigo sea común para ambos. Este factor tan característico de esta red es debido a que prácticamente todos los nodos o bien están conectados al nodo 1 o al nodo 34 (Figura 2), recordamos que son los nodos correspondientes al presidente del club de karate y al instructor respectivamente.

2.4 Distancia entre actores

Las propiedades de la red que hemos examinado hasta el momento tienen que ver principalmente con adyacencias (relaciones directas de un actor con otro). No obstante, la manera en que las personas están inmersas en las redes es más complicada. Dos personas, A y B, pueden tener cada una de ellas cinco amigos. Pero si ninguno de los amigos de A tiene otros amigos además de A, A estaría más aislado que B, en el que cada uno de sus cinco amigos, tiene otros cinco amigos más. La información disponible para B y el potencial para influir en las decisiones es mucho mayor para B. En este caso, ser el amigo de un amigo, también nos puede dar más visibilidad en nuestra red.

Para estudiar cómo las personas están inmersas en las redes, es importante examinar la distancia a la que está situado un actor en relación con los demás. Si dos actores son adyacentes, la distancia entre ellos es una (esto es, la información necesitaría un paso para llegar desde el emisor hasta el receptor). Si A está comunicado con B, B está comunicado con C (y A no está comunicado directamente con C), entonces los actores A y C están a una distancia de dos. Cuántos actores están a diferentes distancias unos de otros puede ser importante para entender las diferencias entre actores y las limitaciones y oportunidades que tiene cada uno como resultado de su posición. A veces, también resulta relevante estudiar de cuántas maneras se pueden conectar dos actores con una distancia dada. Esto es, puede el actor A alcanzar al actor B de más de una manera. Esta característica es relevante porque múltiples conexiones diferentes entre dos actores, nos informa de conexiones más fuertes entre esos actores.

Las distancias también nos dan información de las características de cómo puede ser la comunicación en una red. Cuando las distancias entre actores son grandes, se puede necesitar mucho tiempo para que la información se difunda a través de una población. Si existe mucha variabilidad en las distancias entre actores, se puede producir diferenciación y estratificación entre los actores. Los actores que están próximos pueden ser capaces de ejercer más poder que aquellos que están más distantes. En general, los actores que están conectados con distancias cortas pueden tener relaciones más fuertes. Los actores que tiene múltiples conexiones con los otros actores pueden tener lazos más fuertes. Su conexión estará menos sujeta a intrusiones y por lo tanto serán más estables y fiables.

Recordando los conceptos de distancia que se estudiaron la semana pasada, tenemos: paseos, senderos y caminos. Estos conceptos representan diferentes clases de conexiones que nos permiten analizar las distancias entre actores. Por un lado, los paseos son cualquier secuencia de nodos y aristas que une dos nodos, en los senderos no se pueden repetir ramas y en los caminos no se pueden repetir los vértices. Por reducir los cálculos, en el ARS muchas veces se trabaja con caminos, y más concretamente con caminos mínimos, que en el ARS son ampliamente conocidos como distancia geodésica.

En la figura 5, se muestra un grafo donde hay un paseo de longitud 2 (A, B, C). Hay un paseo de longitud 3 (A, B, D, C). Existen algunos paseos de longitud 4 (A, B, E, D, C; A, B, D, B, C; A, B, E, B, C). A causa de que no están restringidos, los mismo actores y relaciones pueden usarse más de una vez en un paseo dado. No hay ciclos que comiencen y acaben en A. Sí hay algunos ciclos que empiezan y acaban con el actor B (B, D, C, B; B, E, D, B; B, C, D, E, B). Siguiendo con este ejemplo: (A, B, D, B, C) sería un paseo, pero no es un sendero porque la relación entre B y D se usa más de una vez. También, hay un número limitado de caminos entre A y C: (A, B, C; A, B, D, C; A, B, E, D, C) y hay un único camino mínimo entre A y C: (A, B, C).

