Kernelization of vertex cover based on crown structure

U Divya
Intel Technology India Pvt. Ltd.
Bangalore, India
udivya73@gmail.com

***Abstract*—Kernelization is the technique of preprocessing a problem instance to reduce it to a smaller instance. In this paper we discuss the kernelization technique for a graph problem called vertex cover. This technique is based on a graph structure known as crown structure. The current best known solution for kernelization of vertex cover based on crown structure reduces the problem instance to a smaller instance of size $3k$, where k is the size of vertex cover. The solution discussed in this paper reduces the problem instance to an instance of size $2k$. Vertex cover has applications in computer network security(worm propagation) and machine learning(text summarization).**

***Index Terms*—Vertex cover, fixed parameter tractable, kernelization, Gallai Edmonds decomposition**

I. INTRODUCTION

NP-hard problems are a class of problems that are expected to have non-polynomial time complexity, that is the running time of the solution will be an exponential function, or worse, on the size of the problem. Vertex cover problem is one such NP-hard problem. Vertex cover of a graph, say S, is a subset of vertices of the graph such that all edges of the graph have at least one end-point in S. Minimum vertex cover is the the vertex cover of the smallest size among all possible vertex covers. Finding a minimum vertex cover find its applications in keyword based text summarization [1] and also in simulating propagation of computer worms on a network and designing techniques to prevent them [2].

In search for more efficient solution for NP-hard problems Downey and Fellows [3] began to study various hard problems to find if the size is the cause of the exponential factor in the time complexity or some other parameter causes the exponential explosion in the time complexity. This domain of research is called the study of *fixed parameter tractable* problems [6]. To understand the main idea behind this research is to find such solutions which have time complexity of the form $O(poly(size).\alpha^{parameter})$. Here *size* stands for the problem instance size, $poly()$ denotes a polynomial function and *parameter* denotes some parameter(s) other than size on which the problem depends on. This parameter need not be large even for the large size problem instance. For example, for vertex cover, size of the vertex cover can be a parameter. Any solution that has complexity which is polynomial in size of graph and exponential in size of vertex cover will be a fixed parameter tractable solution. So for graphs having vertex cover of small size, this fixed parameter tractable solution gives an efficient solution for vertex cover.

Kernelization is the technique in which the problem instance is reduced in polynomial time to a smaller instance(kernel) whose size is a polynomial function of some non-size parameter, i.e., the size of problem instance itself reduces to $p(parameter)$ where $p()$ is a polynomial function [7]. This way any regular solution will have time complexity $O(poly(size) + f(p(parameter)))$ where f is a non-polynomial function.

In this paper, first the best known solution for kernelization of vertex cover using crown structure is explained. Then the proposed solution is discussed and compared with the best known solution. At the end scope for future work has been discussed along with concluding remarks.

II. LITERATURE REVIEW

A. *Notations and definitions*

We denote a graph as $G(V, E)$ where V is the vertex set and E is the edge set for the graph. We denote a bipartite graph as $G(V_1, V_2; E)$ where V_1 and V_2 are two parts of the vertex set. We will use $G[V']$ to denote the subgraph of G induced on V' where $V' \subseteq V$. We will use $N(S)$ to denote the neighbors of a vertex set S, which do not belong to S. It is called the *open neighbourhood* of S. The *closed neighbourhood* of a vertex set S, denoted by $N[S]$, is the set of vertices which are either in S or in the neighbourhood of some S vertex. If S is a singleton $\{x\}$, then we denote these by $N(x)$ and $N[x]$. By $deg(x)$ we denote the size of $N(x)$. A bipartite graph is a graph with two sets of vertices such that no edge exists between the vertices of same set [9]. A matching in graph is a subset of edges such that all vertices have at least one incident edge in the subset [8]. Maximum matching is the subset of largest size among all possible matchings. Independent set is a set of vertices such that no edge exists between the vertices of the set [10].

B. *Gallai Edmonds decomposition*

Gallai-Edmonds decomposition is partitioning the vertices of a graph into three subsets that satisfy certain properties [4]. One subset, say D, is the set of vertices that will not be present in any vertex cover of the graph. Subset A will be the set of vertices that are present in all vertex covers of the graph. Subset C will be the set of remaining vertices.

C. Known result

A crown structure is said to form from 2 disjoint subsets of V, say I and H, if the following conditions hold :

1) I is an independent set
2) $H = N(I)$ where $N(I)$ is the set of all neighbors of I
3) Edges between H and I contain a matching which matches all vertices in H.

