

Metody zlepšenia lokalnosti weboveho grafu

Report za zimný semester

Mojím cieľom v zimnom semestri bolo získať dostatočne hlboké porozumenie témat, ktorých zvládnutie by mohlo byť potenciálne potrebné na vytvorenie programu schopného dosiahnuť ciele stanovené v tomto projekte.

Konkrétnie išlo o nasledujúce oblasti:

zhlukovanie grafov prostredníctvom analýzy vlastných párov ich maticovej reprezentácie[2]; techniky redukcie dimenzie (napríklad rSVD), ako aj prahovanie „šumu“ vo veľkých dátových množinách[4, 5].

Ked'že obe tieto témy a metódy, ktoré skúmajú, vyžadujú určitý druh výpočtov zahŕňajúcich maticovo-vektorové operácie, rozhodol som sa implementovať takýto framework „od základov“, prispôsobený špecifickým cieľom tohto projektu. Na tento účel bola študovaná numerická lineárna algebra podľa knihy *Numerical Linear Algebra* [1]. Na základe toho boli na implementáciu zvolené známe stabilné a späťne stabilné algoritmy, ako napríklad Lanczosova iteračná metóda v kombinácii s posunutou QR faktORIZÁCIOU ako riešič vlastných hodnôt, a randomizovaná SVD na účely výpočtu singulárneho rozkladu.

S cieľom minimalizovať nadbytočné výpočty boli študované špeciálne štruktúry matíc zapojených do výpočtov realizovaných v projekte, napríklad možnosť redukcie veľkosti Householderových reflektorov pri výpočte QR rozkladu tridiagonálnej matice (v aktuálnom výpočtovom reťazci by sa QR faktorizácia vykonávala výlučne na tridiagonálnych maticiach).

Na účely implementácie uvedených algoritmov boli študované aj intrinsics pre registre AVX, ked'že výpočtovo náročné operácie by boli realizované práve ich pomocou, ako aj systémy merania efektivity paralelných algoritmov a možnosti paralelnej implementácie spomínaných metód.

Zvyšok voľného času som využil na začatie implementácie QR algoritmu.

To je všetko, čo sa mi podarilo zrealizovať počas zimného semestra.

[1] - Lloyd N. Trefethen, David Bau - Numerical Linear Algebra

[2] - Andrew Y. Ng - On Spectral Clustering: Analysis and an algorithm

[4] - Matan Gavish and David L. Donoho - The Optimal Hard Threshold for Singular Values is $4/\sqrt{3}$

[5] - Steven L. Brunton, J. Nathan Kutz - Data Driven Science & Engineering