Resultados de NP-dificuldade e Aproximação para o Problema da $k-$Floresta Geradora Minimamente Rotulada

Kennedy Corrêa[1,2] *, **Manoel Campêlo**[2]

[1]Instituto de Matemática e Estatística
Universidade de São Paulo (USP) – SP – Brasil

[2]Departamento de Estatística e Matemática Aplicada
Universidade Federal do Ceará (UFC) – CE – Brasil.

kennedycorrea2001@gmail.com, mcampelo@ufc.br

Abstract. *Given an edge-labeled graph G, the minimum labeling spanning k-forest problem consists in finding a spanning forest F of G with minimum number of labels and number of components bounded by k. In this work, we present a proof of the NP-completeness of the problem in split graphs and propose an approximation algorithm for solving it.*

Resumo. *Dado um grafo colorido em arestas G, o problema da k-floresta geradora minimamente rotulada consiste em encontrar uma floresta geradora F de G com número mínimo de cores e número de componentes limitado por k. Neste trabalho, apresentamos uma prova da NP-completude do problema em grafos split e propomos um algoritmo aproximativo para solucioná-lo.*

1. Introdução

Um grafo colorido em aresta (GCA) G consiste de uma tupla (V, E, L, ℓ) onde V é o conjunto de vértices, E é um conjunto de arestas, L é um conjunto de cores/rótulos e $\ell : E \to L$ é uma função que colore cada aresta. Dado um GCA G com número de componentes $c(G) \leq k$, O problema da k-floresta geradora minimamente rotulada (kFGMR) consiste em encontrar uma floresta geradora de G com no máximo k componentes e número mínimo de cores. A tupla do GCA G também pode ser representada por (V_G, E_G, L_G, ℓ_G) e o conjunto de cores de um subgrafo H de G pode ser denotado por $\ell_G(H)$.

Tal problema foi apresentado em [Corrêa and Campêlo 2024], onde foram realizados experimentos acerca de um modelo de programação linear inteira mista para a sua solução. O kFGMR surge então como uma generalização do problema da árvore geradora minimamente rotulada (AGMR) proposto em [Chang and Leu 1997], sendo esse o caso onde $k = 1$. O AGMR foi então estudado e generalizado das mais diversas formas. Entre esses estudos incluem-se a aplicação do problema em telecomunicações ([Chwatal et al. 2007]), o estudo da complexidade parametrizada ([Fellows et al. 2010]) e as extensões, como o problema da árvore de Steiner geradora minimamente rotulada ([Cerulli et al. 2006]) e o problema da árvore ponderada geradora minimamente rotulada ([Cerulli et al. 2024]). Neste trabalho, estamos interessado na complexidade do problema e em propor um método menos custoso de solucioná-lo.

*Trabalho realizado parcialmente enquanto aluno da UFC.

2. Complexidade Computacional

Em [Chang and Leu 1997] é provado que o AGMR é NP-completo para grafos completos, sendo esse o caso $k = 1$, aqui, provamos que o resultado vale para qualquer k inteiro positivo em grafos *split* [West 1996].

Um algoritmo verificador para uma solução do kFGMR pode ser obtido através dos seguintes passos: (i) contagem das cores das arestas, (ii) checagem de que essas arestas não formam ciclo, e (iii) verificação de que o número de componentes definidas é no máximo k. Como todos esses passos podem ser feitos em tempo polinomial, o problema está em NP.

Teorema 2.1. *Para todo inteiro positivo* k*, o* k*FGMR é NP-completo em grafos split.*

Demonstração. Como o kFGMR é NP, basta provar que o AGMR em grafos completos se reduz para o kFGMR em grafos *split*.

Tome k inteiro positivo e seja G um GCA completo com função de coloração ℓ_G. Se $k = 1$, os problemas são equivalentes e não há mais o que fazer. Suponha então $k \geq 2$.

Constrói-se um grafo aumentado G' da seguinte forma (Figura 1):

- $V_{G'} = V_G \cup V'$, onde V' é um conjunto de $k-1$ novos vértices artificiais.
- $E_{G'} = E_G \cup E'$, onde E' é um conjunto de $k-1$ arestas ligando cada vértice de V' a algum vértice de $V(G)$.
- $L_{G'} = L_G \cup L'$, com $L' \cap L(G) = \emptyset$ e $|L'| = k - 1$.
- $\ell_{G'}$ é a extensão de ℓ_G obtida de forma que $e \in E' \implies \ell_{G'}(e) \in L'$ e $e, e' \in E'$ com $e \neq e' \implies \ell_{G'}(e) \neq \ell_{G'}(e')$. Em outras palavras, o grafo aumentado adiciona $k-1$ cores distintas para as novas arestas.

Note que, pela construção, G' é grafo *split* conexo.

Seja $q \in \mathbb{Z}^+$. O objetivo é provar que árvore geradora A de G com $|\ell_G(A)| \leq q$ se, e somente se, existir floresta geradora F de G' com no máximo k componentes e $|\ell_{G'}(F)| \leq q$.

