

# HP-MOCD: Viabilizando a Detecção de Comunidades Multiobjetivo em Redes de Larga Escala

Guilherme O. Santos<sup>1</sup>, Lucas S. Vieira<sup>1</sup>  
Carlos H. G. Ferreira<sup>2</sup>, Gladston J. P. Moreira<sup>2</sup>

<sup>1</sup>Departamento de Computação e Sistemas  
Universidade Federal de Ouro Preto – João Monlevade, MG – Brasil

<sup>2</sup>Departamento de Computação  
Universidade Federal de Ouro Preto – Ouro Preto, MG – Brasil

guilherme.os1@aluno.ufop.edu.br, chgferreira@ufop.edu.br

gladston@ufop.edu.br

**Abstract.** *Community detection in complex networks is a fundamental problem, and classical single-objective approaches, such as Louvain and Leiden, although efficient, present important limitations, including the resolution limit and modularity degeneracy, which may lead to suboptimal or inconsistent partitions. In this context, multi-objective evolutionary algorithms (MOEAs) emerge as a promising alternative; however, they are typically restricted to networks with fewer than 10,000 nodes due to the high computational cost. This paper presents and discusses HP-MOCD, a high-performance multi-objective algorithm recently proposed in the literature for community detection, built upon NSGA-II. HP-MOCD combines topology-aware genetic operators, full parallelization, and bit-level optimizations, achieving a time complexity of  $O(G \cdot N_p |V|)$  for sparse graphs. Experimental results on LFR benchmarks and 14 real-world networks show that HP-MOCD is capable of processing graphs with more than one million nodes, outperforming competing MOEAs by up to  $553\times$  in runtime while maintaining competitive or superior partition quality.*

**Resumo.** *A detecção de comunidades em redes complexas é um problema fundamental, e abordagens clássicas de objetivo único, como Louvain e Leiden, embora eficientes, apresentam limitações importantes, incluindo o problema de resolução e a degenerescência da função de modularidade, que podem levar a partições subótimas ou inconsistentes. Nesse contexto, algoritmos evolucionários multiobjetivos (MOEAs) surgem como alternativa promissora, mas, tipicamente, restringem-se a redes com menos de 10.000 nós devido ao elevado custo computacional. Este artigo apresenta e discute o HP-MOCD, um algoritmo multiobjetivo de alto desempenho recentemente proposto na literatura para detecção de comunidades, construído sobre o NSGA-II. O HP-MOCD combina operadores genéticos cientes da topologia, paralelização completa e otimizações em nível de bits, alcançando complexidade  $O(G \cdot N_p |V|)$  para grafos esparsos. Resultados experimentais em benchmarks LFR e 14 redes reais mostram que o HP-MOCD é capaz de processar grafos com mais de um milhão de nós, superando MOEAs concorrentes em até  $553\times$  em tempo de execução e mantendo qualidade competitiva ou superior nas partições.*

## 1. Introdução

A análise de redes complexas é central em diversos domínios, especialmente nos sociais, e a identificação de comunidades, isto é, grupos de nós densamente conectados internamente e esparsos externamente, constitui um dos seus problemas fundamentais [Boccalletti et al. 2006, Newman 2006]. Dado que a detecção de comunidades é um problema NP-difícil [Newman and Girvan 2004], a literatura historicamente adotou heurísticas rápidas de objetivo único, que consistem em maximizar uma medida específica, como a modularidade, a exemplo dos algoritmos de Louvain [Blondel et al. 2008] e Leiden [Traag et al. 2019]. Apesar de sua eficiência, esses métodos apresentam limitações estruturais bem documentadas, como o *limite de resolução*, que impede a detecção de comunidades pequenas [Fortunato and Barthélemy 2007], e a *degenerescência* (do inglês, *degeneracy*), que produz partições distintas com pontuações de qualidade indistinguíveis [Traag et al. 2019, Azevedo et al. 2024]. Essas limitações decorrem do uso de uma única função objetivo, que não captura adequadamente os múltiplos critérios estruturais presentes em redes reais.