Theorem II.1. *There exists a minimum vertex cover of G that contains all vertices of H and no vertices from I.*

Proof. Let C be a minimum vertex cover of G. Let $I_c = I \cap C$ and $\bar{I}_c = I \backslash C$. Let $H_c = N(I_c)$ and $\bar{H}_c = N(\bar{I}_c)$. Since the vertices in $\bar{I}_c$ are not in the vertex cover, vertices from $\bar{H}_c$ are in the vertex cover C to cover the edges incident on $\bar{I}_c$. Since $N(H \backslash H_c) \subseteq I_c$ and there exists a matching in which every H vertex can be matched to an I vertex, we can deduce that $|H \backslash H_c| \leq |I_c|$. So $|C \cap (I \cup H)|$ = $|I_c| + |H_c| \geq |H \backslash H_c| + |H_c| = |H|$. We also know that I is an independent set, so $C' = (C \backslash I) \cup H$ is also a vertex cover. Then $C' \backslash \{I \cup H\} = C \backslash \{I \cup H\}$ and $C' \cap \{I \cup H\} = H$. So $|C'| \leq |C|$. Hence C' is also optimal. □

We can obtain the sets I and H in the following way. Let M_1 be a maximal matching for G. Let U be the set of unmatched vertices in M_1. Let $G[U \cup N(U)]$ be the bipartite graph induced on $U \cup N(U)$. Let M_2 be a maximum matching for the bipartite graph.

Theorem II.2. *Let U' be the set of unmatched vertices in M_2. Let U" be the set of vertices that are reachable from an unmatched vertex by an alternating path of even length in M_2. Then $I = \{U' \cup U"\} \cap U$ and the neighbors of I in G, say H, form a crown structure in G.*

Proof. Let $B(V_1, V_2; E)$ be any bipartite graph. Let D', A', C' form the Gallai-Edmonds decomposition of B. Let $D_1 = D' \cap V_1$ and $A_2 = A' \cap V_2$. The properties of Gallai-Edmonds decomposition [4] implies that set D_1 consists of all unmatched vertices in V_1 in a maximum matching M and all the vertices reachable from an unmatched vertex in V_1 by an even length alternating path in the maximum matching M. It also implies that set D_1 is an independent set and the vertices of the set A_2, that contains all neighbors of D_1, are always matched in all maximum matchings and a matching edge incident on a vertex of A_2 is always of the form (a, d) where $a \in A_2$ and $d \in D_1$. Hence $I = D_1$ and $H = A_2$ satisfies all three conditions required to form a crown structure. □

Hence vertices from set I and H can be removed from consideration and the problem instance can be reduced to (G', k') where $k' = k - |H|$ and $G' = G(V', E')$ where $V' = V \backslash (I \cup H)$ and $E' = \{(x, y) \in E | x, y \in V'\}$. Since the number of edges in a maximum matching is less than or equal to number of vertices in minimum vertex cover, M_1 and M_2 can have at most k edges and $2k$ vertices. Let $n = |V|$. So $|U| \leq n - 2k$, which implies that $|I| \leq n - 2k - k = n - 3k$ because at most k vertices from H can be matched by M_2. Hence $|V'| \leq n - (n - 3k) = 3k$. Thus the reduced instance has at most $3k$ vertices.

III. Proposed solution

A. Reduction

Let U be any independent set in G. Consider a bipartite graph B on the vertex sets U and $N(U)$, the neighbours of U, and edge set being the edges running between them denoted by E_B. Note that there may be edges between $N(U)$ vertices which are clearly not a part of B. Let S'_1, S'_2, S'_3 be the Gallai-Edmonds decomposition of the vertex set of B. In any maximum matching of B there are $|S'_3|/2$ edges which mutually match all the S'_3 vertices and $|S'_2|$ edges which match all the S'_2 vertices to S'_1 vertices, leaving $|S'_1| - |S'_2|$ vertices unmatched. Thus within $U \cup N(U)$ there must be at least $|S'_2| + |S'_3|/2$ vertices in any vertex cover of G just to cover the edges running between U and $N(U)$.

