

Teoría de juegos y expresión genética

Miguel Escribano Pérez

6 mayo 2020

Tema

La teoría de juegos ofrece herramientas que permiten analizar la importancia de genes en un sistema.

Estos análisis usan menos recursos que otros métodos estadísticos.

Contenidos

Teoría de juegos: qué y cómo

Análisis de expresión genética

Análisis de centralidad genética

Teoría de juegos

Estudio de los resultados de la interacción entre agentes que siguen una serie de normas.

Originado en economía; aplicado a sociología, política, biología, computación...

Muchos subtipos.

Elementos de un juego

- Jugadores
- Estrategias
- Información
- Ritmo
- Resultados

	J1 E1	J1 E2
J2 E1	R1	R2
J2 E2	R3	R4

Ejemplo de juego

Dilema del prisionero:

- Delata y sé libre
- Sé delatado y sufrirás

	B no delata	B delata
A no delata	-1, -1	-3, 0
A delata	0, -3	-2, -2

Juego cooperativo

Senado vota mociones por mayoría simple. Hay cuatro partidos: A tiene 3 votos, B 2 votos, C 1 voto y D 1 voto.

Coaliciones ganadoras: AB, AC, AD, ABC, ABD, ACD, BCD, ABCD

¿Cómo repartir los beneficios?

Valor Shapley

Mide el grado de responsabilidad de un jugador en el conjunto de los resultados posibles.

Para un jugador i , compara los resultados de las coaliciones según esté o no presente i .

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (n - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S))$$

Shapley en el senado

En ejemplo anterior, donde A tiene 3 votos, B 2 votos, C 1 voto y D 1 voto:

Shapley de A es 0,5

Shapley de B, C y D es 0,167

Número de votos y poder difieren.

Shapley en conjuntos de juegos

Senado: A 3 votos, B 2 votos, C 1 voto y D 1 voto

Congreso: A 2 votos, B 3 votos, C 1 voto y D 1 voto

Shapley global es la media de Shapley de cada juego.

Shapley	A	B	C	D
Senado	0,5	0,167	0,167	0,167
Congreso	0,167	0,5	0,167	0,167
Global	0,333	0,333	0,167	0,167

Expresión genética en microarray

La técnica de microarray permite conocer el nivel de expresión de un gen en una muestra.

Dadas una muestra normal y otra anómala se pueden seleccionar genes que se expresan de forma diferente.

Aplicando Shapley como en el ejemplo del senado y el congreso se obtienen los genes más relevantes de entre los expresados de forma anómala.

Ejemplo microarray

Dadas las muestras normales (MN), se calcula rango válido con media y desviación típica para cada gen (G):

	MN 1	MN 2	MN 3	MN 4	Rango válido
G1	0,4	0,2	0,3	0,6	[0,20 ; 0,55]
G2	12	10	4	5	[3,89 ; 11,61]
G3	8	13	20	9	[7,05 ; 17,95]
G4	0	-0,5	1,4	1,1	[-0,40 ; 1,40]

Expresiones genéticas anómalas

Dado el microarray de una muestra cancerígena (izq), se compara cada valor con el rango válido y se obtiene una tabla booleana de expresiones anómalas (der).

	M1	M2	M3
G1	0,9	0,4	0,7
G2	4,6	15	18
G3	7	21	12
G4	0,1	-0,4	1,6

	M1	M2	M3
G1	1	0	1
G2	0	1	1
G3	1	1	0
G4	0	0	1

Juegos en el microarray

Cada muestra es un juego y su coalición ganadora los genes expresados de forma anómala. Se puede calcular Shapley global sobre el conjunto de juegos.

El gen 3 es el más relevante para la aparición de este cáncer.

Juego	Coalición ganadora
M1	{G1, G3}
M2	{G2, G3}
M3	{G1, G2, G4}

Gen	Shapley
G1	$(\frac{1}{2} + 0 + \frac{1}{3}) / 3 = 0,278$
G2	$(0 + \frac{1}{2} + \frac{1}{3}) / 3 = 0,278$
G3	$(\frac{1}{2} + \frac{1}{2} + 0) / 3 = 0,333$
G4	$(0 + 0 + \frac{1}{3}) / 3 = 0,111$

Complejidad del juego

Si los rangos de validez están precalculados, para M muestras de G genes:

- Matriz de anomalías: $M \cdot G$ comparaciones
- Shapley parciales: M sumas
- Shapley global: G medias ($M \cdot G$ sumas + G divisiones)

Complejidad: $O(n^2)$ (peor caso de quicksort)

Centralidad en redes genéticas

Dados un conjunto de genes y sus relaciones de co-expresión, identificar los genes más relevantes para afectar a un subconjunto.