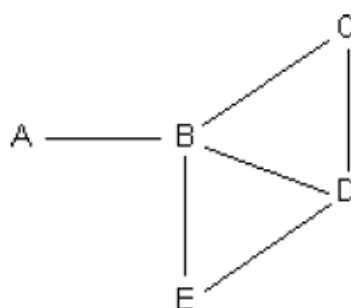


Figura 5. Ejemplo de grafo para evaluar distancias

La distancia geodésica o camino mínimo, como se ha comentado, es el número de relaciones en el camino más corto posible de un actor a otro. Puede haber muchas conexiones entre dos actores en una red. Si consideramos cómo la relación entre dos actores puede proporcionar oportunidades y limitaciones, puede darse el caso que no todas importen. Por ejemplo, supongamos que estoy intentando enviar un mensaje a Sara. Dado que conozco su dirección de correo electrónico, puedo enviarle un mensaje directamente (un camino de longitud 1). Pero también conozco a Luis, que tiene el correo electrónico de Sara. Así, podría enviar un mensaje a Luis para que se lo reenvíe a Sara (un camino de longitud 2). Con estas dos opciones, probablemente se escogerá el camino geodésico (el camino directo) porque da menos problemas y es más rápido y también porque no depende de Luis. Por lo tanto, el camino geodésico (o caminos, puede que haya más de un camino geodésico) es el óptimo entre dos actores y muchos algoritmos parten de estos caminos para realizar el cálculo de otras medidas. En redes con una alta densidad, el camino geodésico es generalmente pequeño. Esto sugiere que la información puede viajar bastante rápido en la red. Cuando una red no está plenamente conectada, no se puede definir las distancias geodésicas entre los pares.

En el ejemplo de la Figura 2, vemos que se trata de una red totalmente conectada, un mensaje que empieza en algún lugar alcanzará cualquier otro. La distancia geodésica media del grafo generado en esta red es de 2.41 con una desviación estándar de 0.93. El diámetro (distancia geodésica más larga existente en la red) es de 5 saltos. El valor del camino medio indica que el flujo de información en la red es muy eficiente al necesitar sólo 2.41 saltos de media en llegar a cualquier nodo, siendo en el caso más extremo cuando se necesitan 5 saltos para llegar de un nodo a otro. Esto sugiere un sistema en el cual la información probablemente alcanzará a cualquier nodo rápidamente. Los nodos 1 y 34 tienen mucho que ver en este aspecto ya que si fueran eliminados del grafo, generarían zonas aisladas y cambiarían radicalmente los datos obtenidos.