Let $S'_{iu} = S'_i \cap U$ and $S'_{in} = S'_i \cap N(U)$. The edges in E_B can be partitioned into E_{12} running between S'_{1u} and S'_{2n}; E_{21} between S'_{2u} and S'_{1n}; E_{22} between S'_{2u} and S'_{2n}; E_{23} between S'_{2u} and S'_{3n}; E_{32} between S'_{3u} and S'_{2n}; and E_{33} between S'_{3u} and S'_{3n}. Let $\delta_G(X)$ denote the set of edges with one end in the vertex set X and the other outside it. From the properties of Gallai-Edmonds decomposition we have facts:

(i) $\delta_G(S'_{1u} \cup S'_{3u}) \subset \delta_G(S'_{2n} \cup S'_{3n})$,

(ii) $|S'_{1u} \cup S'_{3u}| \geq |S'_{2n} \cup S'_{3n}|$,

(iii) In any maximum matching of B there are $|S'_{2n} \cup S'_{3n}|$ matching edges incident on $S'_{2n} \cup S'_{3n}$.

Proof. (i) Set U is an independent set so every edge incident on S'_{1u} is also incident on S'_{2n} and every edge incident on S'_{3u} has its other end in either S'_{3n} or in S'_{2n}.

(ii) and (iii) In any maximum matching of B every vertex in S'_{2n} is matched to some vertex in S'_{1u} so $|S'_{2n}| \leq |S'_{1u}|$ and vertices of S'_{3u} and S'_{3n} are mutually matched so $|S'_{3u}| = |S'_{3n}|$. □

Corollary III.0.1. *If C is a vertex cover of G, then $C' = (C \setminus (S'_{1u} \cup S'_{3u})) \cup S'_{2n} \cup S'_{3n}$ is also a vertex cover and $|C'| \leq |C|$.*

Proof. Consider any maximum matching M of B. So M is also a matching in G. From (iii) we know that M has $|S'_{2n} \cup S'_{3n}|$ edges between $S'_{2n} \cup S'_{3n}$ and $S'_{1u} \cup S'_{3u}$. So C must have at least $|S'_{2n} \cup S'_{3n}|$ vertices from $S'_{1u} \cup S'_{3u} \cup S'_{2n} \cup S'_{3n}$. From (i) $S'_{2n} \cup S'_{3n}$ covers all the edges that any subset of $S'_{1u} \cup S'_{3u} \cup S'_{2n} \cup S'_{3n}$ can cover so C must also be a vertex cover. Also from (ii) $|C'| \leq |C|$. □

This result gives a reduction for the problem of minimum vertex cover of a graph.

Reduction Steps:

(i) Compute the Gallai-Edmonds decomposition of the given graph $G = (V, E)$ into sets S_1, S_2, S_3.

(ii) Let S_1^1 denote the set of isolated vertices of the induced graph $G[S_1]$.

(iii) Taking $U = S_1^1$, compute the bipartite graph B.

(iv) Compute $S'_{1u}, S'_{3u}, S'_{2n}, S'_{3n}$.

(v) Determine the graph $G' = G[V \backslash (S'_{1u} \cup S'_{3u} \cup S'_{2n} \cup S'_{3n})]$. From the corollary we have immediate result.

Lemma III.1. *If C' is a minimum vertex cover of G', then $C' \cup S'_{2n} \cup S'_{3n}$ is a minimum vertex cover of G.*

B. A $2k$ Kernel

Given any graph G we can perform the above reduction until it does not reduce any further. This will happen when $S'_{2u} = U = S_1^1$ and $S'_{3n} = N(U) = N(S_1^1)$. At this stage we know that there exists a maximum matching of B which matches all the vertices of S_1^1 to vertices in $N(S_1^1)$. So the reduced graph has a matching in which all the S_1^1 vertices are matched. Now we will show that the number of vertices in this graph is at most 2 times the number of vertices in the smallest vertex cover for it.

Lemma III.2. *Let $G = (V, E)$ be a graph and S_1, S_2, S_3 be its Gallai Edmonds decomposition. Let S_1^1 be the set of isolated vertices of $G[S_1]$. If there exists a matching of G in which all the S_1^1 vertices are matched, then $|V| \leq 2|C|$ where C is a minimum vertex cover of G.*

Proof. We will denote the set of vertices in the larger components (of size 3 or more) in $G[S_1]$ by S_1^3. So $S_1 = S_1^1 \cup S_1^3$.

Let M_0 be a matching in G in which all the S_1^1 vertices are matched. Starting from M_0 we can apply Edmonds' algorithm to compute a maximum matching M of G. Even after extending a matching by $M" = M' \Delta P$ where P is an augmenting path, the set of matched vertices in M' remain matched in $M"$. Hence every S_1^1 vertex remains matched in M. So G has a maximum matching M in which all the S_1^1 vertices are matched.

