

ISSN: 2319-0124

ALGORITMO GENÉTICO COM FUNÇÃO *FITNESS* PONDERADA PARA DETECÇÃO DE COMUNIDADES EM REDES COMPLEXAS

Eduardo P. AMARAL¹; Diego SAQUI²

RESUMO

Este estudo faz a aplicação de um Algoritmo Genético (AG) para detectar comunidades em redes complexas. Redes complexas são estruturas compostas por nós e conexões entre os mesmos, representando relações entre fatores. Comunidades neste contexto se referem às seções de uma rede, onde há maior quantidade de arestas internas do que externas. As comunidades possuem relevância em áreas nas quais se estudam conceitos de alta complexidade e que são formados por diversas variáveis interconectadas, como é o caso de redes sociais, clima ou sistemas gravitacionais. O AG usado neste estudo, é um método computacionalmente viável por conseguir resolver problemas de alta complexidade em um período de tempo aceitável, além de explorar o espaço de busca de uma maneira inteligente. Para isso, a métrica chamada Community Score foi aplicada junto da Modularidade como uma função de fitness ponderada do AG e assim determinar a qualidade das comunidades obtidas ao longo do processo.

Palavras-chave:

Redes Complexas; Comunidades; Grafos; Algoritmo Genético; Métricas.

1. INTRODUÇÃO

Com a evolução da ciência surgiram estruturas de alta complexidade ao relacionar dados. A densidade de elementos e suas conexões dificultam seu entendimento em microescala, tendo em vista que a conduta geral é fruto de comportamentos díspares, cuja atuação em conjunto forma o fenômeno em larga escala da rede complexa (aqui chamaremos de rede) (BAR-YAM, 2002).

Uma categoria de algoritmos que pode detectar comunidades, que são segmentações em redes, são os Algoritmos Genéticos (AGs). Estes selecionam indivíduos representados por vetores, avalia-os e separa os melhores para cruzamento e mutação. São muito usados devido às suas capacidades: (1) de contornar ótimos locais; (2) de explorar o espaço de busca de uma forma planejada; e (3) limitar seu tempo de execução, o que é adequado para problemas de complexidade elevada, como a detecção de comunidades.

Certas métricas que avaliam comunidades de redes são: a *Modularity* (MD) que relaciona a quantidade de arestas esperadas entre nós com as que realmente existem, mensurando a qualidade de um grafo de se modularizar (BRANDES, et al., 2008); *Betweenness centrality* que define o quão central é cada nó em relação aos demais (BORGATTI; WHITE, 1994); *Clustering coefficient* que

¹ Bolsista FAPEMIG, Eduardo P. AMARAL, Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais - *Campus* Muzambinho. Email: eduardo.pioliam@gmail.com.

² Orientador, Diego Saqui, Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais - *Campus* Muzambinho. Email: diego.saqui@muz.ifsulde Minas.edu.br.

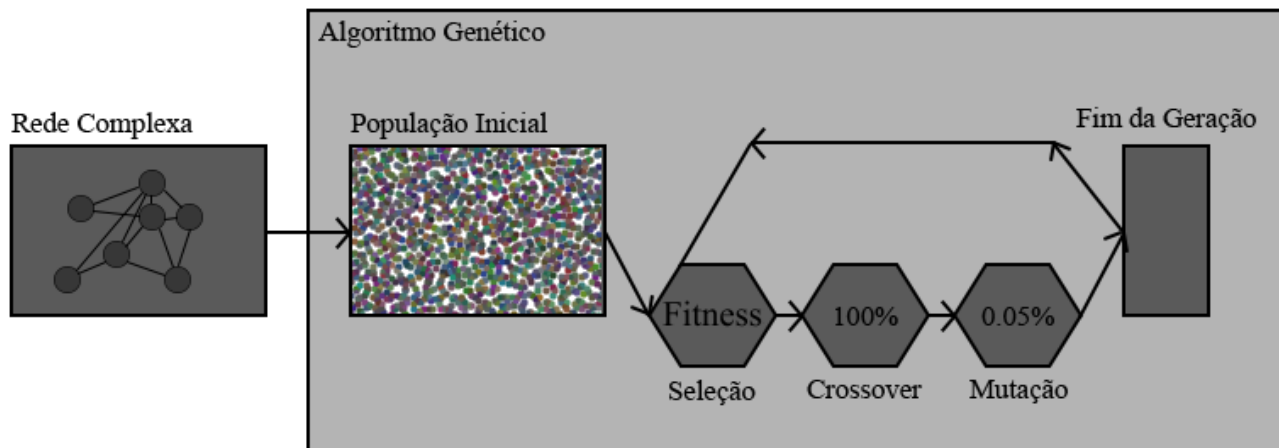
representa uma medida de tendência de agrupamento de nó (HUANG, 2010); e o *Community Score* (CS) que se baseia num co-agrupamento das dimensões de um grafo, detectando submatrizes máximas e densas (PIZZUTI, 2008).

