UNIVERSIDADE FEDERAL DE ALAGOAS
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
NÚCLEO DE INOVAÇÃO TECNOLOGICA

PROGRAMA INSTITUCIONAL DE BOLSAS DE INICIAÇÃO CIENTÍFICA PIBIC/UFAL/FAPEAL/CNPq

RELATÓRIO PARCIAL (2020–2021)

TÍTULO DO PROJETO:
Algoritmos exatos e heurísticos para problemas combinatórios em grafos

TÍTULO DO PLANO DE ATIVIDADES:
Algoritmos exatos para o problema da biclique máxima

Nome do Orientador/Unidade / Campus / Email	Rian Gabriel Santos Pinheiro / Instituto de Computação / Campus A. C. Simões/ rian@ic.ufal.br
Nome do Aluno:	Lucas Carvalho Flores
Email/Fone (aluno)	lcf@ic.ufal.br / (82) 982232497

	BOLSISTA CNPQ		BOLSISTA FAPEAL
	BOLSISTA UFAL	X	COLABORADOR
	BOLSISTA CNPQ-AF		

Obs.: Marcar com um "X" o tipo de bolsa ou colaborador

RESUMO DO PROJETO

Palavras Chave: Biclique, Grafos, Algoritmos Heurísticos.

O foco deste projeto está voltado para algumas aplicações de Otimização Combinatória em problemas de Teoria dos Grafo. Muitos problemas reais de natureza combinatória podem ser modelados por modelos matemáticos discretos, como grafos por exemplo. A solução destes problemas frequentemente exige uma abordagem algorítmica, juntamente com a implementação destes algoritmos em algum sistema computacional. É importante ressaltar a necessidade de obter algoritmos eficientes, pois tipicamente tanto o volume de dados como a frequência de acessos ao sistema são grandes. Se o problema em questão possui complexidade intrinsecamente alta, uma possibilidade é recorrer a um algoritmo aproximativo ou heurístico em vez de um algoritmo exato.

Neste projeto serão abordados o problema da biclique máxima e suas variantes. Dado um grafo $G = (V, E)$, em que V é um conjunto não vazio de vértice e E é um conjunto de arestas. Um par de subconjuntos disjuntos A e B de V é dito biclique se $(a, b) \in E$ para todo $a \in A$ e $b \in B$. Em outras palavras, uma biclique corresponde a um subgrafo bipartite completo de G. Dado como entrada um grafo G, o objetivo do problema da BICLIQUE MÁXIMA é encontrar uma biclique de G que maximize o número de vértices.

OBJETIVOS DO PROJETO DE PESQUISA

Geral

O projeto tem como objetivo geral desenvolver soluções algorítmicas para problemas de Otimização em Teoria do Grafos utilizando abordagens baseadas em métodos consagrados de otimização como a programação matemática e meta-heurísticas.

Específicos

Como objetivos específicos:

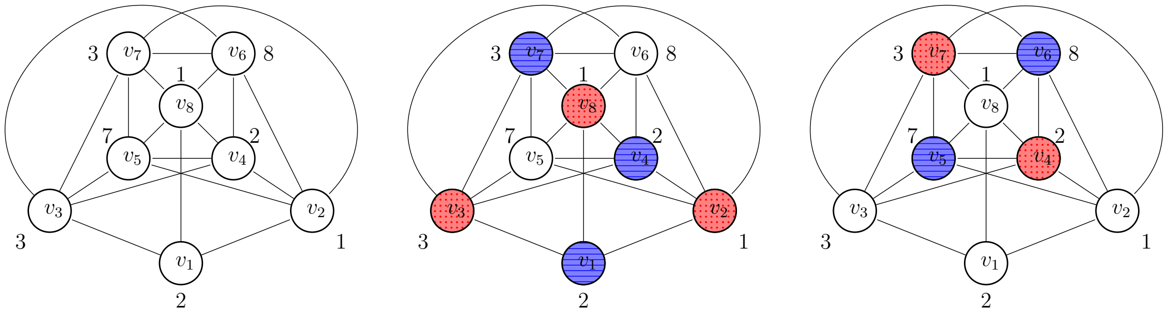
- A construção de algoritmos eficientes para a resolução dos problemas abordados neste projeto;
- Desenvolvimento de *software* de código aberto que possa ser utilizado pela academia e indústria;
- Obtenções de novos resultados científicos e divulgação de pesquisas em veículos de impacto;
- Formação de novos pesquisadores. Este projeto pretende capacitar o aluno a escrever e ler de forma crítica, discutir temas emergentes de pesquisa e capacitá-lo para uma vida acadêmica em pós graduação em nível de mestrado e doutorado, além de incentivá-lo a buscar autonomia em estudar e apresentar soluções a problemas complexos.