Let us consider two induced subgraphs on two disjoints subsets of vertices: $G_1 = G[S_1 \cup S_2]$ and $G_3 = G[S_3]$. Let c_1 and c_3 be the sizes of the smallest vertex covers of G_1 and G_3 respectively. Then the size of the smallest vertex cover of G must be $|C| \geq c_1 + c_3$. Now we will estimate lower bounds for c_1 and c_3 respectively in terms of the number of vertices in these graphs.

Let p^1 be the number of isolated vertices in $G[S_1]$ and p^3 be the number of larger components in $G[S_1]$. Also let n_1^1, n_1^3 denote the number of vertices in these type of components respectively. So $p^1 = n_1^1$ and $n_1^1 + n_1^3 = |S_1|$.

The connected components of $G[S_1]$ are near-perfectly matchable. If any such component has $2r + 1$ vertices for $r \geq 1$, then any vertex cover will require at least $r + 1$ vertices to cover all the internal edges of the component.

Let us split c_1 into 2 parts. c_1^1 denotes the number of vertices to cover the edges incident on S_1^1. So these vertices are either in S_1^1 or in S_2. c_1^3 denotes the number of cover-vertices inside the larger components to cover only the internal edges of these components only.

Since in M all S_1^1 vertices are matched so $c_1^1 \geq n_1^1$. For larger components $c_1^3 \geq (n_1^3 + p^3)/2$. We know that in every maximum matching all S_2 vertices are matched to S_1 vertices such that at most one vertex per component is matched to some S_2 vertex. So $|S_2| \leq p^1 + p^3$. So $|S_1| + |S_2| \leq n_1^1 + n_1^3 + p^1 + p^3$. We have $n_1^1 = p^1$ so $|S_1| + |S_2| = 2n_1^1 + 2(n_1^3 + p^3)/2 \leq 2(c_1^1 + c_1^3) \leq 2c_1$. The number of matching edges in S_3 is exactly $|S_3|/2$ so $|S_3| \leq 2c_3$. So $|V| = |S_1| + |S_2| + |S_3| \leq 2(c_1 + c_3) \leq 2|C|$. This implies that the reduced instance will have at most $2k$ vertices. $\square$

IV. Conclusion and Future Scope

Kernelization helps in designing efficient fixed parameter tractable solutions for NP-hard problems. Kernelization of vertex cover is a significant issue because of the numerous applications of vertex cover. Here one such approach was discussed based on crown structure which is an improvement on the existing best known approach based on crown structure. There is scope of further improvement by utilizing the properties of each partition obtained by Gallai-Edmonds decomposition.

References

[1] Atowar-Ul Islam and Bichitra Kalita, Application of Minimum Vertex Cover for Keywordbased Text Summarization Process, International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 1 (2017), pp. 113-125

[2] Eric Filiol, Edouard Franc, Alessandro Gubbioli, Benoit Moquet and Guillaume Roblot, Combinatorial Optimisation of Worm Propagation on an Unknown Network, Proc. World Acad. Science, Engineering and Technology, Vol 23, August 2007, http://www.waset.org/journals/waset/v34/v34-8.pdf

[3] Rodney G Downey and Michael Ralph Fellows.Parameterized complexity.Springer Science Business Media, 2012.

[4] Douglas B.West,A short proof of the BergeTutte Formula and the GallaiEdmonds Structure Theorem, European Journal of Combinatorics Volume 32, Issue 5, July 2011, Pages 674-676.

[5] Faisal N Abu-Khzam et al. Kernelization Algorithms for the Vertex CoverProblem: Theory and Experiments. In:ALENEX/ANALC69 (2004).

[6] Christian Sloper and Jan Arne Telle, An Overview of Techniques for Designing Parameterized Algorithms, The Computer Journal, Volume 51, Issue 1, January 2008, Pages 122136, https://doi.org/10.1093/comjnl/bxm038

[7] J Guo, R Niedermeier, Invitation to data reduction and problem kernelization, ACM SIGACT News, 2007 - dl.acm.org

[8] Z Galil, Efficient algorithms for finding maximum matching in graphs, ACM Computing Surveys (CSUR), 1986

[9] H Zha, X He, C Ding, H Simon, M Gu, Bipartite graph partitioning and data clustering, Proceeding CIKM '01 Proceedings of the tenth international conference on Information and knowledge management Pages 25-32

[10] M Luby,A simple parallel algorithm for the maximal independent set problem, SIAM journal on computing, 1986 - SIAM