Concepto de pareo en un grafo

- Representación mediante grafos
- Modelo del problema

Un emparejamiento en G = (SUT, A) es un conjunto E de arcos tales que ningún par de ellos tiene un nodo en común.

Emparejamiento máximo o de cardinal máximo -> es el mayor número posible de arcos. NO todos los nodos están emparejados con otro, sino que están emparejados los máximos posibles.

Un nodo = expuesto; si no es punto extremo de ningún arco de E.

Emparejamiento completo -> es máximo y solo puede darse si los dos conjuntos tienen el mismo número de nodos.

Pareo de grafos

- Pareos perfectos
  - Un emparejamiento perfecto es una colección de arcos separados que cubren todos los nodos.
- Pareos maximales
  - Es un emparejamiento que no puede crecer oír agregado de un arco. Cumple la propiedad de que al añadir algún arco que no pertenece al emparejamiento, se deshace el emparejamiento.
  - Si G = (N, E) es un grafo bipartito con N dividido como X U Y, un emparejamiento maximal en G es aquel que relaciona el mayor numero posible de nodos en X con los nodos en Y.
  - Contiene el numero máximo posible de arcos, no puede estar contenido en otro de cardinal mayor.

Con lo anterior

- Puede haber muchos emparejamientos máximos.
- Los emparejamientos máximos son maximales, pero no todos los emparejamientos maximales son máximos.
- Cada emparejamiento perfecto es máximo y maximal

Obtener un emparejamiento máximo bipartito:

- Un camino alterno, camino en el cual sus arcos alternativamente pertenecen y no pertenecen al emparejamiento.
- Un camino incremento, camino alterno que comienza y termina en un nodo libre.

Para lograr esto, se ira incrementando paso a paso el cardinal del emparejamiento mediante caminos incremento hasta obtener un emparejamiento máximo.

Matrimonios estables

- Problema de asignación bipartito.
- Dos conjuntos de personas que tienen el mismo cardinal.

- Encontrar un emparejamiento estable entre ambos grupos de personas.
- Con el algoritmo de Gale-Shapley, siempre y cuando el numero de hombres y mujeres es el mismo, siempre existe un emparejamiento estable entre ellos.
- En términos de matrimonios, el concepto de pareja bloqueante se puede formular de la siguiente manera. Si hay dos matrimonios (hi, mi) y (hj, mj) tales que hi prefiere estar antes con mj que con mi, y mj prefiere estar antes con hi que con hj, entonces (hi, mi) es una pareja bloqueante, ya que tanto hi como mj se divorciaran de sus parejas para estar juntos.

Algoritmo (Gale-Shapley)

1. Cada mujer propone su primera elección.
2. Los hombres con dos o más propuestas responden rechazando a todas menos a la más favorable.
3. Las mujeres rechazadas proponen su segunda elección. Las que no fueron rechazadas continúan con su propuesta.
4. Se repiten los pasos 2 y 3 hasta que ninguna propuesta sea rechazada.

Grafo dirigido

- Es un triple ordenado D = (N(D), A(D) , d) que consiste en un conjunto no vacío de nodos N(D), un conjunto separado de arcos A(D), y una función de incidencia que se asocia con cada arco D con un par ordenado (no necesariamente distinto) de nodos D.

Al igual que con los grafos, los dígrafos tienen una representación grafica simple. Un digrafo se representa mediante un diagrama de su grafo subyacente junto con flechas en sus bordes, cada flecha apuntando hacia la cabeza del arco correspondiente.

Propiedades

- Un recorrido en D es una secuencia finita no nula W cuyos términos son nodos y arcos alternativamente, de manera que, para i = 1, 2, …, k, el arco ai tiene cabeza ui y cola ui-1. Al igual que con los paseos en los grafos, un recorrido a menudo se representa simplemente por su secuencia de nodo.
- Si hay una ruta (u, v) dirigida en D, se dice que el nodo v es accesible desde el nodo u en D.
- Dos nodos están conectados en D si cada uno es accesible desde el otro. Como en el caso de la conexión en grafos, la conexión dirigida es una relación de equivalencia en el conjunto de nodos de D.

Definición 1 Una red es un dígrafo con una fuente designada s y un destino $t \neq s$, además, cada arco e tiene una capacidad positiva, $c(e)$.
Las redes se pueden utilizar para modelar el transporte a través de una red física, de una cantidad física como el petróleo o la electricidad, o de algo más abstracto, como la información.

Definición 2 Un flujo en una red es una función f de los arcos del dígrafo a R, con $0 \leq f(e) \leq c(e)$ para todo e, y tal que:

$$\sum e \in E + vf(e) = \sum e \in E - vf(e)$$

para todo v excepto s y t.
Se desea asignar un valor a un flujo, igual al flujo neto que sale de la fuente, dado que la sustancia que se transporta no puede $recolectarse$ ni $originarse$ en ningún nodo que no sea s y t, parece razonable que este valor también sea el flujo neto hacia el objetivo.
Antes de probar esto, se introduce una nueva notación. Suponer que U es un conjunto de nodos en una red, con $s \in U$ y $t \notin U$.
Sean $\vec{U}$ el conjunto de arcos (v, w) con $v \in U$, $w \notin U$, y $\overleftarrow{U}$ el conjunto de arcos (v, w) con $v \notin U$, $w \in U$.