OBJETIVO ESPECÍFICO DO PLANO DE ATIVIDADES DO ALUNO

O projeto tem como objetivo geral desenvolver e melhorar soluções exatas para o problema da biclique máxima utilizando programação matemática. O presente plano de trabalho tem os seguintes objetivos específicos: (i) estudar modelos e algoritmos exatos existentes; (ii) desenvolver novos modelos exatos mais eficientes.

DETALHAR ETAPAS DO PLANO DE ATIVIDADE ALUNO

Dado um grafo $G = (V, E)$ de entrada, o problema da biclique (subgrafo bipartite completo) máxima balanceada tem como objetivo encontrar uma biclique balanceada (partes com o mesmo número de vértice) que tenha a cardinalidade maximizada. Uma generalização deste problema é a MÁXIMA BICLIQUE BALANCEADA COM PESO NOS VÉRTICES (PMBBPV), nesta variante, além do grafo de entrada, existe uma função de peso $w : V \to \mathbb{R}$ que atribui a cada vértice $v \in V$ um peso $w(v)$. Dessa forma, busca-se encontrar uma biclique balanceada em que a soma de todos os pesos dos vértices pertencentes a solução seja a maximizada. A Figura 1a mostra um grafo G de entrada para o PMBBPV com 8 vértices. Uma solução viável $G' = (V', E')$ com peso 12 é apresentada na Figura 1b, note que G' possui 6 vértices os quais os seus vértices V' são particionados em V'_1 e V'_2 formados pelos vértices em azul tracejado e vermelho pontilhado, respectivamente. G' é uma biclique balanceada uma vez que: (i) para todo par de vértice $v \in V'_1$ e $u \in V'_2$ tem-se que $vu \in E$, (ii) V'_1 e V'_2 são conjuntos independentes e (iii) $|V'_1| = |V'_2|$. Finalmente, a Figura 1c apresenta a biclique G'' com peso 20 que é a solução ótima do problema, mesmo tendo uma cardinalidade menor que G'.



(a) Instância. (b) Sol. viável com peso 12. (c) Sol. ótima com peso 20.

Figura 1: Biclique máxima ponderada.

Existe uma discussão sobre a complexidade de encontrar uma biclique em um grafo bipartido, já que possuem duas variantes principais de problemas relacionados. Uma delas se caracteriza pela propriedade de balanceamento, ou seja, se $|A| = |B| = K$, sendo K um inteiro, então o problema é $\mathcal{NP}$-difícil. A outra variante é quando a biclique não é balanceada, ou seja, o objetivo é maximizar $|A| + |B|$, neste caso o problema pode ser resolvido em tempo polinomial (Garey & Johnson, 1979; Feige & Kogan, 2004; Peeters, 2003). Logo, a complexidade da biclique máxima irá depender da definição utilizada, já que terá uma complexidade para a versão balanceada e outra para a não balanceada. Dessa forma, o problema abordado — biclique balanceada com peso no vértice — sendo uma generalização da versão balanceada sem peso, também é $\mathcal{NP}$-completo. Outros problemas relacionados remoção de arestas/vértices para formação de bicliques são abordados em Yannakakis (1981) e Hochbaum (1998).

Durante o planejamento e escolha do tema, houve um levantamento bibliográfico, onde foram encontrados diversos artigos sobre a biclique máxima balanceada (sem peso) como o Wang et al. (2018), que desenvolveu quatro heurísticas com buscas locais tendo um bom resultado nos testes realizados. Outro artigo encontrado foi o Zhou et al. (2018) que trata sobre um algoritmo exato para resolver a biclique máxima balanceada (sem peso) usando um algoritmo *upper bound propagation* integrado com um *branch-and-bound* que, em alguns casos, ele consegue reduzir o tempo computacional em 4 ordens de magnitude comparando com um algoritmo original *branch-and-bound*.

Uma outra abordagem sobre a biclique máxima balanceada foi feita pelo Zhou & Hao (2017) que utilizou um algoritmo de Constraint-Based Tabu Search (CBTS) e duas técnicas de redução de grafo que atuam em conjunto com a busca tabu apresentada.

Nos primeiros meses foi estudado o problema e realizado buscas na literatura por técnicas para o problema abordado. Foi então escolhida a técnica de Programação Linear Inteira (PLI).

