

## 1 Definição do problema

**Triângulos** são conjuntos de 3 vértices de um grafo tal que, cada possível aresta entre eles, está presente.

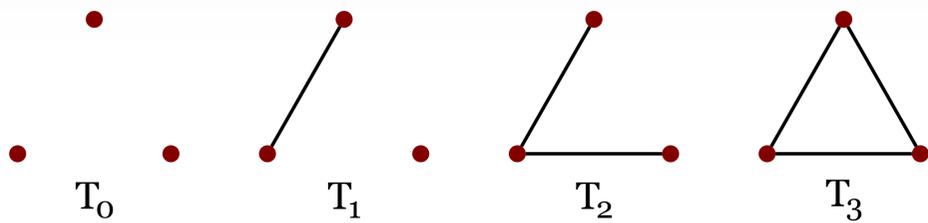


Figura 1: As possibilidades de relação entre três vértices em um grafo.  $T_3$  representa um triângulo.

Os problemas de triângulos são: encontrar, contar e listar triângulos.

**Grafos massivos** são estruturas de dados que representam grafos e não podem ser armazenadas em memória principal de tamanho típico, pois excedem a sua capacidade.

**Algoritmos de Memória Externa** utilizam, tipicamente, um disco rígido, como mídia de armazenamento primário, e a memória principal (RAM), como uma cache desses dados. Todo o processamento é feito em memória principal, porém, as trocas de dados com o HD devem ser minimizadas, pois são custosas devido a sua baixa taxa de transferência — como ilustrado na Figura 2.

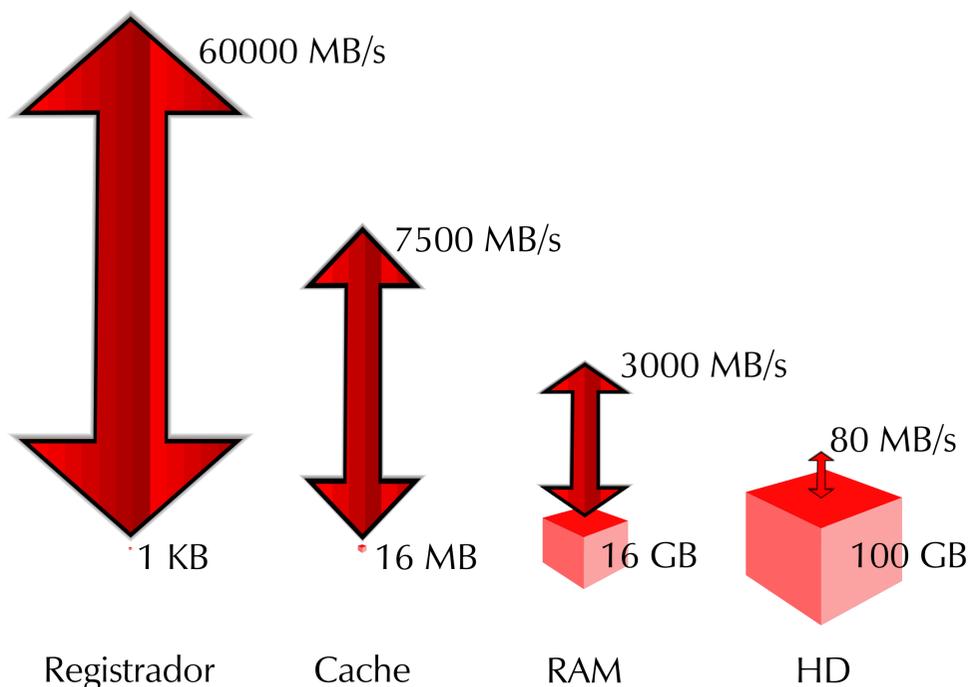


Figura 2: Comparação de diversos níveis de memória. Mídias de maior capacidade possuem taxas de transferência menores.

## 2 Aplicações

Alguns usos do número e da lista de triângulos:

- Análise de redes complexas
- Cálculo de coeficiente de clustering
- Cálculo de coeficiente de transitividade
- Encontrar subgrafos específicos
- Detecção de comunidades, intrusão e spam

## 3 Algoritmo proposto

Em algoritmos de memória externa, o objetivo é reduzir o número de trocas (I/Os) entre memória principal e HD. Devido a esse tempo ser ordens de grandeza maior que o tempo de processamento, analisa-se um algoritmo desses através de sua **complexidade de I/O**. Essa complexidade, para o algoritmo que propusemos, é

$$O\left(\text{Scan}\left(m^{\frac{3}{2}}\right)\right).$$

## 4 Resultados

Tabela 1: Resultados de tempo e transferência total para diversas instâncias testadas. O nosso algoritmo (EMCF) foi comparado ao EMVI, proposto anteriormente por Dementiev (2006).

Instância	Arestas	Tempo		I/O (GiB)	
		EMCF	EMVI	EMCF	EMVI
ip	1.744.214	10,5 s	27,6 s	0,5	1,9
notre-dame	1.935.518	8,5 s	7,9 s	0,5	0,5
actor-2002	15.038.083	2 min 4 s	4 min 36 s	5,1	20,5
p2p	159.870.973	44 min	2 hrs 20 min	116,2	467,7
uk-2005	822.487.051	9 hrs 25 min	6 hrs 9 min	1.939,9	1.296,8
sk-2005	1.810.063.330	22 hrs 44 min	24 hrs 17 min	5.389,3	4.504,8
uk-2007-01	3.433.757.890	62 hrs 26 min		11.642,1	

## 5 Conclusões

- Nós propusemos um novo algoritmo para listagem de triângulos chamado *External Memory Compact-Forward* (EMCF).
- O algoritmo proposto possui a complexidade de I/O melhor do que o algoritmo EMVI, proposto anteriormente por Dementiev (2006).
- Até então, os resultados do EMCF são melhores, em certos casos, e promissores nos demais.
- Ainda é preciso definir uma melhor heurística para iteração de arestas (utilizada internamente no algoritmo).