

REPLICAÇÃO DO ESTUDO DE ESCABILIDADE PARALELA DO ALGORITMO *INFOMAP* PARA DETECCÃO DE COMUNIDADES

Dhruv Babani, Bernardo Balzan,
Eduardo Cardoso

Introdução



O trabalho apresenta um comportamento de análise de desempenho de algoritmos na detecção de comunidades no modelo da memória compartilhada, com o intuito de ter uma visão analítica sobre os limites dos algoritmos propostos.



São realizadas análises significantes, a partir da execução dos algoritmos em grafos de diferentes tamanhos (nodos e arestas) e níveis de complexidade



Pode elevar a quantidade de recursos utilizados e a demanda ao passar do tempo, por conta das complicações que o paralelismo pode trazer



Grafos

- Visão Tradicional

- Estrutura fraca: média do grau interno dos vértices é maior do que a média do grau externo;
- Estrutura forte: os vértices apresentam grau interno maior do que o grau externo.

- Visão Moderna

- Estrutura fraca: média da probabilidade de cada vértice se conectar com outro vértice é maior do que a média da probabilidade de se conectarem com vértices externos;
 - Estrutura forte: os vértices possuem uma probabilidade maior de estarem conectados entre si do que com vértices externos.
-

Categorias de Métodos de Detecção



- Inferência estatística: descrevem modelos generativos para os grafos, tendo o MDL (Minimum Description Length) como o mais utilizado;



- Otimização: procuram maximizar funções, como a função de modularidade das comunidades presentes em um grafo (ex.: Algoritmo de Louvain);



- Simulação de dinâmica: utilizam simulações na estrutura dos grafos, utilizando difusão, caminhamento (ex.: *Infomap*), sincronização, entre outros métodos.

Paralelismo e a sua Escabilidade

- EXECUÇÃO SEQUENCIAL:

- O resultado de uma tarefa é comunicado a outra tarefa;
- Sincronização implícita, em que uma tarefa só é executada após o término da outra.

- EXECUÇÃO CONCORRENTE:

- Área de memória compartilhada acessada por vários fluxos de execução
 - Responsabilidade do usuário fazer uso de mecanismos de sincronização para garantir a correta comunicação.
-

Experimento Original

Os experimentos presente no estudo são realizados em apenas uma máquina multicore, seguindo seguintes parâmetros para realização da análise:

- Tempo de execução do algoritmo;
- Escalabilidade: tempo de execução em relação crescimento do grafo;
- Consumo de recursos: consumo de memória e tempo gasto durante a troca de mensagens em algoritmos paralelos
- Qualidade das comunidades: utilização de grafos reais.

Além disso, foram utilizados grafos com tempo de execução em minutos e com o processamento baseado em 6, 12, 24 e 48 threads. Visando uma análise mais comparativa entre os algoritmos abordados, simulando uma estratégia de memória compartilhada.

Seleção dos Grafos

Grafos	Nodes	Edges
Live Journal	4, 847, 571	68, 993, 773
Pokec	1, 632, 803	30, 622, 564

Algoritmos de Detecção de Comunidades

INFOMAP

- *Infomap* é um algoritmo sequencial amplamente utilizado para o cálculo da estrutura da comunidade de um gráfico. Existe uma gama diversificada de aplicações que o utilizam.
- Possui um número extremamente esparsa de implementações paralelas e distribuídas atualmente disponíveis

RELAXMAP

- Para a etapa de agrupamento, os cálculos realizados para decidir se um nó deve mudar de comunidade são feitos em paralelo
 - Threads movam nós livremente pode resultar em movimentos cíclicos de nós entre comunidades vizinhas, impactando negativamente o MDL
 - Para evitar movimentos que impactem as mesmas comunidades ao mesmo tempo, é utilizada uma secção crítica. Cada vez que um thread encontra um movimento válido, ele tentará acessar a secção crítica e, uma vez dentro, primeiro verificará se o movimento ainda é válido e então prosseguirá com o movimento.
-

Formulas usadas para os resultados

- $\text{Speed Up} = \text{Tempo paralelo} / \text{Tempo sequencial}$
- $\text{Eficiência} = \text{Speed Up} / \text{Threads}$



