γ-Cluster Edge Modification Problems

Antoine Castillon — Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France and Univ Lyon, UCBL, CNRS, INSA Lyon, LIRIS, UMR 5205, F-69622, France
Julien Baste — Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
Clarisse Dhaenens — Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
Mohammed Haddad — Univ Lyon, UCBL, CNRS, INSA Lyon, LIRIS, UMR 5205, F-69622 Villeurbanne, France
Hamida Seba — Univ Lyon, UCBL, CNRS, INSA Lyon, LIRIS, UMR 5205, F-69622 Villeurbanne, France

Abstract

We introduce the γ-Cluster Edge Modification problems, which are variants of the Cluster Editing problem, and defined as: given a graph G, how many edges must be edited (deleted or added) in order to have a disjoint union of quasi-cliques. We provide the complexity classification of most of these problems, and present results on the approximability of some problems as well as a dynamic programming algorithm based on the tree-decomposition of the input graph.

1 Background and motivation

A graph modification problem with respect to a property $\mathcal{P}$ can be defined as follows: given a graph G, find a minimum set of modifications on G such that the modified graph verifies $\mathcal{P}$. These modifications depend on the problem itself and usually are edge deletions, edge additions, or vertex deletions. Graph modification problems have been vastly studied. For instance, in [8, 10] the authors prove that for any non trivial hereditary property $\mathcal{P}$, the $\mathcal{P}$ Vertex Deletion problem is NP-complete, and hard to approximate. Edge modification problems have been studied extensively in [9] which proves the NP-hardness of several problems and gives some approximation algorithms. More specific results have been found for the Cluster Editing/Deletion problems such as kernalizations [5] and better approximations [6]. The properties associated with these problems, are often characterized by a set of forbidden induced subgraphs or minors and thus are hereditary. Such properties impose a strict structure on the modified graph and especially on its small induced subgraphs. Other graph modification problems using non hereditary properties also exist in the literature. For instance, [11] introduced (p, q)-cluster graphs where at most q edges leave each cluster and each cluster misses at most p edges to be a clique. In this work, we introduce similar variants using quasi-cliques allowing in each cluster a number of missing edges relative to the size of that cluster.

Quasi-cliques are a natural way to extend the definition of Cliques to dense graphs. There are usually two ways to formally define a quasi-clique: one focusing on the degree of each node, and the other on the density of edges.

Definition 1 (Degree-based quasi-clique)**.** Given $\gamma \in [0, 1]$, a graph $G = (V, E)$ is a γ-degree-based quasi-clique, if for all $u \in V$, $d(u) \geq \gamma(|V| - 1)$, where $d(u)$ is the degree of u.

Definition 2 (Density-based quasi-clique)**.** Given $\gamma \in [0,1]$, a graph $G = (V, E)$ is a γ-density-based quasi-clique, if $2|E| \geq \gamma|V|(|V|-1)$.

Also, given a graph $G = (V, E)$, we say that a set of vertices $X \subseteq V$ is a quasi-clique of G if $G[X]$, the subgraph induced by X, is a quasi-clique. Various aspects of quasi-cliques have been studied with both definitions. For instance, Maximal Induced Quasi-Clique, which is equivalent to Quasi-Clique Vertex Deletion, is known to be NP-complete for both definitions [1, 2]. Also, several effective enumeration algorithms, or top-k enumeration algorithms, have been developed [3, 4].

Note that quasi-cliques encapsulate well the concept of dense subgraphs without imposing any structure on the small induced subgraphs. Indeed, parameter $\gamma \in [0,1]$ describes how dense the graph is and how close it is to an actual clique and the property "G is a quasi-clique" is not hereditary for both definitions. A notion of "weak heredity", described in Lemma 1, exists for density-based quasi-clique [3] but no such result exists for degree-based quasi-cliques.

Lemma 1. Let $G = (V, E)$ be a graph, X a γ-density-based quasi-clique of size $|X| = k$. For all $k' \leq k$ exists $X' \subseteq X$ a γ-density-based quasi-clique of size k'.

