Financiado por la Unión Europea NextGenerationEU

Alicia Mayor Bal
presenta el resumen del trabajo

Introducing a new edge centrality measure: The connectivity rank index

Jin Zhou, Yanqi Zhang, Jun-an Lu, Guanrong Chen

IEEE Transactions on systems, man and cybernetics, 2757–2764, (2024).

15 de julio de 2024

SERIE DE RESÚMENES EN ESPAÑOL

Palabras clave: Autovalor, Autovector, Centralidad de enlaces, Conectividad algebráica, Red compleja.

Introducción

Encontrar los **nodos o enlaces más influyentes** dentro de una red es una tarea muy importante para el estudio de distintos campos. La centralidad de los nodos se centra en la influencia de los nodos por sí mismos mientras que la centralidad de los enlaces considera también a los vecinos.

La **centralidad de los enlace**s incluye la importancia tanto de los enlaces presentes como de los ausentes. Dependiendo del caso de estudio serán relevantes unos u otros. Si se trata de defender una red ante ataques, será relevante conocer los enlaces presentes más importantes para defenderlos, mientras que si se trata de acelerar la difusión en una red se requerirá identificar los enlaces ausentes más importantes para añadirlos a la red.

Este artíuclo investiga la centralidad de los enlaces para estas dos situaciones con una nueva medida, el **CRI** (Connectivity Rank Index o Índice de clasificación de conectividad en español). Esta medida cuantifica, de forma relativa, el aumento o disminución de la conectividad algebráica (caracterizada por el segundo autovalor más pequeño de la Laplaciana) producido por la adición o eliminación de un enlace.

Además, debido a la alta complejidad que presenta la búsqueda de la variación del segundo autovalor en redes de gran escala, en este trabajo se propone un algoritmo de aproximación de la conectividad algebráica que reduce su complejidad y como consecuencia simplifica el cálculo del CRI.

1. Índice de clasificación de conectividad (CRI)

En primer lugar, se define la variación del segundo autovalor más pequeño de la matriz Laplaciana, al eliminar o añadir un enlace ϵ_{ik} como sigue:

$$\Delta\lambda_{ij}^{-} = \lambda_2(L^{-}) - \lambda_2(L) \quad (1)$$

$$\Delta\lambda_{ij}^{+} = \lambda_2(L^{+}) - \lambda_2(L) \quad (2)$$

El CRI de un nodo presente y de un nodo ausente se definen, respectivamente:

$$I_{ij}^{-} = \frac{\Delta\lambda_{ij}^{-}}{|min_{\epsilon_{mn}\in G}\Delta\lambda_{mn}^{-}|} \quad (3)$$

$$I_{ij}^{+} = \frac{\Delta\lambda_{ij}^{+}}{|min_{\epsilon_{mn}\notin G}\Delta\lambda_{mn}^{+}|} \quad (4)$$

El índice CRI está normalizado, obteniendo valores entre -1 y 1. Cúanto más próximo sea el valor del CRI a 1 o a -1, más importante es el enlace presente o ausente, respectivamente.

2. Aproximación del CRI

Se obtienen cotas inferiores y superiores de la variación en el segundo autovalor más pequeño causado por adición o eliminación de enlaces:

$$\Delta\lambda_{ij}^{+} = \lambda(L^{+}) - \lambda_2(L) \leq (x_{2i}(L) - x_{2j}(L))^2 \quad (5)$$

$$|\Delta\lambda_{ij}^{-}| = \lambda(L) - \lambda_2(L^{-}) \geq (x_{2i}(L) - x_{2j}(L))^2 \quad (6)$$

Donde x_{2i} y x_{2j} son las componentes i y j del autovector unitario $x_2(L)$ correspondiente al segundo autovalor más pequeño de la matriz Laplaciana.