Resultados Comparativos – Original X Replicação(Tabelas)

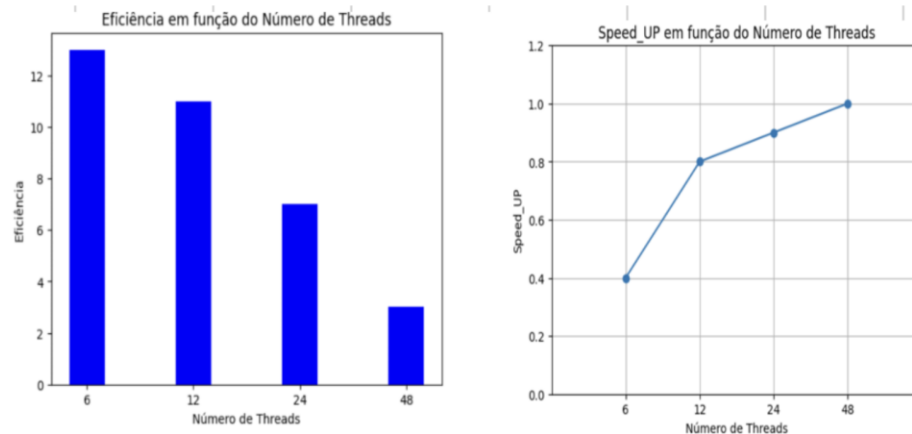
Grafo	INFOMAP(S)	RELAXMAP(S)
Live Journal	1270	90
Pokec	631.57	41.11

Grafo	INFOMAP(S)	RELAXMAP(S)
Live Journal	1150	75
Pokec	629.78	40.76

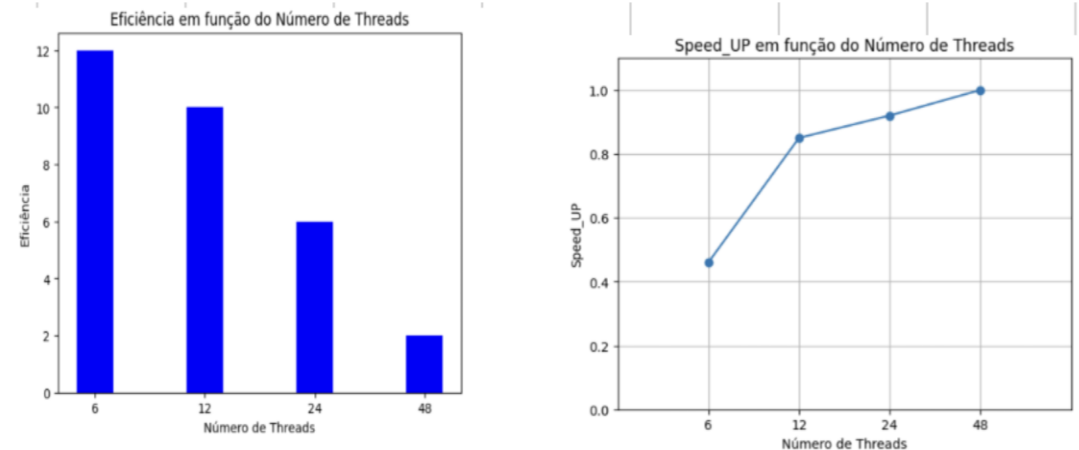
Resultados Comparativos – Original X Replicação(Gráficos)

- Representação grafica da comparação de resultados de *SpeedUp* e *eficiência* dos grafos no algoritmo *RelaxMap* entre este estudo e o original