Moreover, every graph can be an induced subgraph of a quasi-clique. Indeed, one can see that, given $G = (V, E)$ and $\gamma \in [0, 1[$, there exists a graph G' on $\left\lceil \frac{1}{1-\gamma}|V| \right\rceil$ vertices such that G' is a γ-density-based (resp. degree-based) quasi-clique and G is an induced subgraph of G'. If these properties tend to prove that quasi-cliques are indeed interesting dense graphs which do not impose any peculiar structure on their small induced subgraphs, it is important to note that they also prove that it is much harder to find quasi-cliques in practice. Furthermore, other known results tend to confirm it. For instance [3] prove that contrary to cliques, it is NP-hard to decide whether a density-based quasi-clique containing a given set of vertices exists. Note that the same results hold for degree-based quasi-cliques. Also, [4] proved that it is NP-hard to decide whether a given degree-based quasi-clique is maximal.

We focus our study on the quasi-clique variants of the Cluster Editing problem. Such variants can be defined as follows:

Problem 1 (γ-Degree Cluster Editing)**.** Given $G = (V, E)$, find a set $S \subseteq \binom{V}{2}$ of minimum size such that the connected components of $(V, E \Delta S)$ are γ-degree-based quasi-cliques, with $\binom{V}{2}$ being the set of pairs of V and Δ being the symmetric difference.

We define similarly the γ-Degree Cluster Deletion and Completion problems as well as the γ-Density Cluster Editing, Deletion and Completion problems. We sometimes refer to these problems as the γ-Cluster Editing/Deletion/Completion problems if the result discussed holds for both definitions of quasi-cliques. This paper proposes a study of these problems with respect to complexity, approximability and parameterized complexity.

2 Complexity Classification

First, it is important to note that all of these problems are trivially in NP. The complexity classification of these problems is highly similar to the complexity classification of the classic clustering problems. Indeed, we prove that the γ-Cluster Editing and γ-Cluster Deletion problems are NP-complete with both definitions of a quasi-clique.

Theorem 1. For $\gamma > \frac{1}{2}$, the γ-Cluster Editing and γ-Cluster Deletion problems are NP-complete with both definitions of a quasi-clique.

Also, similarly to the Cluster Completion problem, we prove that given a graph G and S an optimal solution of the γ-Degree Cluster Completion problem, S does not contain any edge linking two different connected components of G. Hence, this problem can be solved on each connected component of the input graph separately. Since the γ-Degree Cluster Completion problem can be solved in polynomial time on connected graphs, it can be solved in polynomial time in the general case.

Theorem 2. For $\gamma > \frac{1}{2}$, the γ-Degree Cluster Completion is in P.

However, the same remark does not hold for density-based quasi-cliques. Indeed, sometimes the optimal solution of the γ-Density Cluster Deletion problem involves edges between different connected components. We manage to prove that this problem is harder than the Unary Knapsack problem, and it seems to be very similar to a variation of this problem with multiple knapsacks. However, its complexity remains an open question.

3 Approximation

It is known that the Cluster Deletion problem with a fixed number of clusters k is NP-hard to approximate within any constant factor $c > 1$. Indeed, finding a solution to this problem is equivalent to find a solution of the k-Coloring problem, and thus is NP-hard. We obtain similar results for the γ-Cluster Deletion problems with fixed number of clusters described in Theorem 3. Note that the problem is NP-hard even with $k = 2$, showing again the difficulty brought by quasi-cliques.

Problem 2 ((γ, k)-Cluster Deletion)**.** Given $G = (V, E)$ and $k \in \mathbb{N}$, find a set $S \subseteq \binom{V}{2}$ of minimum size such that $(V, E \backslash S)$ has at most k connected components which are all γ-quasi-cliques.

Theorem 3. For $\gamma > \frac{2}{3}$, and $k \geq 2$, the (γ, k)-Cluster Deletion problem is NP-hard to approximate within any constant factor $c > 1$ with both definitions of quasi-clique.

Several approximation algorithms involving linear programming [7] or combinatorial approaches [6] are known for the Cluster Editing problem. However, these approaches use the property that a cluster graph does not contain any induced P_3. Unfortunately, this property cannot be adapted to the γ-Cluster Editing problem which implies a more complex structure of the modified graph. Finding approximation algorithms for the γ-Cluster Editing and the γ-Cluster Deletion problems will be the object of further work.