Al realizar la aproximación para redes de gran escala, con el objetivo de disminuír la complejidad del algoritmo utilizado, se aproximan los incrementos con sus cotas:

$$\Delta\lambda_{ij}^{+} = \lambda(L^{+}) - \lambda_2(L) = (x_{2i}(L) - x_{2j}(L))^2 \quad (7)$$

$$\Delta\lambda_{ij}^{-} = \lambda(L^{-}) - \lambda_2(L) = -(x_{2i}(L) - x_{2j}(L))^2 \quad (8)$$

Una vez establecidas estas aproximaciones, se propone la siguiente aproximación al CRI para redes de gran escala:

$$I_{ij}^{-} = \frac{-(x_{2i}(L) - x_{2j}(L))^2}{max_{\epsilon_{mn} \in G}(x_{2m}(L) - x_{2n}(L))^2} \tag{9}$$

$$I_{ij}^{+} = \frac{(x_{2i}(L) - x_{2j}(L))^2}{max_{\epsilon_{mn} \notin G}(x_{2m}(L) - x_{2n}(L))^2} \tag{10}$$

Si los denominadores son igual a 0, la medida también será tomada como igual a 0.

Bajo ciertas circunstancias, el autovector unitario correspondiente al segundo autovalor más pequeño de la Laplaciana puede no ser único. Por ello, se realiza una última aproximación de la medida que resulta en:

$$I_{ij}^{-} = \frac{min_{k=1,...,M}\{-(x_{2i}^{k}(L) - x_{1j}^{k}(L))^2\}}{max_{\epsilon_{mn} \in G} max_{k=1,...,M}\{(x_{2m}^{k}(L) - x_{2n}^{k}(L))^2\}} \tag{11}$$

$$I_{ij}^{+} = \frac{min_{k=1,...,M}\{(x_{2i}^{k}(L) - x_{1j}^{k}(L))^2\}}{max_{\epsilon_{mn} \notin G} mink = 1, ..., M\{(x_{2m}^{k}(L) - x_{2n}^{k}(L))^2\}} \tag{12}$$

3. Resultados

Con el objetivo de comparar los valores del CRI exacto y el CRI aproximado, se aplican los cálculos a diferentes redes.

3.1. Red con 13 enlaces

Se calculan los valores exacto y aproximado de I_{ij}^{-} para todos los posibles enlaces a eliminar de la red que unen dos nodos i y j. Además, también se calcula el ranking exacto y aproximado de los enlaces, según los valores del I_{ij}^{-}.

Se observa diferencia entre el valor aproximado y el exacto de I_{ij}^{-} debido al pequeño tamaño de la red. Sin embargo, el ranking aproximado se mantiene consistente con respecto al exacto.

Tras realizar más de 100 simulaciones distintas, aunque el ranking aproximado no es siempre igual al exacto, **ambos rankings mantienen una alta correlación**.

3.2. Red con 170 enlaces

Al incrementar el tamaño de la red, se observa que las centralidades las medidas del CRI exacto y el CRI aproximado están altamente correlacionadas, obteniendo **valores en ambas medidas muy aproximados**.

Sobre los valores obtenidos en esta red se realiza un análisis de correlación de Spearman, obteniendo coeficientes de 0.9980 y 0.9960 para el CRI exacto y aproximado, respectivamente. Esto explica la correlación lineal altamente positiva entre ambas medidas.

Por tanto, para redes de gran escala, es apropiado realizar el cálculo de conectividad basándose en la conectividad algebráica, aplicando la aproximación del CRI.

4. Conclusión

Este artículo propone un índice nuevo para medir la centralidad de los enlaces de una red: el CRI, el cual se normaliza en base a su efecto sobre el segundo autovalor más pequeño de la Laplaciana. En comparación con otros índices existentes, el CRI determina la **importancia tanto de un enlace presente como de uno ausente**.

Además, se ha propuesto una aproximación para el algoritmo, que es capaz de **reducir la complejidad** de $O(N^5)$ a $O(N^3)$, a la hora de calcular los CRIs de todos los enlaces de una red de tamaño N. Este algoritmo funciona correctamente aplicado a redes de gran escala, obteniendo una correlación alta entre los CRI exactos y aproximados de los enlaces de la red.

5. Valoración del documento original

Este artículo tiene una gran base matemática, sin embargo, todos los conceptos se explican de forma clara desde un primer momento por lo que se percibe la relación entre unos y otros para llegar a las aproximaciones finales del CRI.

La variedad de ejemplos presentada al final del artículo es de gran ayuda ya que en ellos se refleja todo el desarrollo matemático tanto en forma de valores como de gráficos.