Nesse contexto, os Algoritmos Evolucionários Multiobjetivo (MOEAs) surgem como uma alternativa mais expressiva, pois permitem otimizar simultaneamente métricas conflitantes, como maximizar a coesão interna e minimizar conexões entre comunidades, resultando em uma *Fronteira de Pareto* com partições diversificadas e de alta qualidade [Moreira and Paquete 2019]. Diversas propostas foram desenvolvidas nessa direção, incluindo MOGA-Net [Pizzuti 2009], Shi-MOCD [Shi et al. 2010], KRM e CCM [Shaik et al. 2021] e CDRME [Dabaghi-Zarandi et al. 2025]. No entanto, apesar de seu potencial teórico, a aplicabilidade prática dos MOEAs permanece limitada, uma vez que o elevado custo computacional dessas abordagens tem restringido seu uso a redes com menos de 10.000 nós, inviabilizando sua adoção em aplicações de larga escala, como aquelas encontradas em plataformas de mídias sociais.

Este artigo apresenta e discute o **HP-MOCD** (*High-Performance Multi-Objective Community Detection*), um algoritmo evolucionário multiobjetivo escalável baseado no *framework* NSGA-II [Deb et al. 2002]. O método integra operadores genéticos cientes da topologia local, apresenta complexidade por geração de  $O(N_p|V|)$  em grafos esparsos e utiliza uma arquitetura *multithreaded* implementada em Rust com interface em Python (via PyO3), estando disponível como biblioteca de código aberto.<sup>1</sup>

Ao apresentar o HP-MOCD à comunidade de análise de redes sociais, este documento destaca seus principais aspectos metodológicos, seu desempenho computacional e seu potencial de aplicação em cenários que envolvem redes de grande escala, como plataformas digitais, sistemas de recomendação e análise de interações entre usuários. Os resultados completos desta pesquisa foram publicados no periódico *Springer Social Network Analysis and Mining* classificado como estrato A1 em Ciência da Computação pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e no quartil Q1 do SCImago Journal Rank (SJR)<sup>2</sup>.

---

<sup>1</sup>URL omitida para revisão.

<sup>2</sup><https://www.scimagojr.com/journalsearch.php?q=19700177337&tip=sid>

## 2. O Algoritmo HP-MOCD

O HP-MOCD é um algoritmo evolucionário multiobjetivo projetado para detecção de comunidades em larga escala, combinando operadores genéticos guiados pela topologia com uma arquitetura paralela eficiente. O método busca explorar o espaço de soluções de forma diversa, produzindo um conjunto de partições que representam diferentes compromissos entre critérios estruturais conflitantes.

### 2.1. Formulação Multiobjetivo

Dado um grafo não-direcionado  $G = (V, E)$ , uma partição comunitária  $\mathcal{C} = \{C_1, \dots, C_k\}$  é um conjunto de subconjuntos disjuntos que cobre  $V$ . A detecção de comunidades é formulada como a minimização simultânea de dois objetivos conflitantes [Shi et al. 2010]: maximizar a coesão interna ( $f_1$ ) e penalizar comunidades excessivamente grandes ( $f_2$ ):

$$f_1(\mathcal{C}) = 1 - \sum_{c \in \mathcal{C}} \frac{|E(c)|}{m}, \quad f_2(\mathcal{C}) = \sum_{c \in \mathcal{C}} \left( \frac{\sum_{v \in c} \deg(v)}{2m} \right)^2 \quad (1)$$

onde  $|E(c)|$  é o número de arestas internas da comunidade  $c$ ,  $m = |E|$  e  $\deg(v)$  é o grau do nó  $v$ . Esses objetivos capturam o compromisso entre coesão estrutural e balanceamento de comunidades, característico do problema.

