Conjunto Dominante Justo*

Leonardo Valente Nascimento[1], **Lehilton Lelis Chaves Pedrosa**[1]

[1]Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Campinas – SP – Brazil
l197899@dac.unicamp.br, lehilton@ic.unicamp.br

Abstract. *In the Dominating Set Problem (DS), given a graph G, the goal is to find a subset $D \subseteq V$ of minimum size such that $N[D] = V$. We consider the β-Fair DS Problem (β-FDS), a variant of DS where V is the disjoint union of red and blue vertices, each vertex must be assigned to a dominating vertex in D, and the ratio between the number of vertices of each color assigned to an element of D must be at least β. This problem has applications in fair clustering, which aims to avoid group underrepresentation that is typically present in solutions produced by standard clustering algorithms. We show that β-FDS is $W[1]$-hard when parameterized by $k + tw$ for each $\beta > 0$, where k is the size of the solution, and tw is the treewidth of the graph. For the special case where $\beta = 1$, we present an $\mathcal{O}(n\Delta)$ algorithm for trees and an $\mathcal{O}(3^{tw}\Delta^{2tw+2}n)$ algorithm for general graphs, where Δ is the maximum degree of the graph.*

1. Introdução

No problema do k-center, dado um conjunto de pontos V em algum espaço métrico e um inteiro positivo k, o objetivo é encontrar k pontos, chamados de centros, de modo a minimizar a maior distância entre um ponto e o centro mais próximo. No problema dual, ao invés do número de centros k, a entrada inclui um valor real r e o objetivo é encontrar um menor conjunto de centros de forma que a distância entre cada ponto e o centro mais próximo seja no máximo r. A versão de decisão do problema dual é equivalente ao Problema do Conjunto Dominante (DS) no grafo que tem conjunto de vértices V e uma aresta entre cada par de pontos que estão a uma distância menor ou igual a r [Hochbaum and Shmoys 1985]. Em DS, dado um grafo $G = (V, E)$ e um inteiro k, o objetivo é decidir se há um subconjunto de vértices D com $|D| \leq k$ tal que $N[D] = V$, onde $N[D]$ é o conjunto formado por D e vizinhos de D.

O k-center e outros problemas de clusterização clássicos não impõem restrições sobre tamanho ou distribuição dos pontos de um cluster, o que pode levar a subrepresentação ou viés contra grupos de elementos em certas aplicações [Chierichetti et al. 2017]. Isso motiva o estudo de versões justas dos problemas, em que cada cluster deve conter um número balanceado de pontos de cada grupo. Enquanto a versão justa do k-center já tem recebido relativa atenção na literatura, uma versão justa de DS ainda não foi estudada. Neste trabalho, introduzimos o problema do Conjunto Dominante Justo (β-FDS), que é uma versão justa de DS em que cada vértice está associado a uma cor (vermelho ou azul) e, para cada cluster, a razão entre o número

*O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, projetos números 312271/2023-9, 404315/2023-2, 124332/2024-2, e FAEPEX/Unicamp, projeto número 2422/23.

de vértices da cor menos representada e o número de vértices da cor mais representada deve ser maior ou igual a β.

Neste trabalho, demonstramos que β-FDS é $W[1]$-difícil quando parametrizado por $k + tw$ para cada $\beta > 0$, em que k é o tamanho da solução e tw é a largura arbórea do grafo. Para o caso especial em que $\beta = 1$, apresentamos um algoritmo de tempo $\mathcal{O}(n\Delta)$ para árvores e um algoritmo de tempo $\mathcal{O}(3^{tw}\Delta^{2tw+2}n)$ para grafos gerais, em que Δ é o grau máximo do grafo.

2. Definições

Dado um grafo $G = (V, E)$ e uma bipartição dos vértices em duas cores R e B, i.e., $V = R \cup B$, o *balanceamento* de um conjunto não vazio $S \subseteq V$ é definido como

$$\text{balance}(S) = \frac{\min\{\,|S \cap R|, |S \cap B|\,\}}{\max\{\,|S \cap R|, |S \cap B|\,\}}.$$

O balanceamento captura a representação mínima de uma cor em um cluster e, em particular, se o balanceamento for 1, então cada cor está igualmente representada.

Dizemos que um conjunto $D \subseteq V$ é conjunto dominante *β-justo* se existir um mapeamento $\alpha : V \to D$ tal que $\text{balance}(\alpha^{-1}(d)) \geq \beta$ para todo $d \in D$, i.e., o balanceamento do cluster associado a d é pelo menos β. Para alguma constante β, com $0 < \beta \leq 1$, o problema do *Conjunto Dominante β-Justo* (*β-Fair Dominating Set*, β-FDS) consiste em, dado um grafo $G = (V, E)$, cujos vértices têm cores R ou B, e um inteiro k, decidir se existe conjunto dominante β-justo de tamanho no máximo k.