Muitos dos algoritmos existentes operam com base em apenas uma métrica de qualidade de comunidades, e com propósito de melhorar a qualidade da detecção de comunidades, neste estudo foi estabelecido uma função ponderada composta pelas métricas MD e CS.

2. MATERIAL E MÉTODOS

A priori, o AG gera aleatoriamente um conjunto (população) com 100 possíveis soluções (indivíduos), e qualifica-as usando a função de *fitness*. Então, os indivíduos de melhor pontuação são selecionados para o cruzamento, que gera novos indivíduos (filhos) com genes escolhidos entre pares selecionados. Estes filhos podem passar pela mutação, que altera um gene aleatoriamente.

Figura 1 - Diagrama do funcionamento geral de um AG.



Fonte: Elaboração do autor (2022).

No total o processo itera (gerações) cem vezes. O modelo geral pode ser observado na Figura 1 junto com as probabilidades de ocorrência de cada operação.

A rede ilustrada na Figura 1 representa o *Random Geometric Graph*³ que retorna um grafo aleatório, gerado pela biblioteca *networkx*⁴ para linguagem Python. O cruzamento forma um vetor binário cujos índices representam um par ou outro. A mutação define partes do indivíduo cujos valores são trocados por outros escolhidos por um *array* de densidade dos dados matriciais do indivíduo. As duas métricas implementadas no AG, o CS e a MD, foram usadas para estabelecer a função de *fitness* conforme a equação (1).

³ [Random Geometric Graph — NetworkX 2.8.4 documentation](#)

⁴ [Software for Complex Networks — NetworkX 2.8.4 documentation](#)

$$M = \sum_{c=1}^n \left[\frac{|I|}{\sum_{i \in I} (a_{ij})^t} + \left[\frac{L_c}{m} - \gamma \left(\frac{k_c}{2m} \right)^2 \right] \right] \quad (1)$$

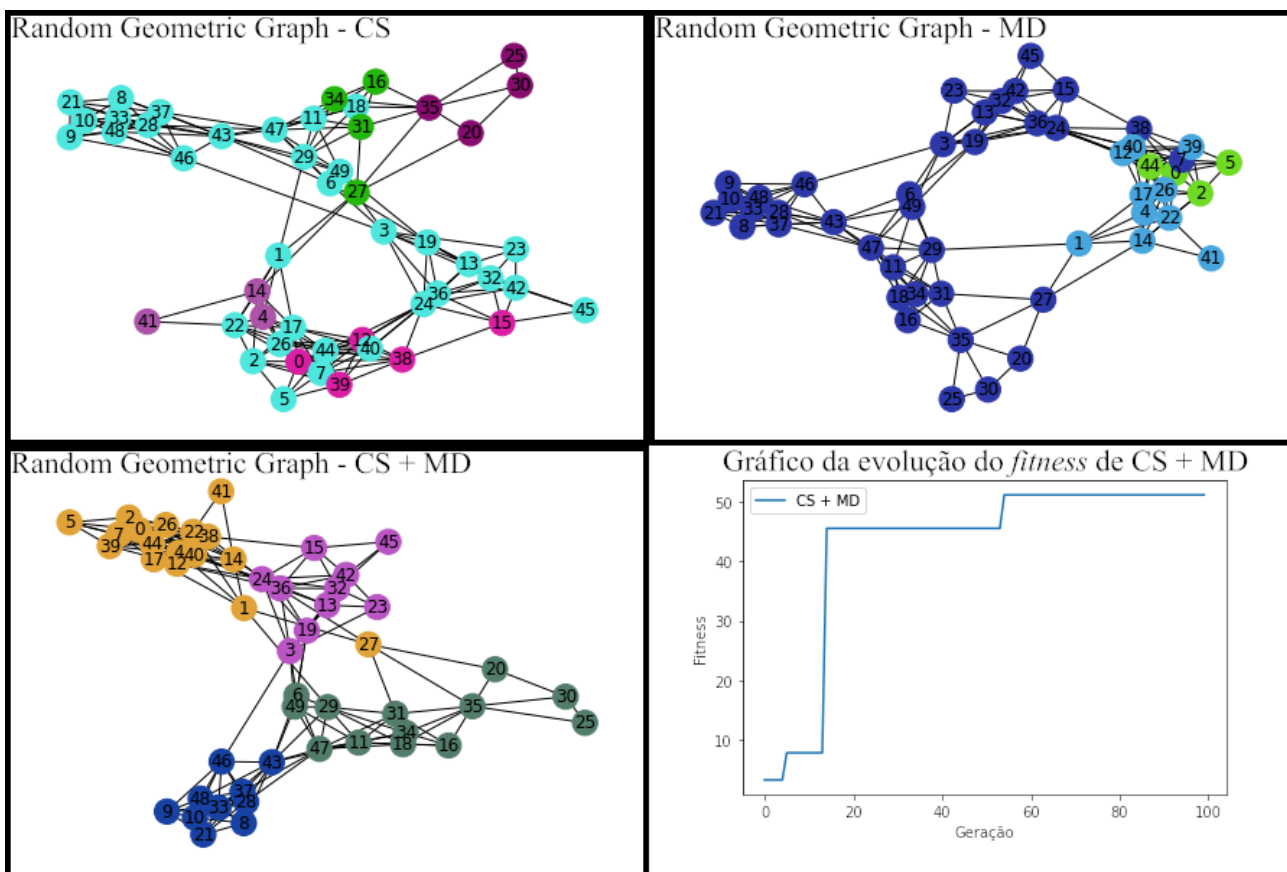
Onde c corresponde à uma comunidade de n existentes; I ao subconjunto das linhas da submatriz; $(a_{ij})^t$ à média da linha i elevada à uma ordem de valor experimental 1,5; L_c à quantidade de intra-arestas de c ; m ao total de arestas do grafo; γ ao valor de resolução de controle da modularidade cujo valor experimental é 1; e k_c à soma dos graus dos nós da comunidade.