Definición 4 El valor de un flujo, denotado $val(f)$, es

$$\sum e \in E + sf(e) - \sum e \in E - sf(e)$$

Un flujo máximo en una red es cualquier flujo f cuyo valor es el máximo entre todos los flujos.
A continuación, se busca formalizar la noción de $cuello\ de\ botella$, a fin de demostrar que el caudal máximo es igual a la cantidad que puede pasar por el cuello de botella más pequeño.

Definición 5 Un **corte** en una red es un conjunto C de arcos con la propiedad de que cada camino de s a t usa un arco en C, es decir, si los arcos en C se eliminan de la red, no hay camino desde s a t.
La capacidad de un corte, denotada $c(C)$, es

$$\sum e \in Cc(e)$$

Un corte mínimo es uno con capacidad mínima, es mínimo si ningún corte está correctamente contenido en C.
Tener en cuenta que un corte mínimo es cuando U es un conjunto de nodos que contiene s pero no t, entonces $\overrightarrow{U}$ es un corte.

Lema 6 Suponer que C es un corte mínimo, entonces hay un conjunto U que contiene s pero no t tal que $C = \overrightarrow{U}$.

Prueba. Sea U el conjunto de nodos v tales que hay un camino de s a v sin arco en C.
Suponer que $e = (v, w) \in C$, dado que C es mínimo, hay un camino P de s a t usando e pero ningún otro arco en C.
Por lo tanto, hay un camino de s a v sin usar un arco de C, entonces $v \in U$.
Si hay un camino de s a w sin arco de C, entonces este camino seguido por la parte de P que comienza con w es un camino de s a t sin arco en C.
Esto implica que hay un camino de s sin usar ningún arco en C, lo que ocasiona una contradicción. Así $w \notin U$ y por lo tanto $e \in U$. Por lo tanto, $C \subseteq \overrightarrow{U}$.
Suponer que $e = (v, w) \in \overrightarrow{U}$, entonces $v \in U$ y $w \notin U$, por lo que cada camino de s a w usa un arco en C.
Dado que $v \in U$, hay un camino de s a v que no usa un arco de C, y este camino seguido por e es un camino de la s a la w. Por lo tanto, el arco e debe estar en C, por lo que $\overrightarrow{U} \subseteq C$.
Ahora hemos demostrado que $C = \overrightarrow{U}$.
Ahora podemos probar una versión del importante teorema de flujo máximo, corte mínimo.

Representaciones de un grafo dirigido

- D normalmente se representa en memoria por una representación enlazada, también llamada estructuras de adyacencia.
- Lista de nodos. Cada elemento de la lista de NODO corresponderá a un nodo de D y será un registro.
- Listas de arcos. Cada elemento de la lista ARCOS corresponde a un arco de D y será registro.

Algoritmos para grafos dirigidos. Algoritmo primero en Anchura y Primero en Profundidad para los grafos dirigidos

Algoritmo de Dijkstra

- Determina el camino más corto dado un nodo origen al resto de nodos en un grafo dirigido y con pesos en cada arco.
- Consiste en ir explorando todos los caminos más cortos que parten del nodo origen y que llevan a todos los demás nodos.

Algoritmo de Floyd-Warshall

- Intenta resolver el problema de encontrar el camino más corto entre todos los pares de nodos o nodos de un grafo.
- Esto es similar a construir una tabla con todas las distancias mínimas entre pares de ciudades de un mapa, indicando la ruta a seguir para ir de la primera ciudad a la segunda.

Algoritmo de búsqueda en anchura BFS

- Recorre o busca elementos en un grafo (usando frecuentemente sobre arboles).
- Se comienza en la raíz (eligiendo algún nodo como elemento raíz en el caso de un grafo) y se exploran todos los vecinos de este nodo.
- Formalmente, es un algoritmo de búsqueda sin información, que expande y examina todos los nodos de un árbol sistemáticamente para buscar una solución.
- No utiliza ninguna estrategia heurística.
- El peso de los arcos para ejecutar BFS debe de ser de igual costo. Si los arcos tienen pesos negativos se aplica el algoritmo de Bellman-Ford.
- Arco de retroceso: es el arco que va de un nodo a uno de sus antecesores. Un arco que va de un nodo hacia si mismo se considera un arco de retroceso.
- Arco cruzado: es el arco que va de un nodo a otro que no es ni antecesor ni descendiente.

Algoritmo de búsqueda en profundidad DFS

- Permite recorrer todos los nodos de un grafo o árbol de manera ordenada pero no uniforme.
- Consiste en ir expandiendo todos y cada uno de los nodos que va localizando, de forma recurrente, en un camino concreto.
- Cuando ya no quedan más nodos que visitar en dicho camino, regresa, de modo que repite el mismo proceso con cada uno de los hermanos del nodo ya procesado.
- Los arcos que llevan a nodos nuevos se conocen como arcos de árbol y forman un bosque abarcador en profundidad para el grafo dirigido dado.
- Arco de retroceso: va de un nodo a uno de sus antecesores. Un nodo que va de un nodo hacia si mismo se considera un arco de retroceso.
- Arco de avance: va de un nodo a uno de sus descendientes.
- Arco cruzado: va de un nodo a otro que no es ni antecesor ni descendiente.