En la figura 6 se muestra las distancias geodésicas entre los nodos de la red de amistad del club de kárate (Figura 2). Se puede ver que existe una distancia geodésica para cada uno de los pares de actores, esto es, el grafo está plenamente conectado y los actores son accesibles por todos los otros (o lo que es lo mismo: existe un camino entre cualquier actor a cualquier actor). Además, las distancias geodésicas entre dos actores suelen ser cortas, entre distancia 2 y 3 suelen alcanzarse la mayoría de los actores de esta red. En la figura 7, se muestra específicamente la frecuencia de caminos geodésicos de longitud 1, 2, 3, 4 y 5 (que sería el camino geodésico más largo en esta red) que tiene esta red.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	0	1	1	1	1	1	1	1	1	2	1	1	1	1	3	3	2	1	3	1	3	1	3	3	2	2	3	2	2	3	2	1	2	2
2	1	0	1	1	2	2	2	1	2	2	2	2	2	1	3	3	3	1	3	1	3	1	3	3	3	3	3	2	2	3	1	2	2	2
3	1	1	0	1	2	2	2	1	1	1	2	2	2	1	2	2	3	2	2	2	2	2	2	2	2	3	3	1	1	2	2	2	1	2
4	1	1	1	0	2	2	2	1	2	2	2	2	1	1	3	3	3	2	3	2	3	2	3	3	3	3	3	2	2	3	2	2	2	2
5	1	2	2	2	0	2	1	2	2	3	1	2	2	2	4	4	2	2	4	2	4	2	4	2	4	3	3	4	3	3	4	3	2	3
6	1	2	2	2	2	0	1	2	2	3	1	2	2	2	4	4	1	2	4	2	4	2	4	4	4	3	3	4	3	3	4	3	2	3
7	1	2	2	2	1	1	0	2	2	3	2	2	2	2	4	4	1	2	4	2	4	2	4	4	4	3	3	4	3	3	4	3	2	3
8	1	1	1	1	2	2	2	0	2	2	2	2	2	2	3	3	3	2	3	2	3	2	3	3	3	3	4	2	2	3	2	2	2	3
9	1	2	1	2	2	2	2	2	0	2	2	2	2	2	2	2	3	2	2	2	2	2	2	2	2	3	3	2	2	2	2	1	2	1
10	2	2	1	2	3	3	3	2	2	0	3	3	3	2	2	2	4	3	2	2	2	3	2	2	3	3	2	2	2	2	2	2	2	1
11	1	2	2	2	1	1	2	2	2	3	0	2	2	2	4	4	2	2	4	2	4	2	4	4	3	3	4	3	3	3	4	3	2	3
12	1	2	2	2	2	2	2	2	2	3	2	0	2	2	4	4	3	2	4	2	4	2	4	2	4	3	3	4	3	3	4	3	2	3
13	1	2	2	1	2	2	2	2	2	3	2	2	0	2	4	4	3	2	4	2	4	2	4	2	4	3	3	4	3	3	4	3	2	3
14	1	1	1	1	2	2	2	2	2	2	2	2	2	0	2	2	3	2	2	2	2	2	2	2	3	3	2	2	2	2	2	2	2	1
15	3	3	2	3	4	4	4	3	2	2	4	4	4	2	0	2	5	4	2	2	2	4	2	2	3	3	2	2	2	2	2	2	1	1
16	3	3	2	3	4	4	4	3	2	2	4	4	4	2	2	0	5	4	2	2	2	4	2	2	3	3	2	2	2	2	2	2	1	1
17	2	3	3	3	2	1	1	3	3	4	2	3	3	3	5	5	0	3	5	3	5	3	5	5	4	4	5	4	4	5	4	3	4	4
18	1	1	2	2	2	2	2	2	2	3	2	2	2	2	4	4	3	0	4	2	4	2	4	4	4	3	3	4	3	3	4	2	3	3
19	3	3	2	3	4	4	4	3	2	2	4	4	4	2	2	2	5	4	0	2	2	4	2	2	3	3	2	2	2	2	2	2	1	1
20	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2	2	0	2	2	2	2	3	3	2	2	2	2	2	2	2	1
21	3	3	2	3	4	4	4	3	2	2	4	4	4	2	2	2	5	4	2	2	0	4	2	2	3	3	2	2	2	2	2	2	2	1
22	1	1	2	2	2	2	2	2	2	3	2	2	2	2	4	4	3	2	4	2	4	0	4	4	3	3	4	3	3	4	2	2	3	3
23	3	3	2	3	4	4	4	3	2	2	4	4	4	2	2	2	5	4	2	2	2	4	0	2	3	3	2	2	2	2	2	2	1	1
24	3	3	2	3	4	4	4	3	2	2	4	4	4	2	2	2	5	4	2	2	2	4	2	0	2	1	2	1	2	1	2	2	1	1
25	2	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	4	3	3	3	3	3	3	2	0	1	3	1	2	3	3	1	2	2
26	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	3	3	3	3	3	3	1	1	0	3	2	2	2	2	1	2	2
27	3	3	3	3	4	4	4	4	2	2	4	4	4	2	2	2	5	4	2	2	2	4	2	2	3	3	0	2	2	1	2	2	2	1
28	2	2	1	2	3	3	3	2	2	2	3	3	3	2	2	2	4	3	2	2	2	3	2	1	1	2	2	0	2	2	2	2	2	1
29	2	2	1	2	3	3	3	2	2	2	3	3	3	2	2	2	4	3	2	2	2	3	2	2	2	2	2	2	0	2	2	1	2	1
30	3	3	2	3	4	4	4	3	2	2	4	4	4	2	2	2	5	4	2	2	2	4	2	1	3	2	1	2	2	0	2	2	1	1
31	2	1	2	2	3	3	3	2	1	2	3	3	3	2	2	2	4	2	2	2	2	4	2	2	3	3	2	2	2	2	0	2	1	1
32	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	2	2	2	1	1	2	2	1	2	0	1	1	
33	2	2	1	2	3	3	3	2	1	2	3	3	3	2	1	1	4	3	1	2	1	3	1	1	2	2	2	2	2	1	1	1	0	1
34	2	2	2	2	3	3	3	3	1	1	3	3	3	1	1	1	4	3	1	1	1	3	1	1	2	2	1	1	1	1	1	1	1	0

Figura 6. Distancias geodésicas de la red de amistad del club de kárate (grafo de la Figura 2)

Frequencies

	1	2
	Freq	Prop
1 1	156	0.139
2 2	530	0.472
3 3	274	0.244
4 4	146	0.130
5 5	16	0.014

5 rows, 2 columns, 1 levels.