¿Cuáles son los circuitos genéticos que acaban activando o desactivando los genes responsables del cáncer?

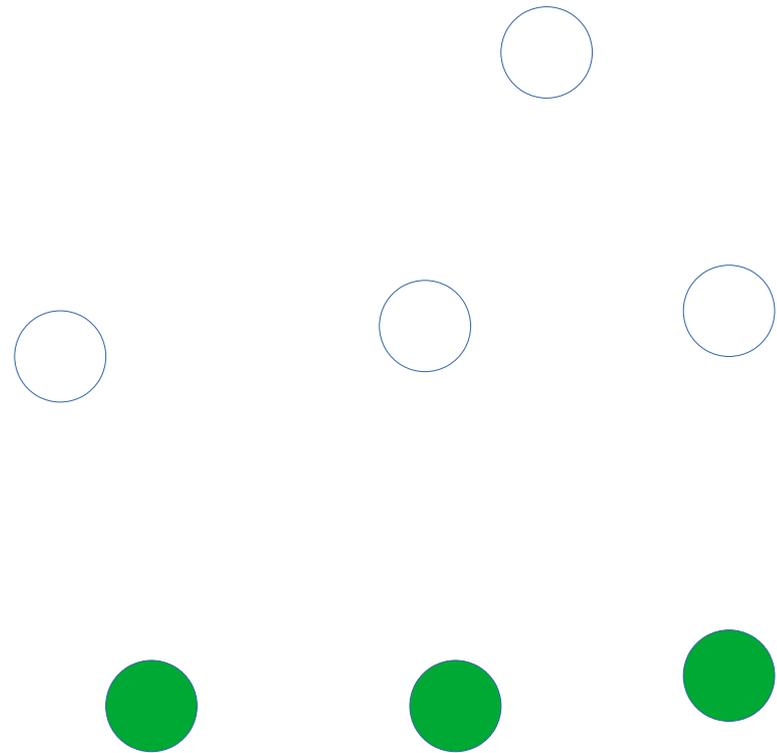
Ejemplo red genética

Conjunto de genes
analizados



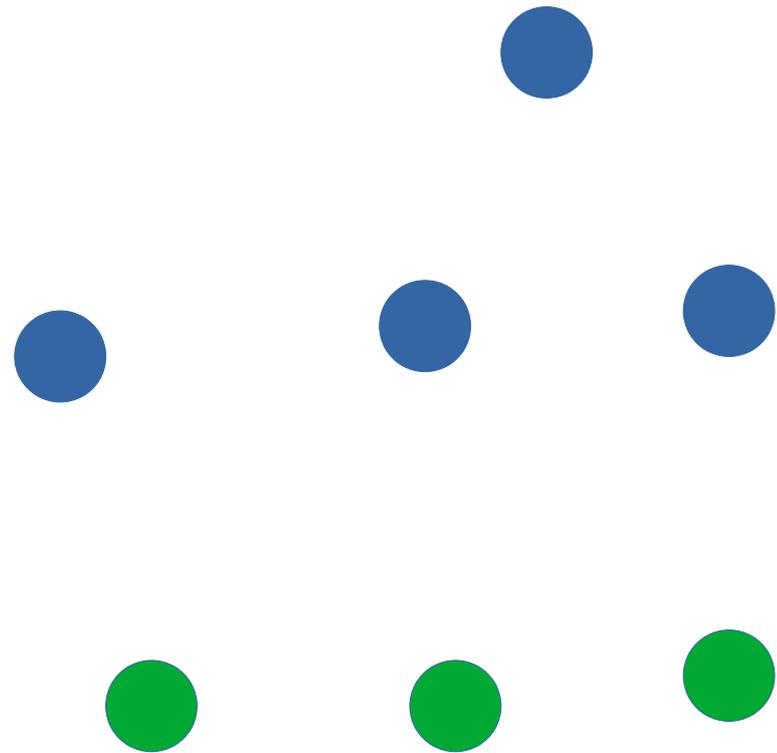
Ejemplo red genética