### 2.2. Operadores Genéticos

O HP-MOCD introduz operadores que exploram diretamente a estrutura topológica da rede, de modo a preservar padrões relevantes e refinar progressivamente as partições:

- **Cruzamento de Votação Majoritária:** O descendente herda, para cada nó  $v$ , o rótulo comunitário mais frequente entre um conjunto de pais, preservando as características topológicas dominantes.
- **Mutação por Vizinhança:** Com probabilidade  $M_R$ , o rótulo de um nó  $v$  é substituído pelo rótulo predominante entre seus vizinhos  $N(v)$ , refinando as bordas das comunidades de forma guiada pela topologia local.

Como resultado, o algoritmo produz um conjunto de soluções não-dominadas (Fronteira de Pareto). Para selecionar uma única partição recomendada, aplica-se um critério de escalarização que maximiza  $Q(\mathcal{C}) = 1 - f_1 - f_2$ , equivalente à modularidade.

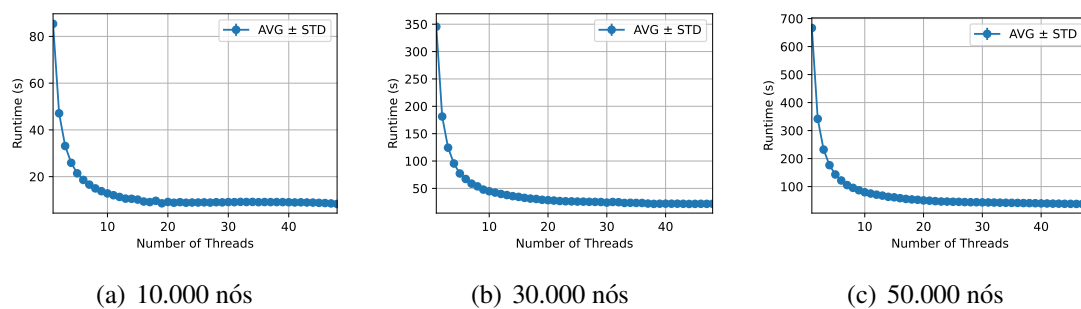
### 2.3. Arquitetura e Complexidade

A eficiência do HP-MOCD decorre da combinação entre estruturas de dados otimizadas e paralelização. A pré-computação de graus e o uso de estruturas *hash* garantem operadores genéticos em  $O(|V|)$  por indivíduo. Para grafos esparsos ( $|E| \approx O(|V|)$ ), a complexidade por geração é  $O(N_p|V|)$ , onde  $N_p$  é o tamanho da população, resultando em complexidade total de  $O(G \cdot N_p|V|)$  para  $G$  gerações. A paralelização *multithreaded* distribui a avaliação de aptidão e os operadores genéticos entre os núcleos disponíveis, permitindo escalabilidade para redes de grande porte.

## 2.4. Impacto da Paralelização Multi-Thread no Desempenho

A Figura 1 evidencia que o HP-MOCD se beneficia fortemente da paralelização, especialmente nas primeiras adições de *threads*. O tempo de execução diminui de forma acentuada até atingir um platô aproximado em 8 *threads* para 10.000 nós, 15 *threads* para 30.000 nós e 20 *threads* para 50.000 nós. Após esses pontos, os ganhos tornam-se marginais, indicando que o *overhead* de coordenação passa a compensar parte dos benefícios da execução paralela.

O contraste com a execução sequencial reforça esse impacto: os melhores tempos paralelos resultaram em *speedups* de  $10,27\times$ ,  $16,45\times$  e  $17,83\times$  para grafos com 10.000, 30.000 e 50.000 nós, respectivamente. Os experimentos foram conduzidos em um AMD Ryzen Threadripper 3960X, variando de 1 a 48 *threads* lógicas, com cinco repetições por configuração e baixo desvio padrão observado.



