Redes Complejas

30 de marzo de 2023

1. Matemáticas de las redes

En este capítulo se introducen los conceptos y herramientas teóricas cruciales para describir y entender las redes. La mayoría de estos conceptos vienen de la teoría de grafos.

1.1. Redes y su representación

Un *grafo* o una *red* es una colección de vertices (sitios) y lados (enlaces). Vamos a denotar el número de sitios en una red por n y el número de enlaces por m.

Una red que no tiene nodos enlazados consigo mismo, o más de un enlace entre dos nodos, se llama *red simple*.

1.2. Matriz de adyacencia

Hay distintas formas de representar matemáticamente una red. Consideremos una red no dirigida con n sitios etiquetados $1, \ldots, n$. Si denotamos un enlace entre dos sitios i, j como (i, j), entonces toda la red puede representarse dando el valor de n y la lista de todos los enlaces. Una descripción así es una *lista de enlaces*.

Una representación más conveniente es la *matriz de adyacencia*. La matriz de adyacencia $\mathbf{A}$ de un grafo simple es la matriz con elementos A_{ij} tal que

$$A_{ij} = \begin{cases} 1, & \text{si hay un enlace entre los sitios } i \text{ y } j, \\ 0, & \text{en otro caso.} \end{cases} \tag{1.1}$$

Para una matriz de un grafo simple, los elementos de la diagonal son cero. Además, esta matriz es simétrica[1].

1.3. Redes pesadas

Muchas de las redes que vamos a estudiar tiene una forma simple *on/off* en las conexiones entre sus vértices (1 o 0). Sin embargo, a veces es útil representar enlaces con un peso, o fuerza, o valor, mediante algún número real. Este tipo de redes pesadas se pueden representar dando a los elementos de la matriz de adyacencia valores iguales a los pesos de las conexiones.

1.4. Redes dirigidas

Una red dirigida o *digrafo* es una red en la que cada enlace tiene una dirección, apuntando desde un sitio hacia otro. Tales enlaces se llaman *enlaces dirigidos*, y pueden ser representados por líneas con flechas.

La matriz de adyacencia de una red dirigida tiene elementos de matriz

$$A_{ij} = \begin{cases} 1, & \text{si hay un enlace desde el sitio } j \text{ hacia } i, \\ 0, & \text{en otro caso.} \end{cases} \tag{1.2}$$

Nótese que la dirección va desde el segundo índice al primero. Esto es contraintuitivo, pero es conveniente matemáticamente, y es la convención utilizada en el libro. Esta matriz, en general, no es simétrica.

1.4.1. Cocitación y acoplamiento bibliográfico

> *Esto va a quedar para anotarlo más adelante. Pero habla sobre dos formas de convertir una red dirigida en una no dirigida.*

[1] Sin embargo, también sería posible representar multienlaces o autoenlaces mediante esta matriz.

1.4.2. Redes acíclicas dirigidas

Un *ciclo* en una red dirigida es un *loop* cerrado con enlaces en cada uno de los sitios apuntando en la misma dirección a lo largo del loop. Algunas redes dirigidas no tienen ciclos, y se las llama redes *acíclicas*. Aquéllas que sí tienen ciclos se llaman *cíclicas*.

Una red típica que no tiene ciclos es una red de citación. Un artículo sólo puede citar otro artículo que ya ha sido escrito. En una red de citación, los enlaces apuntan únicamente hacia atrás en el tiempo.

- En una red acíclica de n vértices, debe haber al menos un sitio en la red que sólo tiene enlaces entrantes, y no salientes.

- A partir de esto, se puede desarrollar un algoritmo que determina si una red tiene o no tiene ciclos.

- Existe un etiquetado de los sitios de una red acíclica tal que su matriz de adyacencia es estrictamente triangular (triangular con ceros en la diagonal).

- Todos los autovalores de una matriz de adyacencia son nulos sí y sólo si la red es acíclica .

1.5. Hipergrafos

En algunos tipos de redes los enlaces unen a más de dos sitios a la vez. Tales enlaces se conocen como hiperenlaces, y una red con hiperenlaces es un hipergrafo. Los hipergrafos pueden ser representados de una manera más conveniente, conocida como red bipartita

1.6. Redes bipartitas

La asociación o pertenencia de los sitios a grupos, representados por hiperenlaces en un hipergrafo, puede ser representada como una red bipartita. En estas redes hay dos tipos de enlaces, unos representando los originales, y otros representando los grupos a los que pertenecen.