Subconjunto de genes
cuya expresión se quiere
influenciar



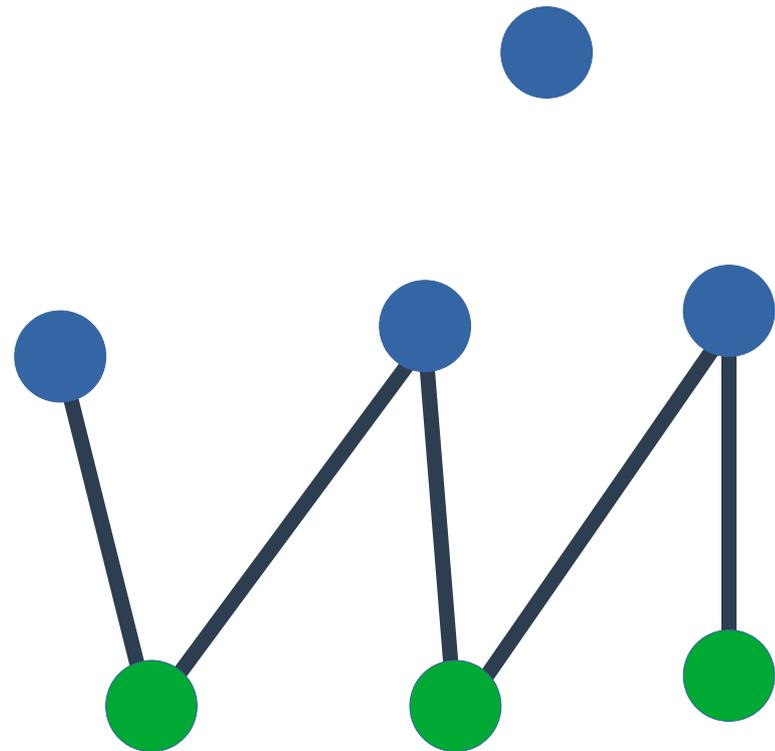
Ejemplo red genética

Resto de genes



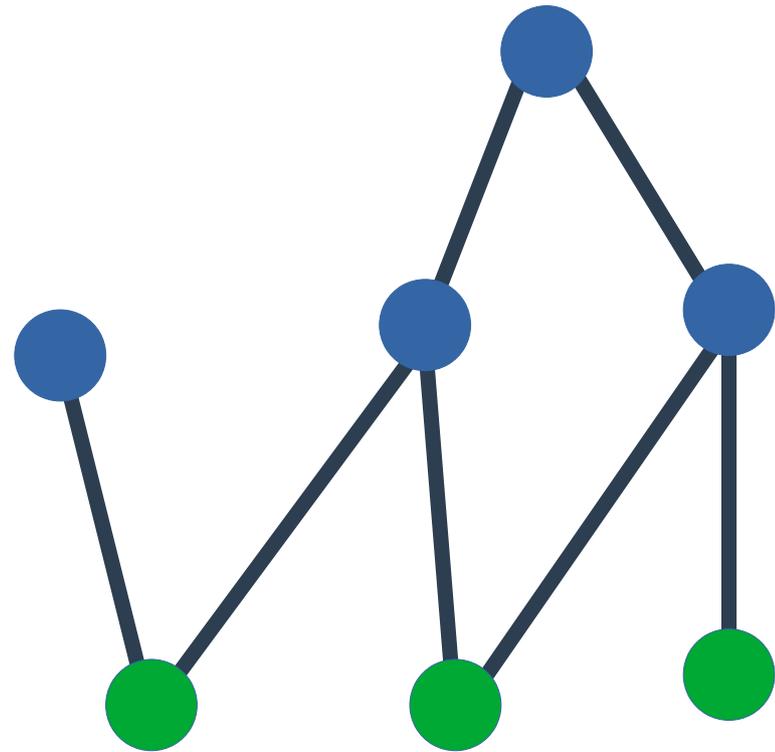
Ejemplo red genética

Relaciones de co-expresión de genes importantes



Ejemplo red genética