Tendo em vista a tendência de queda da CS, seu inverso foi usado em alinhamento com o aumento natural da MD.

3. RESULTADOS E DISCUSSÕES

A Figura 2 abaixo demonstra como cada métrica dividiu um mesmo grafo em comunidades.

Figura 2 - Comunidades detectadas por métrica e gráfico de evolução de CS + MD.



Fonte: Elaboração do autor (2022).

Nota-se que a CS encontrou mais comunidades que a MD, porém com maior mistura de nodos, e o inverso na MD. Contudo, a união das duas expôs mais comunidades mantendo a coerência de nodos. Pelo gráfico vê-se a frequência de saltos de ótimos locais prematuramente,

sendo um deles de alta magnitude. Podemos inferir que a união das métricas acaba por juntar os melhores aspectos de ambas métricas ademais com baixo custo adicional.

5. CONCLUSÕES

Através de um estudo de complementaridade entre métricas em um AG, verificou-se que a operação destas juntas pode formar efeitos de interferência construtiva. Para tais casos a função não só é refinada como suscita maior manipulação da mesma, tendo em vista a existência de constantes controladoras de aspectos específicos de cada métrica, além de abrir portas para o desenvolvimento de heurísticas de maior complexidade e eficiência.

Com a detecção de comunidades aprimoradas, o estudo de fenômenos e sistemas de alta complexidade é facilitado, possibilitando o estudo minucioso dos efeitos micro e suas inter-relações, além de pontuar momentos ou relações cruciais entre segmentos onde há a transição de comportamentos micro para macro.

AGRADECIMENTOS

Agradecemos à FAPEMIG, ao LabSoft e ao IFSULDEMINAS- Campus Muzambinho pela oportunidade e estrutura concedidas para realização dessa pesquisa.

REFERÊNCIAS

BAR-YAM, Yaneer. **General features of complex systems**. Encyclopedia of life support systems. [S.I.], v.1. 2002.

BRANDES, U., et al. **On modularity clustering**. IEEE, [S.I.], v.20. 2008.

BORGATTI, Stephen P.; WHITE, Douglas R. **Betweenness centrality measures for directed graphs**. ScienceDirect, [S.I.], v.16. 1994.

HUANG, Zan. **Link prediction based on graph topology: The predictive value of generalized clustering coefficient**. SSRN, [S.I.], 2010. Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1634014>. Acesso em: 01/Jun/2022.

PIZZUTI, Clara. **GA-Net: A genetic algorithm for community detection in social networks**. ResearchGate, Roma. 2008. Disponível em: <https://www.researchgate.net/publication/220701568_GA-Net_A_Genetic_Algorithm_for_Community_Detection_in_Social_Networks>. Acesso em: 02/Jun/2022.