El equivalente a la matriz de adyacencia en las redes bipartitas es una matriz rectangular llamada *matriz de incidencia*. Si n es el número de sitios de una red y g el número de grupos, entonces la matriz de incidencia $\mathbf{B}$ es una matriz $g \times n$ con elementos B_{ij} tales que

$$B_{ij} = \begin{cases} 1, & \text{si el vértice } j \text{ pertenece al grupo } i, \\ 0, & \text{en otro caso.} \end{cases} \tag{1.3}$$

Aunque una red bipartita puede dar la representación más completa de una red particular, a menudo es conveniente trabajar directamente con conexiones entre sitios del mismo tipo. A partir de la red bipartita, podemos inferir esas conexiones, creando una proyección de un sólo modo (*one-mode projection*).

As an example, consider again the case of the films and actors. We can perform a projection onto the actors alone by constructing the n-vertex network in which the vertices represent actors and two actors are connected by an edge if they have appeared together in a film. The corresponding one-mode projection onto the films would be the g-vertex network where the vertices represent films and two films are connected if they share a common actor. Figure 6.5 shows the two one-mode projections of a small bipartite network.

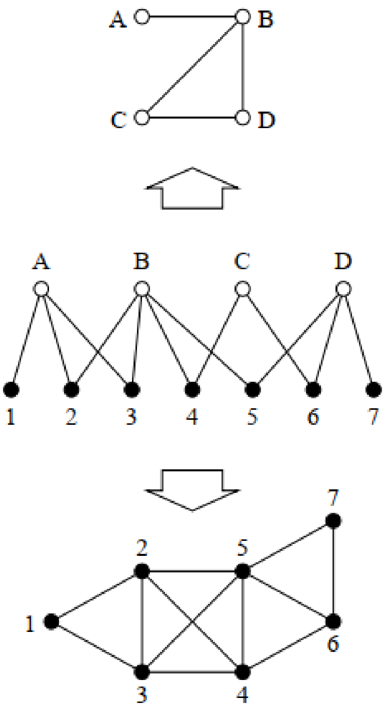

Figure 6.5: The two one-mode projections of a bipartite network. The central portion of this figure shows a bipartite network with four vertices of one type (open circles labeled A to D) and seven of another (filled circles, 1 to 7). At the top and bottom we show the one-mode projections of the network onto the two sets of vertices.

Con el producto $\mathbf{P} = \mathbf{B^T B}$ $(n \times n)$ o $\mathbf{P}' = \mathbf{BB^T}$ (que es una matriz $g \times g$) obtenemos las matrices de adyacencia de estas dos proyecciones de la red bipartita[2]

1.7. Árboles

Para anotar más adelante...

1.8. Redes planares

Para anotar más adelante...

1.9. Grado

El *grado* de un sitio en un grafo es el número de sitios conectados a él. Vamos a denotar el grado de un sitio i por k_i. Para un grafo no dirigido de n sitios el grado puede ser escrito en términos de la matriz de adyacencia como

$$k_i = \sum_{j=1}^{n} A_{ij} \tag{1.4}$$

De aquí se puede derivar que para una red con m enlaces, se tiene que

$$2m = \sum_{i=1}^{n} k_i \tag{1.5}$$

El grado medio c de un sitio en un grafo no dirigido es

$$c = \frac{1}{n}\sum_{i=1}^{n} k_i = \frac{2m}{n} \tag{1.6}$$

El número máximo de enlaces en un grafo simple es $\binom{n}{2} = \frac{1}{2}n(n-1)$. La densidad ρ de un grafo es la fracción de esos enlaces que está presente:

$$\rho = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)} = \frac{c}{n-1}. \tag{1.7}$$

[2]También podríamos hacer pesada la matriz $\mathbf{B}$ para capturar algo más de información de la red bipartita. Aún así, estas proyecciones no pueden capturar toda la información original.

La densidad cumple estrictamente $0 \leq \rho \leq 1$ [3].

- Se dice que una red es *densa* si ρ tiende a un valor constante para $n \to \infty$.
- Se dice que una red es *dispersa* si ρ tiende a cero para $n \to \infty$

Ocasionalmente nos vamos a cruzar redes en las que todos los sitios tienen el mismo grado. En teoría de grafos, estas son conocidas como redes regulares. Un grafo regular donde sus sitios tienen grado k se le dice k-regular (ej: red cuadrada infinita es 4-regular).