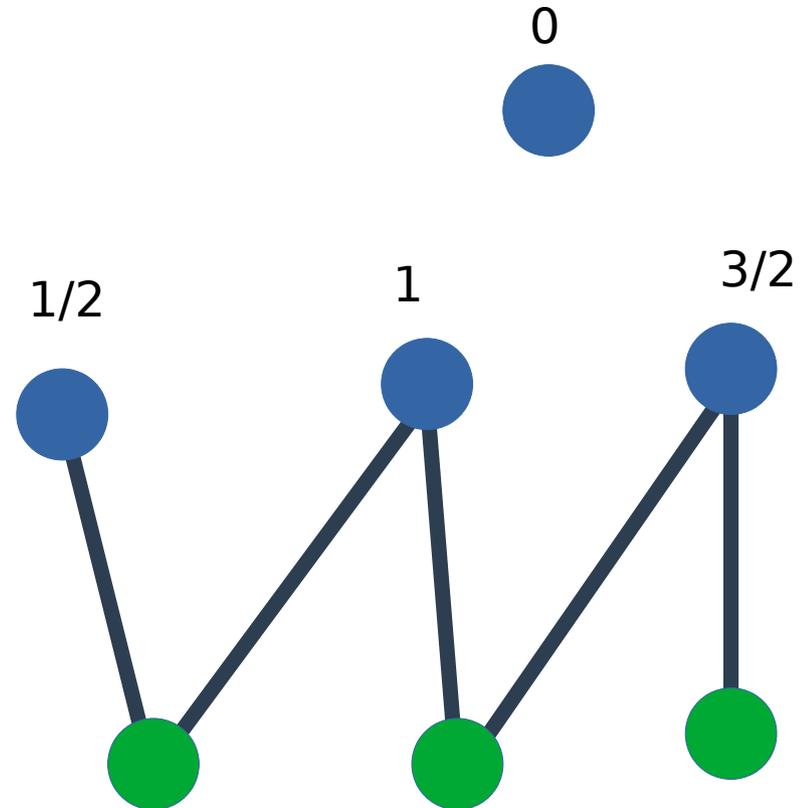
Relaciones de co-expresión
del resto de genes



Shapley red genética

Relaciones de co-expresión de genes importantes

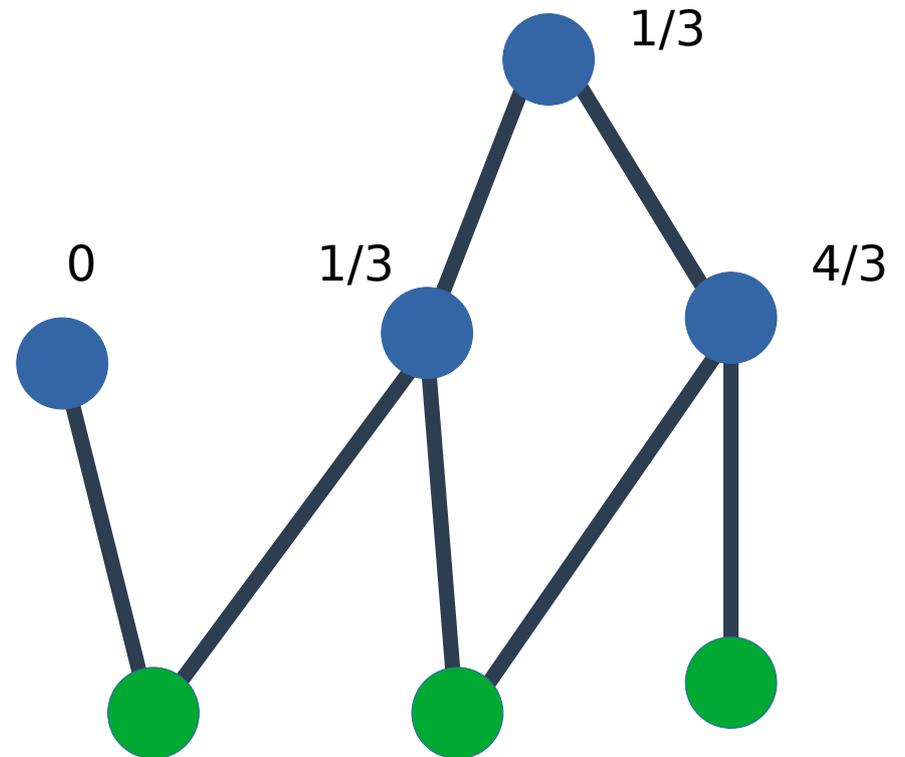
Shapley de conjunto de juegos (un juego por gen verde)