**Figura 1. Tempo médio de execução (média  $\pm \sigma$ ) em função do número de *threads* lógicas, para grafos sintéticos LFR com 10.000 (a), 30.000 (b) e 50.000 (c) vértices. Comparações conduzidas no mesmo *hardware* descrito no texto.**

## 3. Avaliação Experimental e Domínios de Aplicação

O HP-MOCD foi avaliado em redes sintéticas e reais de diferentes domínios, com destaque para redes sociais. O método foi comparado a MOEAs, como MOCD, MOGA-Net, KRM, CCM e CDRME, além de heurísticas clássicas de objetivo único, incluindo Louvain, Leiden e ASYN-LPA. A configuração experimental utilizou  $C_P = 0,8$ ,  $M_P = 0,2$ ,  $E_S = 4$ ,  $N_p = 100$  e 100 gerações.

A avaliação considerou as métricas NMI, AMI, F1-score pareado e modularidade, com significância estatística verificada por testes-*t* pareados com correção de Bonferroni. Nas redes sintéticas LFR, foram analisados cenários de robustez, variando o parâmetro de mistura  $\mu$ , e de escalabilidade, variando o número de nós. Cada configuração foi replicada 20 vezes. Também foram utilizadas quatorze redes reais, abrangendo desde redes sociais clássicas, como *Karate*, até redes de larga escala, como *Youtube*, além de redes de e-commerce, literatura científica e sistemas biológicos. Essa diversidade permitiu avaliar o desempenho do HP-MOCD em diferentes tamanhos, densidades e estruturas topológicas.

## 4. Resultados

Esta seção resume os principais achados experimentais do HP-MOCD.

#### 4.1. Redes Sintéticas

Nos experimentos de robustez ao parâmetro de mistura  $\mu$ , o HP-MOCD manteve alta acurácia (NMI e AMI estáveis) até  $\mu = 0,4$ , superando expressivamente os demais MOEAs. Para  $\mu \leq 0,3$ , rivalizou estatisticamente com Louvain e Leiden ( $p < 0,01$ ), conforme ilustrado na Figura 2.

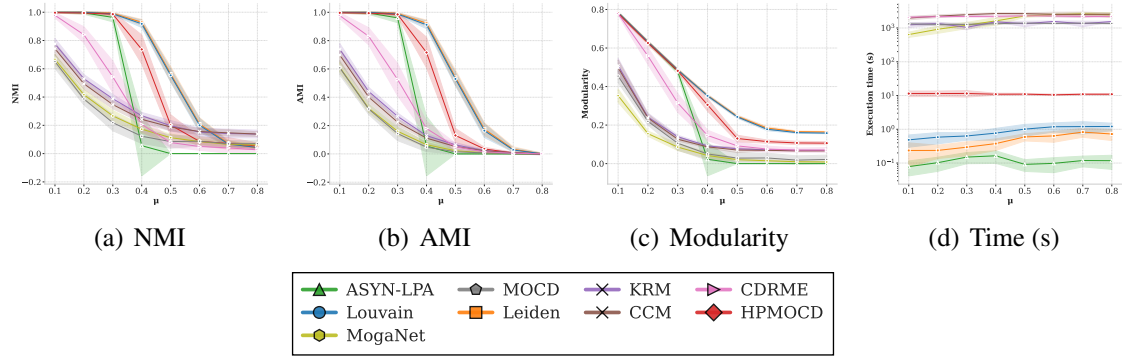


Figura 2. Robustez estrutural ( $\mu$ ) em redes LFR (1.000 nós). Médias de 20 execuções com IC de 95%.

Nos testes de escalabilidade, o HP-MOCD manteve  $NMI \approx 1,0$  em redes de até 20.000 nós, processando-as em  $148 \pm 13$  segundos. Em contraste, os demais MOEAs apresentaram degradação severa de qualidade e estouro de tempo (limite de 32h). Os ganhos variaram de  $132 \times$  a  $259 \times$  em relação a KRM, CCM e MOCD em redes de 13k–15,5k nós, e  $553 \times$  em relação ao CDRME em 2.000 nós, conforme Figura 3.