Note que pode não existir nenhum conjunto dominante β-justo independentemente do tamanho, por exemplo, se $\beta = 1$ e o número de vértices de cada cor não for o mesmo. Observamos que, pelo menos para $\beta = 1$, é possível verificar a existência de algum conjunto dominante β-justo em tempo polinomial (Teorema 2). No algoritmo da Seção 4, iremos supor que existe um conjunto dominante 1-justo e recebemos também uma decomposição em árvore de largura mínima do grafo de entrada. Indicamos [Cygan et al. 2015] para as definições completas.

3. W[1]-dificuldade

Para estudar a dificuldade de β-FDS, iremos descrever uma redução do Problema do Conjunto Dominante Capacitado (CDS). Dado um grafo $G = (V, E)$ e uma função de capacidade $c : V \to \mathbb{N}$, um subconjunto de vértices $D \subseteq V$ é um conjunto dominante capacitado se existir um mapeamento $\alpha : V \to D$ tal que $|\alpha^{-1}(d)| \leq c(d)$ para todo $d \in D$. O problema CDS consiste em, dado um grafo, capacidades nos vértices e um inteiro k, decidir se existe um conjunto dominante capacitado de tamanho no máximo k.

Sabemos que CDS é $W[1]$-difícil quando parametrizado por $k + tw$, em que tw é a largura arbórea do grafo em questão e k é o tamanho da solução [Dom et al. 2008][*]. Em seguida, iremos demonstrar que β-FDS também é $W[1]$-difícil para cada β racional via uma redução parametrizada.

---

[*]Em [Dom et al. 2008], um conjunto dominante capacitado é definido em termos de um mapeamento $\alpha : (V \setminus D) \to D$, mas é conveniente considerar um mapeamento $\alpha : V \to D$ no contexto de particionamento justo e é possível demonstrar que ambas as variantes são $W[1]$-difíceis modificando ligeiramente a redução apresentada no artigo original.

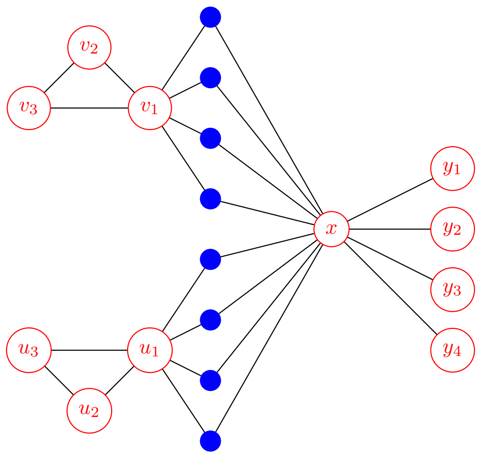

Figura 1. Redução de CDS para β-FDS.

Teorema 1. *Para cada β racional tal que $0 < \beta \leq 1$, β-FDS é $W[1]$-difícil quando parametrizado por $k + tw$, sendo tw a largura arbórea do grafo de entrada e k o tamanho da solução.*

Ideia da demonstração. Como β é racional entre 0 e 1, podemos escrevê-lo como uma fração irredutível $\frac{p}{q}$, onde p e q são constantes inteiras positivas e $p \leq q$. Dada uma instância (G, c, k) de CDS, em que $G = (V, E)$ é um grafo de largura arbórea tw, vamos construir uma instância (G', R, B, k') de β-FDS em que $G' = (V', E')$ é um grafo com largura arbórea no máximo $tw' = q \cdot tw + 1$ e $k' = k + 1$ (veja Figura 1).

Para cada vértice $v \in V$, crie uma clique com q vértices em R (vermelhos) denotados por $v_1, v_2, \ldots, v_q$. Além disso, para cada aresta $(u, v) \in E$, adicione uma aresta entre u_i e v_j para todo $1 \leq i, j \leq q$. Observe que, nesse momento, os vértices de cada clique são vizinhos gêmeos e G' é igual ao produto entre G e K_q.

Agora, para cada vértice $v \in V$, crie $p \cdot c(v)$ vértices em B (azuis) denotados por $v'_1, v'_2, \ldots, v'_{p \cdot c(v)}$ e ligue cada vértice v'_i ao vértice v_1, que é a primeira cópia do vértice v na clique. Defina $C = \sum_{v \in V} c(v)$ e observe que o número de vértices azuis criados é $p \cdot C$.

Crie um vértice x vermelho e ligue esse vértice a cada um dos vértices azuis, i.e., cada vértice v'_i para $v \in V$ e $1 \leq i \leq p \cdot c(v)$. Finalmente, crie $q \cdot (C - |V|) - 1$ vértices vermelhos, denotados por $y_1, y_2, \ldots, y_{q \cdot (C - |V|) - 1}$, e ligue cada um deles a x. Isso completa a construção de G'.