Shapley red genética

Relaciones de co-expresión del resto de genes

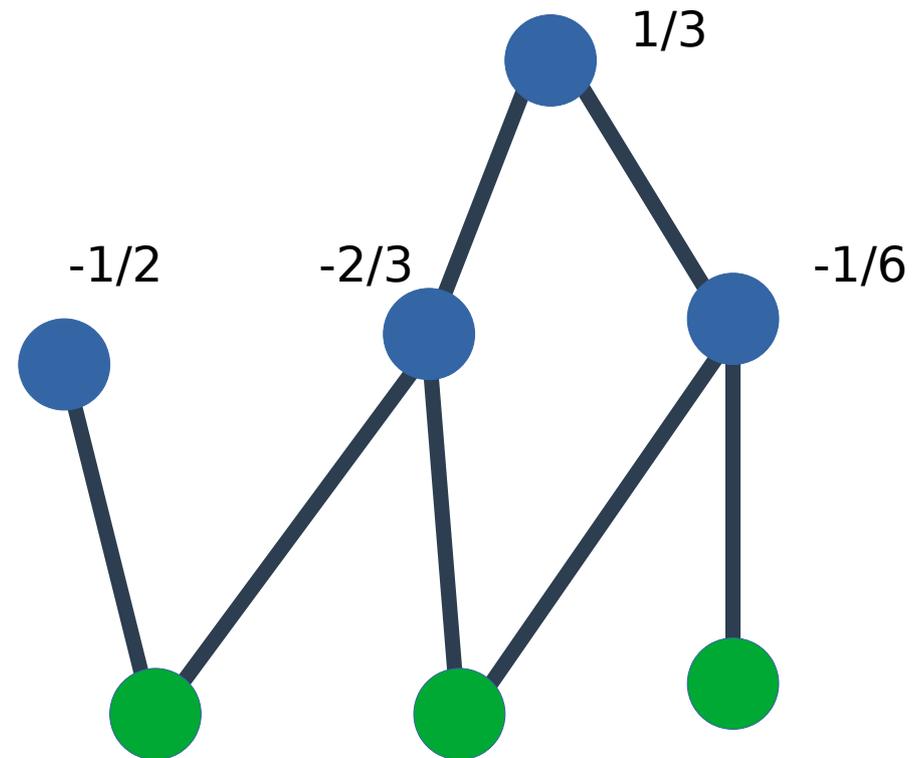
Shapley de conjunto de juegos (un juego por gen verde)



Centralidad red genética

Relaciones de co-expresión
del resto de genes

Medida de centralidad
(diferencia de los Shapley
previos)



Complejidad del juego

Muy alta, hay que calcular el bosque de relaciones de los genes.

Existen formas de aproximar el resultado calculando un subconjunto del problema.

Conclusión

La teoría de juegos ofrece herramientas muy potentes para identificar elementos de gran influencia en un conjunto.

El uso de Shapley o índices de centralidad puede ayudar a acotar un problema extenso a un subconjunto, mejorando la eficiencia de los estudios.

Referencias

Moretti S. (2006). *Game Theory applied to gene expression analysis*. PhD thesis
<https://www.lamsade.dauphine.fr/~moretti/contenuti/phd1.pdf>

Moretti S., Bonassi S. (2010). *Using coalitional games on biological networks to measure centrality and power genes*. ResearchGate
https://www.researchgate.net/publication/46149106_Using_coalitional_games_on_biological_networks_to_measure_centrality_and_power_of_genes

Cesari G., Algaba E., Moretti S., Nepomuceno J. (2018). *An application of the Shapley value to the analysis of co-expression networks*. SpringerOpen
<https://doi.org/10.1007/s41109-018-0095-y>

Wikipedia, Shapley value, https://en.wikipedia.org/wiki/Shapley_value