Average: :2.408
Std Dev: :0.930

Figura 7. Estadísticas de los caminos geodésicos

A veces la redundancia de la conexión es una importante característica de la estructura de una red. Si hay muchos caminos eficientes conectando dos actores, las probabilidades de que una señal vaya de un sitio a otro, mejoran. Por supuesto, si dos actores son adyacentes, solamente existiría este camino. En la Figura 8, puede verse el número de caminos geodésicos que hay entre cada par de actores. Se puede apreciar que, entre determinados nodos, tenemos muchos caminos geodésicos diferentes para alcanzarlos. Normalmente, se corresponde con aquellos que tienen una distancia geodésica mayor. De este modo, no solamente las distancias geodésicas son de corta distancia, sino que hay múltiples caminos más cortos entre los actores. Esto nos permite concluir que el flujo de información probablemente no se interrumpirá a causa de la existencia de múltiples caminos entre esos nodos. Será necesario un estudio más exhaustivo, para determinar si entre esos múltiples caminos, siempre hay algún actor fijo, ya que eso le convertiría en un intermediario poderoso. Esto será algo que estudiemos en las siguientes lecciones. A priori, el hecho de tener caminos alternativos e igual de eficientes, nos podría hacer que podamos obviar pasar por determinados actores.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	7	7	2	1	7	1	7	1	7	9	1	1	4	1	2	7	2	1	3	4
2	1	1	1	1	1	1	1	1	3	1	1	1	2	1	5	5	2	1	5	1	5	1	5	6	2	1	3	1	1	5	1	1	2	3
3	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	2	1	2	1	2	1	2	1	6	7	1	1	1	3	3	1	6
4	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	2	2	2	2	2	2	2	2	3	2	1	1	1	1	2	1	1	1	
5	1	1	1	1	1	3	1	1	1	1	1	1	1	1	7	7	1	1	7	1	7	1	7	9	1	1	4	1	2	7	2	1	3	4
6	1	1	1	1	3	1	1	1	1	1	1	1	1	1	7	7	1	1	7	1	7	1	7	9	1	1	4	1	2	7	2	1	3	4
7	1	1	1	1	1	1	1	1	1	1	3	1	1	1	7	7	1	1	7	1	7	1	7	9	1	1	4	1	2	7	2	1	3	4
8	1	1	1	1	1	1	1	1	2	1	1	1	2	4	1	1	2	2	1	2	1	2	1	2	2	1	15	1	1	1	1	1	1	14
9	1	3	1	2	1	1	1	2	1	2	1	1	1	3	2	2	2	1	2	2	2	1	2	2	5	5	1	2	2	2	1	3	1	1
10	1	1	1	1	1	1	1	1	2	1	1	1	2	2	1	1	2	2	1	1	1	2	1	1	3	2	1	2	2	1	1	1	2	1
11	1	1	1	1	1	1	3	1	1	1	1	1	1	1	7	7	1	1	7	1	7	1	7	9	1	1	4	1	2	7	2	1	3	4
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	7	7	2	1	7	1	7	1	7	9	1	1	4	1	2	7	2	1	3	4
13	1	2	2	1	1	1	1	2	1	2	1	1	1	2	9	9	2	1	9	1	9	1	9	12	1	1	5	2	3	9	3	1	4	5
14	1	1	1	1	1	1	1	4	3	2	1	1	2	1	1	1	2	2	1	3	1	2	1	1	4	3	1	2	2	1	2	2	2	1
15	7	5	1	2	7	7	7	1	2	1	7	7	9	1	1	2	14	12	2	1	2	12	2	2	3	4	1	1	1	2	2	2	1	1
16	7	5	1	2	7	7	7	1	2	1	7	7	9	1	2	1	14	12	2	1	2	12	2	2	3	4	1	1	1	2	2	2	1	1
17	2	2	2	2	1	1	1	2	2	2	1	2	2	2	14	14	1	2	14	2	14	2	14	18	2	2	8	2	4	14	4	2	6	8
18	1	1	2	2	1	1	1	2	1	2	1	1	1	2	12	12	2	1	12	2	12	2	12	15	1	1	7	2	3	12	1	1	5	7
19	7	5	1	2	7	7	7	1	2	1	7	7	9	1	2	2	14	12	1	1	2	12	2	2	3	4	1	1	1	2	2	2	1	1
20	1	1	2	2	1	1	1	2	2	1	1	1	1	3	1	1	2	2	1	1	1	2	1	1	3	3	1	1	1	1	2	2	1	1
21	7	5	1	2	7	7	7	1	2	1	7	7	9	1	2	2	14	12	2	1	1	12	2	2	3	4	1	1	1	2	2	2	1	1
22	1	1	2	2	1	1	1	2	1	2	1	1	1	2	12	12	2	2	12	2	12	1	12	15	1	1	7	2	3	12	1	1	5	7
23	7	5	1	2	7	7	7	1	2	1	7	7	9	1	2	2	14	12	2	1	2	12	1	2	3	4	1	1	1	2	2	2	1	1
24	9	6	2	3	9	9	9	2	2	1	9	9	12	1	2	2	18	15	2	1	2	15	2	1	2	1	2	1	1	1	2	3	1	1
25	1	2	1	2	1	1	1	2	5	3	1	1	1	4	3	3	2	1	3	3	3	1	3	2	1	1	2	1	1	5	3	1	1	2
26	1	1	6	1	1	1	1	1	5	2	1	1	1	3	4	4	2	1	4	3	4	1	4	1	1	1	3	2	1	1	4	1	2	2
27	4	3	7	1	4	4	4	15	1	1	4	4	5	1	1	1	8	7	1	1	1	7	1	2	2	3	1	1	1	1	1	1	2	1
28	1	1	1	1	1	1	1	1	2	2	1	1	2	2	1	1	2	2	1	1	1	2	1	1	1	2	1	1	2	2	1	2	3	1
29	2	1	1	1	2	2	2	1	2	2	2	2	2	3	2	1	1	4	3	1	1	1	3	1	1	1	1	1	2	1	1	1	3	1
30	7	5	1	2	7	7	7	1	2	1	7	7	9	1	2	2	14	12	2	1	2	12	2	1	5	1	1	2	1	1	2	2	1	1
31	2	1	3	1	2	2	2	1	1	1	2	2	3	2	2	2	4	1	2	2	2	1	2	2	3	4	1	1	1	2	1	2	1	1
32	1	1	3	1	1	1	1	1	3	1	1	1	1	1	2	2	2	2	1	2	2	1	2	3	1	1	1	2	1	2	2	1	1	1
33	3	2	1	1	3	3	3	1	1	2	3	3	4	2	1	1	6	5	1	1	1	5	1	1	1	2	2	3	3	1	1	1	1	1
34	4	3	6	1	4	4	4	14	1	1	4	4	5	1	1	1	8	7	1	1	1	7	1	1	2	2	1	1	1	1	1	1	1	1