Para completar a demonstração, é possível mostrar que cada solução para a instância de CDS corresponde a uma solução de β-FDS e vice-versa. A observação principal é que o número de vértices vermelhos é exatamente $q \cdot C$ enquanto o número de vértices azuis é exatamente $p \cdot C$. Isso implica que, para um conjunto dominante β-justo D com mapeamento α, temos balance$(\alpha^{-1}(d)) = \beta$ para cada vértice $d \in D$. Segue que cada cluster tem um múltiplo de q vértices vermelhos e um múltiplo de p vértices azuis. Além disso, é possível verificar que existe uma solução em que o número de vértices de G associados a cada d é no máximo $c(d)$. $\square$

4. Algoritmos para $\beta = 1$

Nesta seção, lidamos com o caso particular em que $\beta = 1$. Neste caso, é conveniente representar a cor de um vértice por meio de uma coloração $\chi : V \to \{R, B\}$ e considerar uma função f com $f(R) = 1$ e $f(B) = -1$, de forma que um conjunto S terá balance$(S) = 1$ se e somente se $\sum_{v \in S} f(\chi(v)) = 0$.

Essa definição tem algumas propriedades interessantes, que permitem verificar se um conjunto é balanceado sem necessariamente armazenar quantos elementos de cada cor são dominados por cada vértice. Em seguida, definimos $n = |V|$ e $m = |E|$.

Teorema 2. *Existe algoritmo que decide se um grafo $G = (V, E)$ com coloração $\chi : V \to \{R, B\}$ admite conjunto dominante 1-justo em tempo $\mathcal{O}(n^{5/2} + m)$.*

Este teorema mostra que a dificuldade do problema não está apenas em achar um conjunto que satisfaça as condições de justiça, mas sim em garantir que esse conjunto tenha tamanho menor ou igual a k. Em seguida, mostramos que o problema é FPT quando parametrizado por $tw + \Delta$, sendo Δ o maior grau entre os vértices do grafo.

Teorema 3. *Existe algoritmo que resolve 1-FDS tempo $\mathcal{O}(3^{tw}\Delta^{2tw+2}n)$ para grafos de grau no máximo Δ quando equipado com uma decomposição em árvore de largura tw.*

Para árvores, obtemos um algoritmo especializado, também baseado em programação dinâmica.

Teorema 4. *Existe algoritmo que resolve 1-FDS para árvores em tempo $\mathcal{O}(n\Delta)$.*

O algoritmo considera, para cada vértice v, um subproblema de encontrar um melhor cluster $\alpha^{-1}(v)$ tal que $\sum_{u \in \alpha^{-1}(v)} f(\chi(u)) = 0$. Isso pode ser feito resolvendo-se uma instância do Problema da Mochila 0-1 em tempo $\mathcal{O}(\delta^2(v))$, em que $\delta(v)$ é o grau de v. Assim, o problema completo pode ser resolvido em tempo $\mathcal{O}(\sum_{v \in V} \delta^2(v)) = \mathcal{O}(n\Delta)$.

Também, é possível obter um algoritmo aleatorizado baseado no Teorema 4, resolvendo-se cada subproblema em tempo $\mathcal{O}(\delta^{3/2}(v))$ com alta probabilidade de acerto, o que leva a uma complexidade de $\mathcal{O}(n\sqrt{\Delta})$ no total. Para isso, embaralhamos a ordem das arestas da árvore, de forma que as somas parciais de $f(\chi(u))$ para um cluster $\alpha^{-1}(v)$ ótimo se assemelhe às distâncias percorridas por um passeio aleatório de $\delta(v)$ passos. A esperança do maior valor absoluto de um passeio aleatório de k passos é $\Theta(\sqrt{k})$ [Palacios 2008], então podemos restringir cada instância do problema da mochila correspondente para que considere apenas valores entre $-c\sqrt{\Delta}$ e $c\sqrt{\Delta}$ para alguma constante c. Não conseguimos cotas satisfatórias para a probabilidade de acerto até o momento, mas em experimentos para $c = 2$ obtivemos chance de erro menor do que 1%.

Referências Bibliográficas

Chierichetti, F., Kumar, R., Lattanzi, S., and Vassilvitskii, S. (2017). Fair clustering through fairlets. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Cygan, M., Fomin, F., Kowalik, L., Lokshtanov, D., Marx, D., Pilipczuk, M., Pilipczuk, M., and Saurabh, S. (2015). *Parameterized Algorithms*. Springer.

Dom, M., Lokshtanov, D., Saurabh, S., and Villanger, Y. (2008). Capacitated domination and covering: a parameterized perspective. In *Proceedings of the 3rd International Conference on Parameterized and Exact Computation*, IWPEC'08, page 78–90, Berlin, Heidelberg. Springer-Verlag.

Hochbaum, D. and Shmoys, D. (1985). A best possible heuristic for the k-center problem. *Mathematics of Operations Research*, 10:180–184.

Palacios, J. L. (2008). On the simple symmetric random walk and its maximal function. *The American Statistician*, 62(2):138–140.