- Estudo Original - Live jornal



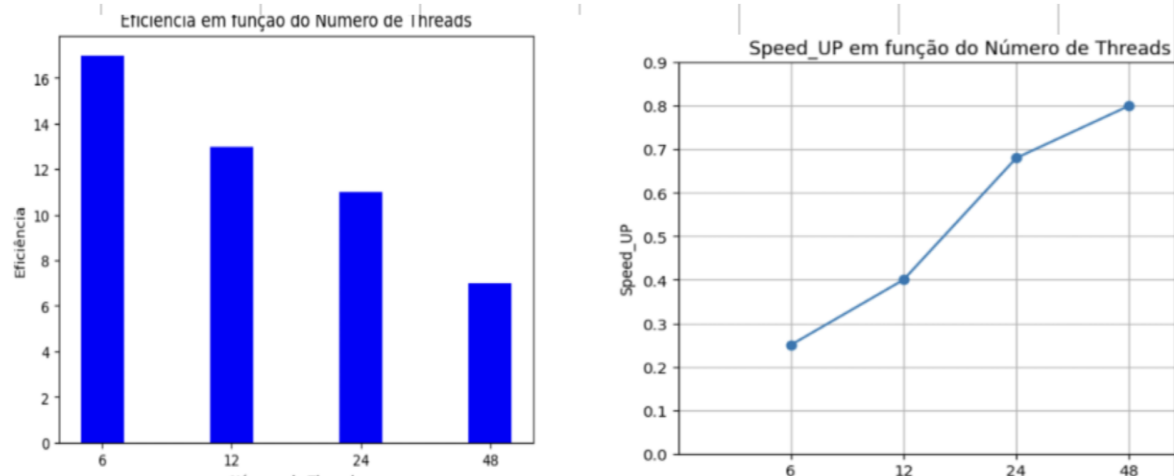
- Replicação - Live Jornal



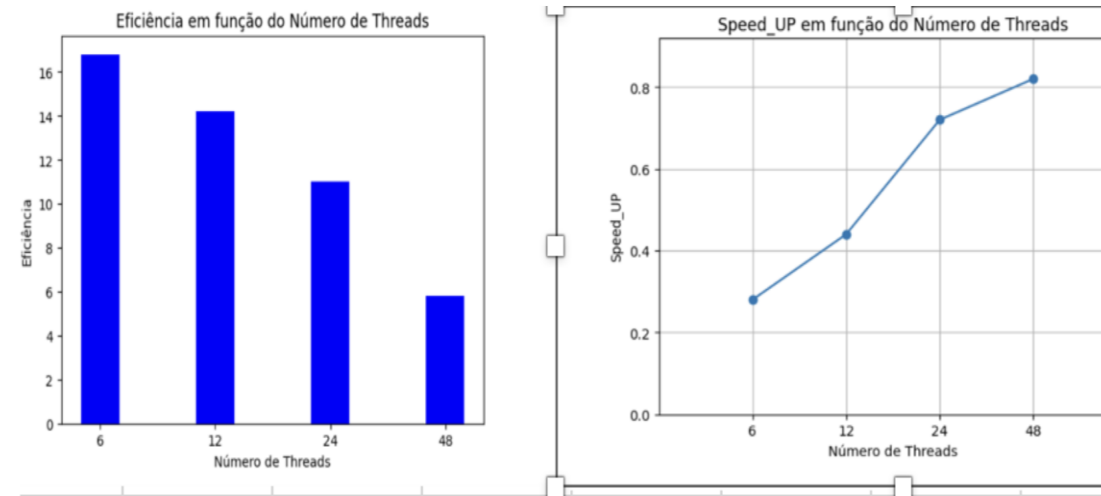
Resultados Comparativos – Original X Replicação(Gráficos)

- Representação grafica da comparação de resultados de *SpeedUp* e *eficiência* dos grafos no algoritmo *RelaxMap* entre este estudo e o original

- Estudo Original - Pokec



- Replicação - Pokec



Conclusões

- Portanto, o processo de estudo realizado a partir das análises de desempenho dos algoritmos destacados construiu-se de maneira satisfatória, onde foi possível desenvolver o aprendizado a respeito da identificação e detecção de comunidades em grafos.
 - Os resultados obtidos com base nos experimentos de replicação foram capazes de demonstrar as métricas e manipulações contidas no estudo original, além das abordagens de paralelismo, mesmo com uma quantidade de núcleos.
 - Além disso, a partir das análises realizadas em relação ao desempenho dos algoritmos, possibilitou-se encontrar a melhor alternativa para os diferentes grafos, que apresentam um comportamento mais econômico para cenários que não necessitem um maior número de núcleos e opções para a máxima eficiência em processos de detecção
-