Um Problema de Programação Inteira é um modelo de programação matemática no qual algumas ou todas as variáveis do problema pertencem ao conjunto dos números inteiros. Quando todas as variáveis são inteira o modelo é denominado programação inteira pura; caso contrário, é denominado programação inteira mista (Hillier, 2013). Abaixo, um exemplo de um problema de programação inteira.

$$\max \quad \sum_{i=1}^{n} c_i x_i \tag{1}$$

$$\text{s. a} \quad \sum_{i=1}^{n} a_{ij} x_i \leq b_i \qquad \forall j = 1, \ldots, m \tag{2}$$

$$x_i \in Z \qquad \forall i = 1, \ldots, n. \tag{3}$$

De acordo com essa definição, a MÁXIMA BICLIQUE BALANCEADA COM PESO NOS VÉRTICES (PMBBPV) pode ser modelada como um problema de programação linear inteiro puro (sendo os pesos inteiros), como o proposto em Assunção & Pinheiro (2020).

Atualmente o trabalho está em melhorar o algoritmo inserindo novos cortes, colocando restrições durante a execução ao invés de criá-las para todas as variáveis de decisão logo no início, permitindo que o procedimento de branch-and-bound realize podas mais cedo para chegar à solução ótima mais rápido.

APRESENTAÇÃO E DISCUSSÃO DOS PRINCIPAIS RESULTADOS OBTIDOS DURANTE O PRIMEIRO SEMESTRE DA PESQUISA

Algoritmo de PLI

O algoritmo exato foi implementado aplicando PLI, e foi utilizada a formulação matemática proposta em Assunção & Pinheiro (2020):

$$
\begin{aligned}
\max \quad & \sum_{v_i \in V} w_i x_i + \sum_{v_i \in V} w_i y_i && && (4)\\
\text{s. a} \quad & x_i + y_i \leq 1 && \forall v_i \in V && (5)\\
& x_i + x_j \leq 1 && \forall v_i v_j \in E && (6)\\
& y_i + y_j \leq 1 && \forall v_i v_j \in E && (7)\\
& x_i + y_j \leq 1 && \forall v_i v_j \notin E && (8)\\
& y_i + x_j \leq 1 && \forall v_i v_j \notin E && (9)\\
& \sum_{v_i \in V} x_i = \sum_{v_i \in V} y_i && && (10)\\
& x_i, y_i \in \{0,1\} && \forall v_i \in V. && (11)
\end{aligned}
$$

A formulação utiliza variáveis decisão binárias x_i e y_i para cada vértice $v_i \in V$. A variável x_i vale 1 se e somente se o vértice v_i pertencerá à parte V_1 na solução final. Da mesma forma, y_i determina se v_i pertencerá à parte V_2.

- A função objetivo (4) contabiliza o peso da biclique.
- As restrições (5) determinam que um vértice só pode pertencer a uma parte.
- As restrições (6) e (7) determinam que vértices adjacentes não podem pertencer a mesma parte.
- As restrições (8) e (9) determinam que vértices não-adjacentes não podem pertencer a partes distintas.
- A restrição (10) força a biclique ser balanceada. Finalmente, as restrições (11) determinam os domínios das variáveis.

Resultados

O modelo foi implementado com o CPLEX 12.9 na linguagem C++, compilado com g++ v10.2.0 (GCC), com a opção de optimização -O3 e em um computador Manjaro Linux 20 de 64 bits, 16GB de RAM e processador Intel Core i7 2.8GHz .

A Tabela 1 representa os resultados obtidos com alguns dos grupos de instâncias DIMACS e BHOSLIB. Os resultados obtidos utilizando programação matemática para cada instância são apresentados pela coluna `Resultado` (indicando o resultado ótimo da biclique) e pela coluna `Tempo(s)`, em segundos.

Tabela 1: Resultados para as instâncias do DIMACS e BHOSLIB.

Instância	Resultado	Tempo (s)
Instâncias DIMACS		
C125-9	673	2.19
C250-9	1180	8.57
C500-9	1373	571.98
MANN-a9	255	0.04
MANN-a27	1173	1.14
MANN-a45	1191	15.79
MANN-a81	1194	185.33
keller4	2166	67.69
brock200-1	1542	174.97
Instâncias BHOSLIB		
frb30-15-1	5340	23.12
frb30-15-2	5565	18.01
frb30-15-3	5565	10.78
frb30-15-4	5415	17.00
frb30-15-5	5415	17.29
frb35-17-1	6222	37.60
frb35-17-2	6290	34.31
frb35-17-3	6222	29.81

CRONOGRAMA DE ATIVIDADES

Atividade 1: Levantamento do estado da arte a respeito do problema nas bases científicas descritas;

Atividade 2: Planejamento e análise (teórica) dos algoritmos propostos;

Atividade 3: Implementação e execução de algoritmo exato, bem como melhorias a soluções já existentes;

Atividade 4: Comparação e análise dos resultados obtidos com os resultados já disponíveis na literatura;

Atividade 5: Escrita do relatório final e artigos a serem submetidos a conferências ou revistas.