En una red dirigida, cada sitio tiene dos grados. El *in-degree* es el número de enlaces entrantes, y el *out-degree* es el número de enlaces salientes.

$$k_i^{\text{in}} = \sum_{j=1}^{n} A_{ij}, \qquad k_j^{\text{out}} = \sum_{i=1}^{n} A_{ij}. \tag{1.8}$$

Para grafos dirigidos, valen las siguientes relaciones:

$$m = \sum_{ij} A_{ij}. \tag{1.9}$$

$$c_{\text{in}} = c_{\text{out}}. \tag{1.10}$$

$$c = \frac{m}{n}. \tag{1.11}$$

1.10. Caminos

Un camino en una red es cualquier secuencia de sitios tal que cada par consecutivo de sitios en la secuencia está conectada por un enlace en la red. Es una ruta a lo largo de la red que va de un sitio a otro a través de los enlaces de la red.

La *longitud* de un camino en una red es el número de enlaces recorridos a lo largo del camino. Es relativamente sencillo calcular el número de caminos de longitud r en una red [4]. El número total $N_{ij}^{(r)}$ de caminos de longitud r desde j hasta i es

$$N_{ij}^{(r)} = [\mathbf{A}^r]_{ij}. \tag{1.12}$$

Desde este último resultado, se sigue que el número L_r de loops de longitud r en cualquier lugar de la red es

$$L_r = \operatorname{Tr} \mathbf{A}^{\mathbf{r}}. \tag{1.13}$$

1.10.1. Caminos geodésicos

Para anotar más adelante...

1.11. Componentes

Es posible que no exista un camino entre un par dado de sitios en una red. Cuando esto ocurre, se dice que la red está *desconectada*, o *conectada* cuando esto si ocurre.

Los subgrupos de una red que están conectados se llaman *componentes*. Técnicamente una componente es un subconjunto de sitios de una red tal que existe al menos un camino desde cada miembro de ese subconjunto a todos los otros miembros.

La matriz de adyacencia de una red con más de una componente se puede escribir en forma diagonal por bloques. Existen algoritmos que pueden agarrar una red con un etiquetado arbitrario de los índices e identificar rápidamente sus componentes, de forma tal que su matriz de adyacencia quede en forma diagonal por bloques.

1.11.1. Componentes en redes dirigidas

Para anotar más adelante...

1.12. Caminos independientes, conectividad, y conjuntos de corte

Para anotar más adelante...

[3] Cuando n es muy grande, vale bien la aproximación $\rho = c/n$.

[4] Hay una descripción medio inductiva en el libro, que me la voy a saltear.

1.13. Laplaciano de un grafo, random walks

Para anotar más adelante...

2. Medidas y métricas

2.1. Centralidad de grado

Un gran volumen de investigación en redes se ha dedicado a estudiar el concepto de centralidad. ¿Cuáles son los sitios más importantes en una red?

Quizá la medida más simple de centralidad es el grado de un sitio, el número de enlaces conectados a él. Es una medida simple pero poderosa.

2.2. Centralidad por proximidad

La medición de centralidad por proximidad (*closeness*) mide la distancia media de un sitio a otros sitios. Supongamos que d_{ij} es la longitud de un camino geodésico desde i hasta j. Entonces la distancia geodésica media de i hasta j, promediada sobre todos los sitios j de la red, es

$$l_i = \frac{1}{n} \sum_j d_{ij} \tag{2.1}$$

En la literatura de redes sociales, es común calcular la *closeness centrality* C_i:

$$C_i = \frac{1}{l_i} = \frac{n}{\sum d_{ij}} \tag{2.2}$$

Un problema que tiene esta medida de centralidad es que el rango dinámico de los valores para los sitios es a menudo muy pequeño, y es difícil identificar diferencias.

2.3. Centralidad betweenness

Un concepto diferente es la *centralidad betweenness*, que mide el grado en que un sitio está en caminos entre otros sitios. Los sitios con alta centralidad betweenness podrían tener influencia considerable dentro de una red en virtud de su control sobre la información que pasa entre otros.

Matemáticamente: consideremos el conjunto de todos los caminos geodésicos en una red no dirigida. Entonces la centralidad betweenness del sitio i es el número de esos caminos que pasan por i. Sea n^i_{st} 1 si el sitio i está en el camino geodésico entre s y t, o 0 si no existe ese camino. Entonces la centralidad betweenness x_i está dada por

$$x_i = \sum_{st} n^i_{st}. \tag{2.3}$$

Si hay más de un camino geodésico posible entre s y t, lo que puede hacerse es sumar cada contribución de forma fraccionaria, dividiendo por el nro. total de caminos geodésicos g_{st}:

$$x_i = \sum_{st} \frac{n^i_{st}}{g_{st}} \tag{2.4}$$