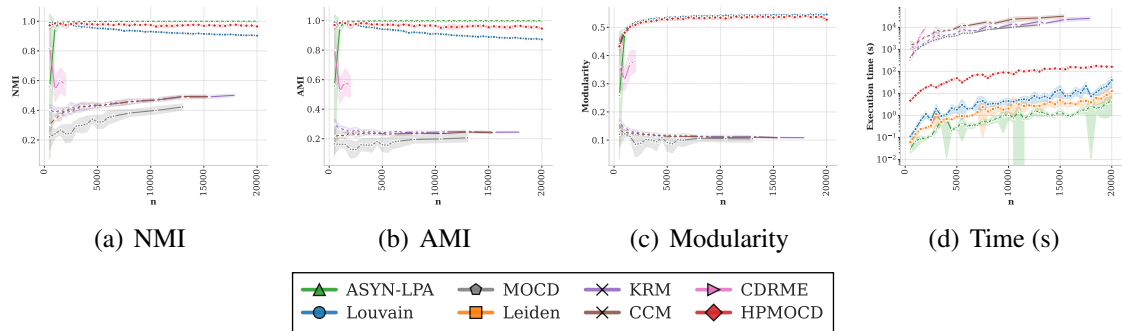


Figura 3. Escalabilidade em redes LFR ( $\mu = 0.3$ ). Lacunas indicam estouro do tempo limite (32h).

#### 4.2. Redes Reais

Em redes de pequeno e médio porte, o HP-MOCD foi um dos métodos mais consistentes: na rede *Football*, alcançou os maiores valores de AMI (0,881), NMI (0,912) e *F1-score* (0,864). Uma análise estatística revelou superioridade significativa em 13 de 32 combinações de métrica/dataset e equivalência em outras 14, totalizando dominância ou equivalência em aproximadamente 85% das avaliações contra MOEAs. O resultado mais expressivo reside na escalabilidade: o HP-MOCD foi o único MOEA capaz de processar as seis maiores redes dentro do limite de 15 horas, evidenciando sua aplicabilidade em cenários reais de larga escala. A Tabela 2 sintetiza os ganhos percentuais do HP-MOCD sobre os *baselines* multiobjetivo nos três cenários experimentais.

**Tabela 1. Desempenho do HP-MOCD nas 14 redes reais. Redes com †: nenhum outro MOEA completou em 15h.**

Rede	$ V $	AMI	NMI	Modulari.	F1
Youtube†	1.134.890	0,301 ± 0,016	0,538 ± 0,037	0,660 ± 0,009	0,033 ± 0,004
Amazon†	334.863	0,402 ± 0,006	0,667 ± 0,001	0,762 ± 0,012	0,007 ± 0,001
AS†	23.752	0,345 ± 0,010	0,419 ± 0,008	0,503 ± 0,016	0,070 ± 0,018
Cora†	23.166	0,441 ± 0,007	0,524 ± 0,004	0,660 ± 0,023	0,145 ± 0,021
CiteSeer†	3.327	0,199 ± 0,004	0,318 ± 0,003	0,792 ± 0,013	0,033 ± 0,009
Prog.†	3.109	0,342 ± 0,008	0,391 ± 0,007	0,669 ± 0,019	0,150 ± 0,039
OS	2.176	0,282 ± 0,011	0,308 ± 0,010	0,580 ± 0,014	0,195 ± 0,063
Network	1.249	0,254 ± 0,014	0,285 ± 0,012	0,633 ± 0,020	0,176 ± 0,040
Databases	1.046	0,313 ± 0,007	0,369 ± 0,006	0,683 ± 0,023	0,171 ± 0,032
Eu-core	1.005	0,494 ± 0,026	0,563 ± 0,025	0,399 ± 0,009	0,313 ± 0,047
Football	115	0,881 ± 0,013	0,912 ± 0,010	0,584 ± 0,009	0,864 ± 0,023
Polbooks	105	0,502 ± 0,018	0,521 ± 0,016	0,525 ± 0,002	0,743 ± 0,024
Dolphins	62	0,532 ± 0,074	0,550 ± 0,068	0,518 ± 0,006	0,544 ± 0,077
Karate	34	0,658 ± 0,046	0,676 ± 0,041	0,417 ± 0,004	0,698 ± 0,033