Seja A uma árvore geradora de G satisfazendo $|\ell_G(A)| \leq q$. Construa F simplesmente adicionando os vértices V' em A para que F seja floresta geradora de G', mas sem modificar arestas. Em outros termos, os vértices artificiais são componentes triviais (vértices isolados) em F. Com isso, número de cores não se modifica, e o número de componentes de F será exatamente k, pois há uma componente formada pela árvore original e $k-1$ componentes obtidas pelos vértices artificiais isolados. Dessa forma, F é floresta geradora de G' com $|\ell_{G'}(F)| \leq q$ e $c(F) = k$.

Reciprocamente, seja F uma floresta geradora de G' satisfazendo $c(F) \leq k$ e $|\ell_{G'}(F)| \leq q$. Sem perda de generalidade, pode-se admitir $c(F) = k$ (caso $c(F) < k$ seria suficiente remover arestas de F). Seja k' o número de vértices de V' que são árvores em F e seja F' a subfloresta obtida de F retirando estes vértices. Note que $c(F') = c(F) - k' = k - k'$ e $|\ell_{G'}(F')| = |\ell_{G'}(F)| \leq q$. Além disso, o número de vértices de V' presentes em F' é $(k-1) - k'$, cada um deles sendo folha em F' (já que V' define um conjunto independente em G'). Sendo assim, $F'' = F'[V]$ é uma floresta geradora de G com $c(F'') = c(F') = k - k'$ árvores, $\ell_G(F'') = \ell_{G'}(F'')$ e $|\ell_G(F'')| = |\ell_{G'}(F')| - ((k-1) - k')$.

Como G é conexo, a floresta F'', geradora de G, pode ser estendida para uma árvore geradora A de G adicionando $c(F'') - 1 = (k - k') - 1$ arestas de G. Logo, $|\ell_G(A)| \leq |\ell_G(F'')| + (k - k') - 1 = |\ell_{G'}(F')| - ((k-1) - k') + (k - k') - 1 = |\ell_{G'}(F')| \leq q$.

Com isso, a redução está feita e o kFGMR é NP-completo. $\square$

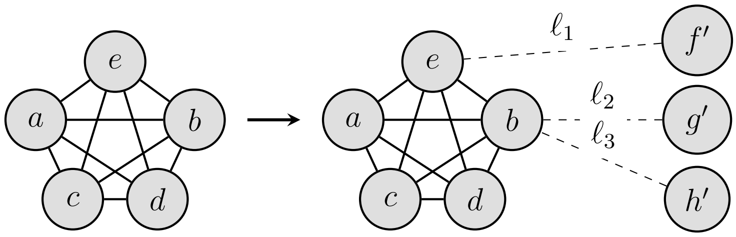

Figura 1. Exemplo da redução para $k = 4$.

3. Algoritmo Aproximativo

Em [Krumke and Wirth 1998] há uma versão revisada do *Maximum Vertex Cover Algorithm* (MVCA) que foi originalmente proposto em [Chang and Leu 1997] para resolver o problema AGMR. Já em [Xiong et al. 2005], há uma prova de que tal algoritmo possui fator de aproximação H_b, onde $b = \max_{l \in L} |\{e \in E : \ell(e) = l\}|$ e $H_b = \sum_{i=1}^{b} \frac{1}{i}$, além disso, é provado que o fator de aproximação é apertado, i.e, existe GCA G tal que a solução obtida pelo MVCA em G é H_b vezes a solução ótima. Dessa forma, propomos o Algoritmo 1, o qual chamamos de MVCA Generalizado. Quando $k = 1$, o algoritmo é equivalente ao MVCA.

Algoritmo 1: MVCA Generalizado

Entrada: Um GCA G com vértices V, arestas E e cores L e um inteiro positivo k.

Seja $C \leftarrow \emptyset$ o conjunto de cores utilizadas;
Seja $H = G[C]$ subgrafo gerador de G com arestas induzidas por C;
enquanto *H tiver mais que k componentes*
  para cada $i \in L - C$ **faça**
    Determine o número de componentes conexas ao inserir todas as arestas de cor i em H;
  fim
  Escolha cor i com o menor número de componentes resultantes;
  $C \leftarrow C \cup \{i\}$;
  $H \leftarrow G[C]$;
fim

Teorema 3.1. *O Algoritmo 1 tem um fator de aproximação H_b para o problema kFGMR.*

A prova desse resultado é bastante longa para ser apresentada neste trabalho, sendo similar a prova dada no trabalho original para o fator de aproximação do MVCA. Ela consiste em limitar as funções de custo de uma sequência de problemas de programação linear que simulam os passos do algoritmo. A diferença em nossa demonstração é considerar que a floresta deverá possuir $|V| - k$ arestas ao invés de $|V| - 1$ arestas como considerado no trabalho original.

