Complementi di Algoritmi e Strutture Dati

L'algoritmo di Karger per il taglio minimo in un grafo

Docente: *Nicolò Cesa-Bianchi* versione 18 aprile 2018

Un taglio in un grafo non orientato $G = (V, E)$ è l'insieme $\Gamma(S,T) \equiv \{(u,v) \in E : u \in S,\, v \in T\}$ dove S,T è una partizione ammissibile dei vertici (ammissibile significa che $S,T \not\equiv \emptyset$). In un grafo non pesato, il costo di un taglio corrisponde alla sua cardinalità $\bigl|\Gamma(S,T)\bigr|$. In questa nota consideriamo grafi non pesati e senza cappi (self-loop); tuttavia, ammettiamo la presenza di archi multipli fra coppie di nodi (useremo quindi il termine multigrafo invece di grafo). Il problema del taglio minimo (MinCut) in un multigrafo è definito nel modo seguente.

Problema **MinCut.**
Istanza: Un multigrafo $G = (V, E)$.
Soluzione: Una partizione S,T di V tale che $\bigl|\Gamma(S,T)\bigr|$ è minimo fra tutte le partizioni ammissibili di G.

Il problema MinCut ha tantissime applicazioni. Per esempio, in un sistema distribuito dove i nodi rappresentano processi e gli archi canali di comunicazione fra di essi, il taglio minimo corrisponde ad assegnare i processi a due CPU in modo che la comunicazione interprocessore —che è tipicamente lenta— sia minimizzata. Una seconda applicazione è la segmentazione di immagini. Qui i nodi rappresentano pixel e gli archi del grafo connettono pixel simili. Il taglio minimo corrisponde allora ad una segmentazione dell'immagine in due parti che sono fra loro il più dissimili possibile.

Il problema di MinCut è facilmente risolvibile in tempo polinomiale deterministico, per esempio usando l'algoritmo di Stoer–Wagner che ha un tempo di esecuzione dell'ordine $\mathcal{O}\bigl(|E||V| + |V|^2 \log|V|\bigr)$. Mostriamo ora un semplice algoritmo probabilistico Monte Carlo, l'algoritmo di Karger, che trova il taglio minimo con probabilità almeno $1 - \varepsilon$ in tempo pari a $\mathcal{O}\bigl(|E||V|^2 \log \frac{1}{\varepsilon}\bigr)$.

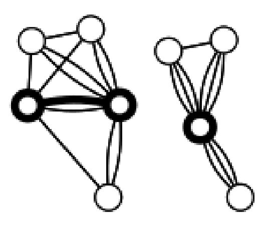

Figura 1: La contrazione di un arco in un multigrafo (da Wikipedia).

L'algoritmo di Karger è basato sull'operazione di contrazione di un arco (si veda la Figura 1). La contrazione di un arco (u,v) in un multigrafo G produce un multigrafo $G/(u,v)$ definito come:

1. Un nuovo vertice z è aggiunto al grafo.
2. Ogni arco $(w,u) \in E$ con $w \neq v$ è sostituito da un arco (w,z).
3. Ogni arco $(w,v) \in E$ con $w \neq u$ è sostituito da un arco (w,z).
4. Gli archi del tipo (u,v) e i vertici u,v sono rimossi.

Nel seguito, diciamo che z è un supervertice che contiene u e v. Quando uno o entrambi i due vertici agli estremi di un arco contratto sono a loro volta supervertici, allora i nodi in essi contenuti diventano parte del nuovo supervertice.

Introduciamo ora l'algoritmo di Karger "base" che ripetutamente contrae archi a caso del multigrafo fino a quando il numero di vertici rimanenti è pari a due. Dato che la contrazione di un arco riduce di uno il numero di vertici, l'algoritmo si fermerà esattamente dopo $|V| - 2$ passi. A questo punto, l'algoritmo produce il cut corrispondente all'unica partizione ammissibile dei due vertici. Dato che i vertici del multigrafo finale corrispondono ad una partizione dei vertici del multigrafo iniziale, abbiamo ottenuto un taglio del multigrafo iniziale.