4 Parameterized Complexity

Since trees are really sparse graphs and the solution on these graphs is usually simple, implying the deletion of most of the edges, we decide to first use the treewidth, which measures the distance between the graph and a tree, to parameterize our problems. We obtain an XP algorithm parameterized by treewidth for the γ-Cluster Editing and the γ-Cluster Deletion problems. The same algorithm becomes FPT algorithms when the number of quasi-cliques or the size of the largest quasi-clique is bounded.

Theorem 4. For $\gamma > \frac{1}{2}$, γ-Cluster Editing and γ-Cluster Deletion can be solved in time $n^{O(\mathsf{tw})}$, where n is the size of the input graph and tw its treewidth.

Corollary 1. For $\gamma > \frac{1}{2}$, γ-Cluster Editing and γ-Cluster Deletion can be solved in time $2^{O(\mathsf{tw})}n^{O(1)}$ if the number of quasi-cliques is bounded, or in time $\mathsf{tw}^{O(\mathsf{tw})}n^{O(1)}$ if the size of the largest quasi-clique is bounded, where n is the size of the input graph and tw its treewidth.

Since several polynomial and even linear kernalizations with respect to the number of modifications are already known for the Cluster Editing and the Cluster Deletion problems [5], we except similar results for our problems. That will also be the object of further work.

5 Concluding Remarks

This paper introduces several new edge modification problems which are based on quasi-cliques and thus cannot be defined by an hereditary property contrary to prior graph modification problems. If this fact can remove constraints on the modified graph, it also complicates the problem. We provide the complexity of most of the introduced problems along with a result on the difficulty of the approximation of such problems and a dynamic programming algorithm based on the tree decomposition of the input graph.

Acknowledgement: This work is supported by Agence Nationale de la Recherche (ANR-20-CE23-0002).

References

[1] Pattillo Jeffrey, Veremyev Alexander, Butenko Sergiy and Boginski Vladimir. On the maximum quasi-clique problem. *Discrete Applied Mathematics*. 2013.

[2] Pastukhov Grigory, Veremyev Alexander, Boginski Vladimir and Prokopyev Oleg. On maximum degree-based γ-quasi-clique problem: Complexity and exact approaches. *Networks*. 71. 2017.

[3] Takeaki Uno. An efficient algorithm for enumerating pseudo cliques. In *Proc. of the 18th int. conf. on Algorithms and computation* (ISAAC'07), pp. 402–414. 2007.

[4] Seyed-Vahid Sanei-Mehri, Apurba Das and Srikanta Tirthapura. Enumerating Top-k Quasi-Cliques. *IEEE Int. Conf. on Big Data (Big Data)*, pp. 1107-1112. 2018.

[5] Gramm Jens, Guo Jiong, Hüffner Falk and Niedermeier Rolf. Graph-Modeled Data Clustering: Fixed-Parameter Algorithms for Clique Generation. *Theor Comput Syst*, 38. 2003.

[6] Nikhil Bansal, Avrim Blum, and Shuchi Chawla. Correlation Clustering. In *Proc. of the 43rd Symposium on Foundations of Computer Science (FOCS '02)*. IEEE Computer Society, USA, 238. 2002.

[7] Moses Charikar, Venkatesan Guruswami and Anthony Wirth. Clustering with qualitative information, *Journal of Computer and System Sciences*, Volume 71, Issue 3, pp. 360-383. 2005.

[8] Mihalis Yannakakis. Node-and edge-deletion NP-complete problems. In *Proc. of 10th annual ACM symposium on Theory of computing (STOC '78)*, pp. 253–264. 1978.

[9] Assaf Natanzon, Ron Shamir and Roded Sharan, Complexity classification of some edge modification problems, *Discrete Applied Mathematics*, Volume 113, Issue 1, pp. 109-128. 2001.

[10] Carsten Lund and Mihalis Yannakakis. The Approximation of Maximum Subgraph Problems. In *Proc. of the 20th Int. Coll. on Automata, Languages and Programming (ICALP '93)*, pp. 40–51. 1993.

[11] Lokshtanov Daniel and Marx Dániel. Clustering with Local Restrictions. *Information and Computation*, pp. 278-292. 2013.