Digrafos y lenguaje natural

Estándar. Cada palabra corresponde a la etiqueta de un nodo y si una palabra a inmediatamente precede a una palabra b en una sección s, entonces existe una arista que comienza en a y termina en b etiquetada con s. En este caso se tiene en cuenta la puntuación y las secciones propias de un documento como el titulo o resumen, entre otros.

Simple. A diferencia de la estándar, no se etiquetan las aristas con las secciones y no se tienen en cuenta todas las secciones sino aquellas que sean más visibles.

Distancia n. Se buscan las n palabras siguientes a partir de un término dado, y las aristas se etiquetan con la distancia al punto inicial. El parámetro n es definido por el usuario.

Distancia simple. Es similar a la anterior, con la diferencia de que las aristas no son etiquetadas y lo único que se sabe es que la distancia entre dos nodos conectados es menor que n.

Frecuencia absoluta. Es similar a la representación simple, pero cada nodo y arista son etiquetados con una medida de frecuencia. Para un nodo esto indica cuántas veces los términos aparecen en el documento; para las aristas esto significa el número de veces que dos términos conectados aparecen en el orden especifico. Bajo esta representación, el tamaño del grafo se define como la suma de las frecuencias de los nodos más la suma de las frecuencias de las aristas.

Frecuencia relativa. Es similar a la frecuencia absoluta con la diferencia de que las frecuencias son normalizadas entre 0 y 1. Los nodos se normalizan por el valor máximo de frecuencia de los nodos y las aristas similarmente por el máximo valor de frecuencia en las aristas.

Grafos conceptuales Un grafo conceptual es un grafo bipartito, esto es, que tiene dos tipos de nodos: conceptos y relaciones conceptuales, y cada arco une solamente a un concepto con una relación conceptual. Existen dos notaciones para los grafos conceptuales, la forma lineal (textual) y los diagramas o display form, que presentan tres tipos de elementos notacionales: rectángulos, que marcan los nodos de concepto, círculos, que marcan los nodos de relación, flechas direccionadas, que marcan los arcos:

La intención del uso de los grafos conceptuales para la representación del lenguaje natural es importante para la aplicación del esquema al análisis del lenguaje.

Los grafos conceptuales ponen el énfasis en la representación semántica (en el sentido lingüístico de la palabra), por ejemplo, tomar la siguiente expresión:

4.5.3. Grafo acíclico dirigido

Un grafo acíclico dirigido (DAG por sus siglas en inglés), es un grafo dirigido finito sin ciclos dirigidos.
Esto significa que sus nodos están conectados por arcos con una dirección específica y el recorrido de todo el grafo nos lleva de un punta A al B, sin tener la posibilidad de regresar al punto en A de ninguna forma
Aciclico significa que no hay bucles (es decir, *ciclos*) en el gráfico, de modo que para cualquier nodo dado, si sigue un arco que conecta ese nodo con otro, no hay camino en el gráfico para volver a ese nodo inicial.

Propiedades de los dagS

1. Tienen un punto de partida (origen) y un punto de llegada o final (sumidero). Al ser dirigidos, esto garantiza que su recorrido siempre vaya de un punto de origen a un punto final, y no se puede regresar sobre dicho recorrido.

2. La modificación de una relación entre nodos, reescribe la totalidad del DAG, debido a que su estructura y peso ha variado

3. Son paralelizables. Un DAG puede tener generación paralela y recorridos de diferente valor entre distintos nodos. Esto optimiza su generación y la capacidad de verificar la relación entre los nodos y la información que puedan contener.

4. Son reducibles. Una propiedad única de los DAG es que su estructura puede ser reducida a un punto óptimo en que su recorrido cumpla con todas las relaciones especificadas en el mismo sin ninguna pérdida. Esto significa que es posible reducir las relaciones de los nodos hasta un punto mínimo en que dicha reducción no afecta la capacidad de verificar la información de ningún nodo en ningún momento.

Un DAG también puede utilizarse para representar conceptualmente una serie de actividades.

El orden de las actividades se representa mediante un gráfico, que se presenta visualmente como un conjunto de círculos, cada uno de los cuales representa una actividad, algunos de los cuales están conectados por líneas, que representan el flujo de una actividad a otra, cada círculo es un $nodo$ y cada línea es un $arco$.
$Dirigido$ significa que cada arco tiene una dirección definida, por lo que representa un flujo direccional único de un nodo a otro.
Los DAG son útiles para representar muchos tipos diferentes de flujos, incluidos los flujos de procesamiento de datos.
Al pensar en los flujos de procesamiento a gran escala en términos de DAG, se pueden organizar los diversos pasos y el orden asociado para estos trabajos.