**Tabela 2. Percentual de superioridade do HP-MOCD sobre MOEAs concorrentes e melhoria média.**

Cenário	Métrica	Sup. (%)	Melhoria (%)
Escalabilidade	Modulari.	34,1	111,91
	NMI	74,3	50,86
	AMI	74,3	113,24
Robustez ( $\mu$ )	Modulari.	69,0	227,93
	NMI	63,8	144,88
	AMI	67,2	230,94
Redes Reais	Modulari.	55,0	58,35
	NMI	51,7	13,30
	AMI	55,0	65,60

## 5. Conclusão

Este artigo apresentou o HP-MOCD, um algoritmo evolucionário multiobjetivo eficiente para detecção de comunidades em redes complexas. Ao combinar operadores genéticos cientes da topologia, paralelização e otimizações de implementação, o método alcança complexidade  $O(G \cdot N_p |V|)$  em grafos esparsos e processa redes com mais de um milhão de nós em tempo prático, superando concorrentes em até  $553\times$ .

Os resultados mostram que o HP-MOCD mantém alta qualidade nas soluções, com dominância ou equivalência estatística em 85% das avaliações contra MOEAs, além de desempenho competitivo frente a heurísticas clássicas em cenários de baixa mistura. A geração de uma Fronteira de Pareto amplia sua utilidade ao oferecer múltiplas partições de qualidade, adequadas a diferentes critérios estruturais e contextos de aplicação.

Assim, o HP-MOCD demonstra que abordagens multiobjetivo podem ser aplica-

das de forma prática à detecção de comunidades em larga escala, ampliando seu potencial de uso em redes sociais, sistemas biológicos, redes de informação e plataformas digitais.

## Agradecimentos

Este trabalho foi apoiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-307151/2022-0). Agradecemos também à UFOP pelo apoio.

## Referências

- Azevedo, B. F., Rocha, A. M. A. C., Fernandes, F. P., Pacheco, M. F., and Pereira, A. I. (2024). Comparison between single and multi-objective clustering algorithms: Mathe case study. In *Optimization, Learning Algorithms and Applications*.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4):175–308.
- Dabaghi-Zarandi, F., Afkhami, M. M., and Ashoori, M. H. (2025). Community detection method based on random walk and multi objective evolutionary algorithm in complex networks. *Journal of Network and Computer Applications*.
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197.
- Fortunato, S. and Barthélemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104.
- Moreira, G. and Paquete, L. (2019). Guiding under uniformity measure in the decision space. In *2019 IEEE Latin American Conference on Computational Intelligence (LACCI)*, pages 1–6.
- Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582.
- Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113.
- Pizzuti, C. (2009). A multi-objective genetic algorithm for community detection in networks. In *21st IEEE International Conference on Tools with Artificial Intelligence*, pages 379–386.
- Shaik, T., Ravi, V., and Deb, K. (2021). Evolutionary multi-objective optimization algorithm for community detection in complex social networks. *SN Computer Science*, 2(1):1.
- Shi, C., Zhong, C., Yan, Z., Cai, Y., and Wu, B. (2010). A multi-objective approach for community detection in complex network. In *IEEE Congress on Evolutionary Computation*, pages 1–8.
- Traag, V. A., Waltman, L., and van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1):5233.