

PRL 2. projekt – Preorder

Tomáš Kantor (xkanto14)

1. května 2022

1. Rozbor a analýzu algoritmu

Algoritmus, který vypočte preorder pořadí symbolů na vstupu se skládá z těchto částí:

1. Vytvoření seznamu sousedů.
2. Výpočet Eulerovy cesty stromem. Každé orientované hraně stromu je přiřazena další hrana, která na ni navazuje, a to tak, že celkově tvoří kružnici, která prochází všemi vrcholy.
3. Zavedení kořene. Hrana, která vede do kořene stromu je dostane jako následníka samu sebe. Toto z kružnice vytvoří seznam hran, které začínají a končí v kořeni stromu.
4. Výpočet suffix sum. Vstupem je Eulerova cesta z předchozího kroku. Počáteční váhy pro jednotlivé hrany jsou dány takto: 1 pro dopředné hrany a 0 pro zpětné hrany.
5. Přiřazení preorder pořadí vrcholům.

1.1. Odvození časové složitost

Jádro algoritmu, které má na vstupu seznam sousedů a na výstupu preorder pořadí vrcholů stromů má následující složitost:

1. Výpočet Eulerovy cesty stromem má konstantní časovou složitost, protože neobsahuje, žádné cykly. $O(1)$.
2. Zavedení kořene má konstantní složitost. $O(1)$
3. Výpočet suffix sum. Každý procesor provede $\log_2(2n-2)$ iterací vnitřního cyklu. $O(\log(n))$
4. Přiřazení preorder pořadí vrcholům. Je provedeno v konstantním čase. $O(1)$

Složitost jádra algoritmu je $\log(n)$.

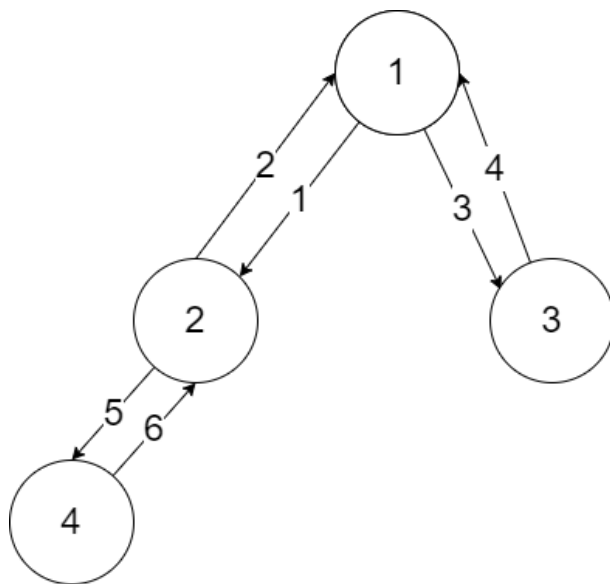
1.2. Odvození počtu procesorů

Každý symbol na vstupu algoritmu představuje jeden vrchol stromu, jejich počet je označen n . Pro každou orientovanou hranu ve stromu je přiřazen procesor pro její zpracování. Z teorie grafů vyplývá, že počet neorientovaných hran ve stromu je $n - 1$, kde n je počet vrcholů. Na každou neorientovanou hranu připadají 2 orientované, celkově jich je $2n - 2$. Z toho vyplývá, že $p(n) = 2n - 2$.

2. Implementace

Seznam sousedů je implementován jako pole struktur, které obsahují atributy `edge_num`, `reverse_edge_num`, `from_vertex`, `to_vertex`, `next`. Atribut `next` obsahuje číslo další hrany vedoucí ze stejného vrcholu. Seznam sousedů je indexován podle čísla hrany, takže je možné v konstantním čase nalézt jakoukoliv hranu. Pro nalezení hran vedoucích z vrcholu je použito další pole, které obsahuje číslo první hrany v seznamu sousedů pro každý vrchol. Obě pole jsou vytvořeny ve funkci `create_adj_list`.

Pojmenování orientovaných hran se řídí podle následujícího vzorce: hrana, která vede z rodiče do vrcholu k má číslo $2k - 3$. Reverzní hrana je pak $2k - 2$. Toto číslování má tu výhodu, že dopředné hrany jsou vždy liché a zpětné sudé. Také je velmi snadné zjistit název hrany pro každý vrchol.



Obrázek 1: Pojmenování hran stromu

Každý procesor pro svou hranu vypočte následující hranu v Eulerově cestě voláním funkce `compute_etour`.

Procesory vypočtou hodnotu suffix sum ve funkci `suffix_sum_forward_edges`. Synchronizace hodnot polí `value` a `succ` je provedena pomocí funkce `MPI_Allgather`.

Procesory dopředných hran pak vypočtou hodnotu preorder pro vrchol, do kterého daná hrana vede. Tato informace je pak zasláno procesoru č. 0, který vypíše hodnoty přiřazené vrcholům (hodnoty na vstupu programu) ve správném pořadí.