Figura 8. Número de caminos geodésicos entre cada nodo

2.5 Coeficiente de clustering

El coeficiente de clustering o coeficiente de agrupamiento de un nodo, i , se define como el ratio del número de enlaces existentes sobre todos los enlaces posibles entre sus vecinos. A nivel de red general, se estudiaba el coeficiente de agrupamiento medio, ahora en esta semana veremos el coeficiente de agrupamiento de cada nodo, que se calcula:

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

Donde,

- k_i : grado del nodo i , número de enlaces que los conecta con otros
- N : número de nodos de la red
- L_i : número de enlaces entre los vecinos del nodo i
- C_i : coeficiente de agrupamiento del nodo i .
 - $C_i =$
0 indica que ninguno de los vecinos de i están conectados entre sí
 - $C_i = 1$ indica que todos están conectados

En la red de amistad de los miembros del club de kárate de la figura 2, se muestra la información del coeficiente del clustering en la Tabla 2. Un valor de 1 indica que todos sus vecinos están conectados entre ellos, se puede ver que existen varios nodos donde se alcanza dicho valor. Además, se puede dividir diferente tipo de nodos, nodos con un valor de 1, cuyos adyacentes estarían totalmente conectados entre ellos, solamente el nodo 12 tendría valor 0, donde ninguno de sus vecinos estaría conectado y finalmente, nodos con valores entre 0.4 y 0.6 con conexiones intermedias entre sus adyacentes. Hay que ver que aquellos nodos que tenían un alto grado, como 1 y 34, tienen muchos enlaces que hace menos probable que existan conexiones entre todos sus vecinos al ser muchos.

