

## Resumo

\* Este caderno de pesquisa tem como objetivo documentar o processo de replicação do estudo realizado por Dhruv Babani, Bernardo Balzan e Eduardo Cardoso sobre a detecção de comunidades em grafos utilizando o algoritmo Infomap. O estudo original investigou a escalabilidade do algoritmo Infomap em ambientes de memória compartilhada e comparou os resultados com o experimento original.

## 1.Introdução

\*O estudo original abordou a importância do desempenho de algoritmos que lidam com estruturas de dados, especialmente em grafos, para fins acadêmicos e analíticos. O artigo atual busca replicar e comparar os resultados obtidos no estudo original, bem como descrever o processo de replicação.

## 2.Fundamentação Teórica

\*Nesta seção, são apresentados os conceitos teóricos relevantes para a compreensão do estudo, incluindo informações sobre grafos, detecção de comunidades em grafos e paralelismo para melhorar a escalabilidade.

### 2.1 Grafos

- Definição de um grafo como modelo de estrutura de dados.
- Compreensão das comunidades de rede com base no grau dos vértices.
- Descrição das diferentes visões para a detecção de comunidades em grafos, incluindo estrutura fraca e estrutura forte.
- Classificação dos métodos de detecção de comunidades em grafos, como inferência estatística, otimização e simulação de dinâmica.

### 2.2 Paralelismo para Melhoria de Escalabilidade

- Explicação do paralelismo em memória compartilhada.
- Discussão sobre execução sequencial e concorrente.
- Introdução das métricas de Speed Up e Eficiência para medir o desempenho em ambientes paralelos.

## 2.3 Experimento original

- Descrição dos experimentos realizados no estudo original, incluindo os parâmetros de análise, como tempo de execução, escalabilidade, consumo de recursos e qualidade das comunidades.
- Apresentação dos algoritmos utilizados (InfoMap e RelaxMap).

## 3. Processo de Replicação

### 3.1 Seleção dos Grafos

Para a replicação do estudo, é fundamental escolher grafos apropriados que correspondam aos utilizados no estudo original. Isso garantirá que os resultados sejam comparáveis.

**Passo 1:** Identificar os grafos originais utilizados no estudo original (no caso, "Live Journal" e "Pokec") e obter informações sobre o número de vértices, arestas e outras métricas relevantes desses grafos.

**Passo 2:** Encontrar conjuntos de dados ou fontes confiáveis que disponibilizem os grafos escolhidos para download ou acesso. Certifique-se de que os conjuntos de dados estejam nas mesmas condições que os utilizados no estudo original.

**Passo 3:** Baixar os grafos selecionados e verificar se os arquivos de dados estão em formato apropriado (por exemplo, arquivos de arestas no formato correto).

**Passo 4:** Analisar os grafos para verificar se correspondem às informações fornecidas no estudo original (número de vértices, arestas, etc.). Faça isso para garantir que os dados estejam corretos.

### 3.2 Experimento para Detecção de Comunidades

Após a seleção dos grafos, é hora de executar o experimento de detecção de comunidades com os algoritmos InfoMap e RelaxMap, seguindo as configurações e métricas do estudo original.

#### **Passo 5:** Configurar o ambiente de experimentação:

- Certificar-se de que o ambiente de execução seja consistente com o ambiente do estudo original (no caso, macOS).
- Verificar se todas as bibliotecas necessárias para a manipulação de grafos e os algoritmos InfoMap e RelaxMap estão instaladas.
- Certificar-se de que o número de threads ou núcleos utilizados para a execução seja o mesmo usado no estudo original (por exemplo, 6, 12, 24, 48 threads).

#### **Passo 6:** Executar os algoritmos de detecção de comunidades:

- Iniciar a execução dos algoritmos InfoMap e RelaxMap nos grafos selecionados.
- Registrar o tempo de execução para cada configuração (número de threads) e para cada algoritmo.
- Monitorar o consumo de recursos, como uso de CPU e memória durante a execução.

#### **Passo 7:** Coletar os resultados:

- Registrar os resultados da detecção de comunidades para cada execução, incluindo informações sobre as comunidades encontradas e suas métricas (por exemplo, modularidade).
- Certificar-se de que os resultados estejam estruturados e documentados de forma adequada.

#### **Passo 8:** Analisar os resultados:

- Comparar os resultados obtidos na replicação com os resultados do estudo original.
- Avaliar a escalabilidade dos algoritmos, observando como o tempo de execução varia com o aumento do número de threads.
- Geração de gráficos demonstrando o percurso dos parâmetros de Eficiência e Speed Up, em relação ao número de threads.

Ao seguir esses passos, você estará preparado para replicar o estudo com precisão e gerar resultados comparáveis aos do estudo original.

### Codigos utilizados:

<https://github.com/mapequation/infomap> - InfoMap

<http://uwescience.github.io/RelaxMap/> - RelaxMap