Algorithm 1 (Karger base)

Input: Multigrafo $G = (V, E)$

while $|V| \geq 2$ **do**
    Scegli un arco a caso $(u, v) \in E$
    $G \leftarrow G/(u, v)$
end while
Output l'unico taglio rimasto in G

Karger base può essere implementato in tempo $\mathcal{O}(|E|)$ rappresentando la sequenza di contrazioni tramite una permutazione causale degli archi di G (dettagli omessi).

Vediamo ora qual è la probabilità che Karger base ritorni un taglio Γ^* di costo minimo k. Per prima cosa, si noti che Karger base ritorna Γ^* se e solo se nessun arco nel taglio viene contratto. La probabilità che il primo arco contratto X_1 sia nel taglio è pari a

$$\mathbb{P}(X_1 \in \Gamma^*) = \frac{k}{|E|} \leq \frac{k}{|V|k/2} = \frac{2}{|V|}$$

dove $|E| \geq |V|k/2$ perché $|E|$ è sicuramente più grande di $|V|d_{\min}/2$, dove $d_{\min}$ è il grado minimo dei nodi del grafo, e $d_{\min} \geq k$ perché altrimenti il costo del taglio minimo sarebbe $d_{\min}$.

Quindi, $\mathbb{P}(X_1 \notin \Gamma^*) \geq 1 - \frac{2}{|V|}$ e la probabilità che il secondo arco contratto X_2 non sia nel taglio, dato $X_1 \notin \Gamma^*$, è

$$\mathbb{P}\big(X_2 \notin \Gamma^* \mid X_1 \notin \Gamma^*\big) \geq 1 - \frac{k}{(|V| - 1)k/2} = 1 - \frac{2}{(|V| - 1)}$$

In generale, indicando con A_i l'evento $X_i \notin \Gamma^*$ per $i \geq 2$,

$$\mathbb{P}\big(A_i \mid A_1, \ldots, A_{i-1}\big) \geq 1 - \frac{2}{(|V| - i + 1)}$$

indicando convenzionalmente $\mathbb{P}(A_1 \mid X_0) = \mathbb{P}(A_1)$ possiamo scrivere

$$\mathbb{P}\big(\text{Krager base restituisce } \Gamma^*\big) = \mathbb{P}\left(\bigcap_{i=1}^{|V|-2} A_i\right) = \prod_{i=1}^{|V|-2} \mathbb{P}\big(A_i \mid A_0, \ldots, A_{i-1}\big)$$

$$\geq \prod_{i=0}^{|V|-3} \left(1 - \frac{2}{|V| - i}\right) = \prod_{i=0}^{|V|-3} \frac{|V| - i - 2}{|V| - i} = \frac{\prod_{i=1}^{|V|-2} i}{\prod_{j=3}^{|V|} j} = \frac{(|V| - 2)!}{\frac{|V|!}{2!}} = \frac{1}{\binom{|V|}{2}}$$

Se ripetiamo Karger base per $T = \binom{|V|}{2} \ln \frac{1}{\varepsilon} = \mathcal{O}\big(|V|^2 \ln \frac{1}{\varepsilon}\big)$ volte e scegliamo il taglio di costo minimo fra quelli prodotti la probabilità che questo non sia Γ^* è al più

$$\left(1 - \frac{1}{\binom{|V|}{2}}\right)^T \leq e^{-\ln \frac{1}{\varepsilon}} \leq \varepsilon$$

dove abbiamo usato la maggiorazione $1 - x \leq e^{-x}$ per ogni $x \in \mathbb{R}$. Il tempo totale di esecuzione, considerando anche il costo di ciascuna contrazione, è quindi $\mathcal{O}\big(|E||V|^2 \ln \frac{1}{\varepsilon}\big)$.

Una versione più sofisticata, nota come algoritmo di Karger-Stein, trova un taglio di costo minimo con probabilità almeno $1 - \varepsilon$ in tempo $\mathcal{O}\big((|V| \ln |V|)^2 \ln \frac{1}{\varepsilon}\big)$.