Id	Label	Coeficiente Clustering	triangles
1	1	0,150	18
2	2	0,333	12
3	3	0,244	11
4	4	0,667	10
5	5	0,667	2
6	6	0,500	3
7	7	0,500	3
8	8	1,000	6
9	9	0,500	5
10	10	0,000	0
11	11	0,667	2
12	12	0,000	0
13	13	1,000	1
14	14	0,600	6
15	15	1,000	1
16	16	1,000	1
17	17	1,000	1
18	18	1,000	1
19	19	1,000	1
20	20	0,333	1
21	21	1,000	1
22	22	1,000	1
23	23	1,000	1
24	24	0,400	4
25	25	0,333	1
26	26	0,333	1
27	27	1,000	1
28	28	0,167	1
29	29	0,333	1
30	30	0,667	4
31	31	0,500	3
32	32	0,200	3
33	33	0,197	13
34	34	0,110	15

Tabla 2. Coeficientes de clustering (red de amistad club de Kárate, Figura 2)

3. Resumen

En esta lección se han estudiado alguno de los indicadores a nivel de nodo que nos permiten describir las conexiones y distancias entre los actores. En el ARS a nivel local se suele estudiar el vecindario inmediato: las diadas y las triadas en las cuales están inmersos. El grado de un actor, y el grado de entrada y de salida (si las relaciones son dirigidas) nos explica hasta qué punto un actor puede estar limitado o puede limitar a otros. Este conocimiento puede ser útil para describir las oportunidades de un actor. A partir de dicha información podemos describir *tipos* de actores de acuerdo con las posiciones que ocupan en la estructura: aislados, grandes emisores o grandes receptores.

Las conexiones locales de los actores son importantes para comprender la conducta social del conjunto de la población y también para entender a cada individuo. Tanto si los actores son accesibles por otros, como si las relaciones tienden a ser recíprocas o transitivas también es significativo para describir el conjunto de la población. Las poblaciones con muchas conexiones recíprocas y transitivas responden de forma diferente a los desafíos que aquéllas que tienen pocas conexiones o presentan una gran variedad de las conexiones que presenta cada uno de sus individuos. En estos últimos casos, se tienden a desarrollar más diferenciación y estratificación entre sus individuos.

Además, se ha introducido el análisis de las distancias entre pares de actores considerando tanto paseos, como senderos, como caminos. Las distancias nos proporcionan información valiosa acerca del modo en el que los individuos están inmersos en las redes, considerando un contexto más amplio que los vecindarios locales de los actores. Una distancia especialmente estudiada para analizar la cercanía entre los actores en el ARS es la distancia geodésica. La distancia geodésica media de un actor al resto, las variaciones en las distancias geodésicas de los diferentes actores, así como el número de distancias geodésicas entre cada par de actores puede ayudarnos a describir importantes similitudes y diferencias entre los actores y cómo de cerca están conectados la población en su conjunto. En este contexto, no solamente la distancia geodésica es relevante, sino también cualquier otra medida de distancia, nos permite conocer todas las posibilidades de comunicación (aunque no sean optimizadas) que existen entre los actores y la fortaleza de dichas conexiones.

Finalmente, el estudio del coeficiente de clustering en cada nodo nos proporciona información acerca de la conexión que existen entre sus nodos adyacentes y nos permite determinar si solamente dependen de nosotros para comunicarnos con ellos o no, lo que nos determina si tenemos más o menos influencia en la información que reciben.

Referencias

- [1] Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). *Analyzing social networks*. Sage.
- [2] Dey, N., Borah, S., Babo, R., & Ashour, A. S. (2018). *Social network analytics: computational research methods and techniques*. Academic Press.