Tabela 2: Cronograma de atividades do Aluno 2

ATIV.	**Meses**											
	2020					**2021**						
	SET	OUT	NOV	DEZ	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO
1	OK	OK	OK	OK	OK	OK						
2	OK											
3		OK	OK	OK	OK	OK						
4												
5	OK	OK	OK									

O levantamento bibliográfico foi uma atividade realizada durante todo esse tempo de pesquisa para verificar se novos artigos estão sendo escritos sobre o tema e é uma atividade que irá durar até o final da pesquisa. A fase de planejamento dos algoritmos e modelos matemáticos aconteceu durante o primeiro mês.

Na atividade 3 ocorreu a implementação de um modelo de programação matemática que teve a sua primeira versão em dezembro, sendo usado para testes e obtenção dos resultados parciais. O algoritmo ainda terá novas versões desenvolvidas, pois será necessária uma otimização.

O resultado obtido será comparado com as versões antigas do trabalho desenvolvido. A escrita do relatório final será a última parte da pesquisa, sendo realizada nos últimos meses e terá um artigo submetido para uma revista.

Com a implementação do modelo exato, serão propostos novas maneiras de acelerar o algoritmo exato. Para isso, serão desenvolvidos novos cortes e algoritmos de separações para o problema. Isso é, o modelo pode ser inicializado sem algumas restrições, e a medida que que soluções inteiras forem encontradas, serão executados os algoritmos de separação para determinar se alguma das restrições não adicionadas foi violada. Em caso, positivo, adiciona-se a restrição e o processo continua.

RELACIONE OS PRINCIPAIS FATORES POSITIVOS E NEGATIVOS QUE INTERFERIRAM NA CONDUÇÃO DO PROJETO E PLANO DE ATIVIDADES

Um fator determinante no andamento do projeto foi o auxílio do professor e orientador Rian Pinheiro, presente em todas as etapas do projeto explicando problemas, soluções e dando o direcionamento da pesquisa.

O maior impasse durante a pesquisa foi o afastamento momentâneo do estudante Lucas Carvalho das suas atividades no projeto devido a contrair o Covid-19. A dificuldade de se encontrar artigos pertinentes ao problema da biclique máxima balanceada com peso no vértice também foi um impasse.

Referências

Assunção, L. M. A. & Pinheiro, R. G. S. (2020), Algoritmo grasp-vnd para o problema da máxima biclique balanceada com peso no vértice, *in* 'ANAIS DO LII SIMPóSIO BRASILEIRO DE PESQUISA OPERACIONAL'.

Feige, U. & Kogan, S. (2004), 'Hardness of approximation of the balanced complete bipartite subgraph problem'.

Garey, M. R. & Johnson, David S., .-a. (1979), *Computers and intractability : a guide to the theory of NP-completeness*, San Francisco : W.H. Freeman. Includes indexes.

Hillier, F. S. (2013), *Introdução à Pesquisa Operacional*, Porto Alegre: McGraw-Hill.

Hochbaum, D. S. (1998), 'Approximating clique and biclique problems', *Journal of Algorithms* **29**(1), 174 – 200. URL `http://www.sciencedirect.com/science/article/pii/S0196677498909646`.

Peeters, R. (2003), 'The maximum edge biclique problem is np-complete', *Discrete Applied Mathematics* **131**(3), 651 – 654. URL `http://www.sciencedirect.com/science/article/pii/S0166218X03003330`.

Wang, Y., Cai, S. & Yin, M. (2018), 'New heuristic approaches for maximum balanced biclique problem', *Information Sciences* **432**, 362 – 375. URL `http://www.sciencedirect.com/science/article/pii/S0020025517311337`.

Yannakakis, M. (1981), 'Edge-deletion problems', *SIAM Journal on Computing* **10**(2), 297–309. URL `https://doi.org/10.1137/0210021`.

Zhou, Y. & Hao, J.-K. (2017), 'Combining tabu search and graph reduction to solve the maximum balanced biclique problem', *ArXiv*.

Zhou, Y., Rossi, A. & Hao, J.-K. (2018), 'Towards effective exact methods for the maximum balanced biclique problem in bipartite graphs', *European Journal of Operational Research* **269**(3), 834 – 843. URL `http://www.sciencedirect.com/science/article/pii/S0377221718302194`.