Como o fator de aproximação do MVCA é apertado, podemos mostrar que o fator de aproximação do Algoritmo 1 também é apertado para todo k inteiro positivo. Para isso, dado um grafo G e um inteiro k, podemos expandi-lo para um grafo G^* tal que $c(G^*) = k$ e cada componente de G^* seja uma cópia G, mas que os conjuntos de cores de suas componentes sejam disjuntos dois a dois. Um exemplo disso é dado na Figura 2.

Lema 3.2. *Suponha que o problema AGMR em G tenha solução ótima A com valor v. Então uma solução ótima para o problema kFGMR em G^* é obtida escolhendo uma floresta F tal que cada componente de F seja isomorfa à árvore A. O número de cores de tal solução será igual a kv.*

Demonstração. Por construção, F é uma floresta geradora de G^* com $c(F) = k$ e $|L(F)|$ é a soma da quantidade de cores em cada componentes, ou seja, kv. Suponha que exista uma solução F' com $|L(F')| < kv$, então existe uma árvore A' componente de F' tal que $|L(A')| < v$ e A' é cópia de alguma árvore geradora A'' de G. Entretanto, $|L(A'')| = |L(A')| < v = |L(A)|$, contradizendo a minimalidade de A. Logo, F é solução ótima para o kFGMR em G^* com valor igual a kv. $\square$

Teorema 3.3. *Para qualquer inteiro positivo k, o fator de aproximação H_b do Algoritmo 1 para resolver o kFGMR é apertado.*

Demonstração. Seja G um GCA tal que o MVCA retorna uma árvore geradora A de G com número de cores $H_b v$, onde v é o número de cores da solução ótima.

Se $k = 1$, não há o que fazer. Suponha então $k \geq 2$ e construa G^* com k componentes. Como a cor escolhida em cada passo do Algoritmo 1 é aquela que gera o menor número de componentes e considerando que as componentes de G^* possuem conjuntos de cores disjuntos dois a dois, temos que o algoritmo pode utilizar os mesmos passos em cada componente para encontrar uma árvore geradora. Em alguma possibilidade, é gerada uma floresta geradora F de G^* cujas componentes são cópias de A, ou seja, $|\ell_{G^*}(F)| = k|\ell_G(A)| = kH_b v$. Como o Lema 3.2 afirma que o valor ótimo para G^* é kv, a solução F tem número de cores H_b vezes o ótimo, ou seja, o fator de aproximação é alcançado para G^*, sendo então um fator apertado. $\square$

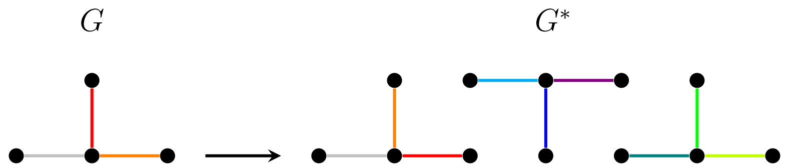

Figura 2. Extensão do GCA G para G^* for $k = 3$.

4. Conclusão

Este trabalho apresentou uma prova da NP-completude do kFGMR e também um algoritmo aproximativo com fator de aproximação estrito para resolver o problema. Futuros trabalhos podem envolver o estudo de novas heurísticas ou a análise da complexidade parametrizada do kFGMR.

Referências

Cerulli, R., Ambrosio, C., Serra, D., and Sorgente, C. (2024). The budgeted labeled minimum spanning tree problem. *Mathematics*, 12(2).

Cerulli, R., Fink, A., and Gentili, M. (2006). Extensions of the minimum labelling spanning tree problem. *Journal of Telecommunications and Information Technology*, 26(4):39–45.

Chang, R.-S. and Leu, S.-J. (1997). The minimum labeling spanning trees. *Inf. Process. Lett.*, 63(5):277–282.

Chwatal, A. M., Raidl, G. R., and Dietzel, O. (2007). Compressing fingerprint templates by solving an extended minimum label spanning tree problem.

Corrêa, L. and Campêlo, M. (2024). O problema da k-floresta geradora minimamente rotulada. In *Anais do LVI SBPO*, Fortaleza.

Fellows, M. R., Guo, J., and Kanj, I. (2010). The parameterized complexity of some minimum label problems. *Journal of Computer and System Sciences*, 76(8):727–740.

Krumke, S. O. and Wirth, H.-C. (1998). On the minimum label spanning tree problem. *Information Processing Letters*, 66(2):81–85.

West, D. (1996). *Introduction to Graph Theory*. Introduction to Graph Theory. Prentice Hall.

Xiong, Y., Golden, B., and Wasil, E. (2005). Worst-case behavior of the mvca heuristic for the minimum labeling spanning tree problem. *Operations Research Letters*, 33(1):77–80.

Este trabalho foi parcialmente financiado por CNPQ (132983/2024-9, 442977/2023-9, 404479/2023-5 e 312417/2022-5) e FUNCAP (ITR-0214-00096.01.00/23).