On the complexity of finding a potential community

Cristina Bazgan[1*], Thomas Pontoizeau[1], and Zsolt Tuza[2,3]

[1] Université Paris-Dauphine, PSL Research University, CNRS, LAMSADE, 75016 Paris, France
bazgan, thomas.pontoizeau@lamsade.dauphine.fr

[2] Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Budapest, Hungary

[3] Department of Computer Science and Systems Technology, University of Pannonia, Veszprém, Hungary
tuza@dcs.uni-pannon.hu

Abstract. An independent 2-clique of a graph is a subset of vertices that is an independent set and such that any two vertices inside have a common neighbor outside. In this paper, we study the complexity of finding an independent 2-clique of maximum size in several graph classes and we compare its complexity with the complexity of maximum independent set. We prove that this problem is NP-hard on apex graphs, APX-hard on line graphs, not $n^{1/2-\epsilon}$-approximable on bipartite graphs and not $n^{1-\epsilon}$-approximable on split graphs, while it is polynomial-time solvable on graphs of bounded degree and their complements, graphs of bounded treewidth, planar graphs, (C_3, C_6)-free graphs, threshold graphs, interval graphs and cographs.

Keywords: combinatorial optimization, complexity, algorithm, independent set, inapproximability

1 Introduction

Community detection is a well established research field in the area of social networks. It can find many applications in this area with the recent development of social networks like Facebook or Linkedin. A social network can be easily modeled by a graph in which vertices represent members and edges represent relationships between those members.

There are several ways to define a community. Intuitively, a community corresponds to a dense subgraph, that is to say a subgraph with a lot of edges. If a community is defined as a group of maximum size such that all members know each other, it corresponds to the well known NP-hard problem of finding a maximum clique. However, such a condition is strong and is not always relevant to describe a community.

*Institut Universitaire de France

Another way to define a community is to relax the strong condition of a clique and focus on the distance between members of a social network. Different measures have been studied to describe it. Luce introduced in [12] the notion of k-cliques while Mokken extended this notion in [13] by defining k-clubs. A k-clique (resp. a k-club) of G is a subgraph S in which any two vertices are at distance at most k in G (resp. in the subgraph induced by S). The standard term 'clique' means both a 1-clique and a 1-club.

With the recent development of social networks and particularly online dating services, it could be interesting to investigate the detection of some group of people who do not know each other, but are related by their other relationships. Such a group could be considered as a 'potential' community since it does not form a community in the first place, but could become one thanks to their proximity. This may find various applications in online dating and meet-up services in which members expect not to know the other members.

More precisely, considering a graph G, we want to define potential communities by looking at independent sets in which any two members are related within a specified distance in G. Contrary to a k-club, the distance between two vertices must be realized via vertices outside of the subgraph. We call such a subset of vertices an independent k-clique, where k is the largest distance between vertices of S in the original graph. In this paper, we study the problem of finding an independent 2-clique of maximum size.

We investigate the complexity of finding an independent 2-clique of maximum size in several graph classes. Since this problem is close to finding an independent set of maximum size, we also compare the hardness of the two problems. Figure 1 summarizes the results we prove in the paper.

The paper is structured as follows. In Section 2 we introduce formally some notation and definitions. In Section 3 we show that the complexity of Max Independent 2-Clique jumps from polynomial-time solvable to NP-hard when the input class is extended from planar graphs to apex graphs. In Section 4 we present polynomial algorithms to solve Max Independent 2-Clique in some graph classes. In Section 5 we show NP-hardness and non-approximability of Max Independent 2-Clique in some other graph classes. Conclusions are given in Section 6. Due to space limit, some proofs will be given in a journal version.

2 Preliminaries

In this paper, all considered graphs are undirected. The *complement* $\overline{G} = (V, \overline{E})$ of a graph $G = (V, E)$ is the graph in which $uv \in E$ if and only if $uv \notin \overline{E}$, for all vertex pairs $u, v \in V$. A *k-cycle* is a cycle of length k. The maximum degree of a vertex in a graph G will be denoted by the usual notation $\Delta(G)$.

We recall that a *clique* in a graph is a set of mutually adjacent vertices. A set of vertices is called a 2-*clique* if any two vertices of the set are at distance at most 2 in G. An *independent set* in a graph is a set of vertices such that no

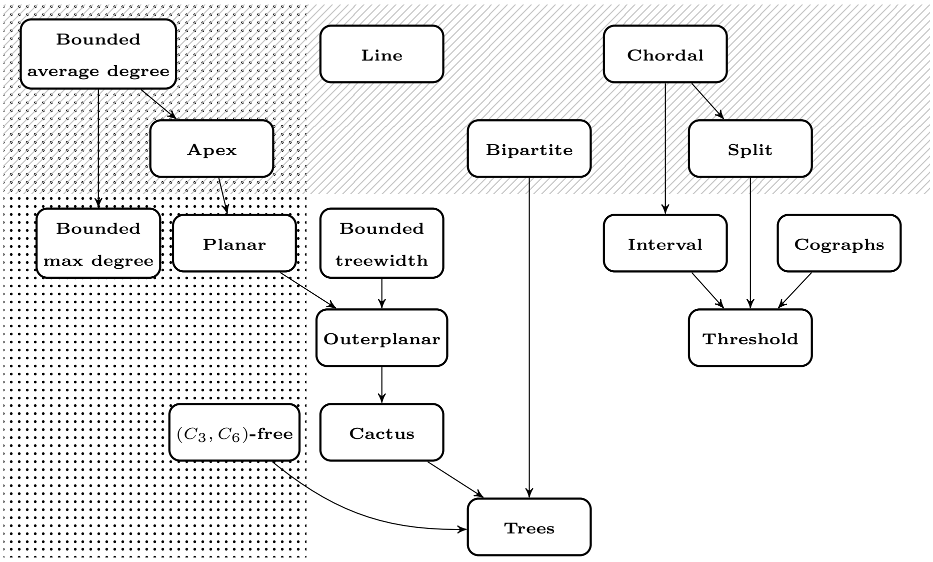

Fig. 1. Relationship among some classes of (connected) graphs, where each child is a subset of its parent. We compare the hardness of Max Independent 2-Clique and Max Independent Set in studied graph classes. Max Independent 2-Clique is NP-hard on graph classes at the top (hatched area) and is polynomial-time solvable on graph classes at the bottom (non-hatched area). Max Independent Set is NP-hard on graph classes on the left (dotted area) and is polynomial-time solvable on graph classes on the right (non-dotted area).

two of them are joint by an edge. An *independent 2-clique* is a subset of vertices which is an independent set and a 2-clique at the same time.

In this paper we are interested in the following optimization problem:
Max Independent 2-Clique
Input: A graph $G = (V, E)$.
Output: A subset $S \subset V$ which is an independent 2-clique of maximum size.

The Max Independent 2-Clique problem is closely related to another well known one:
Max Independent Set
Input : A graph $G = (V, E)$.
Output : A subset $S \subset V$ such that S is an independent set of maximum size.

Given a graph G, the standard notation for the maximum size of an independent set in G is $\alpha(G)$. The maximum number of vertices in an independent 2-clique of G will be denoted by $\alpha_{=2}(G)$. The subscript '=2' intends to express that the distance between any two vertices of the independent set is exactly 2.

Note that $\alpha_{=2}(G) \geq 2$ whenever at least one connected component of G is not a complete graph. (Indeed, any such component contains two vertices at distance exactly two, hence forming an independent 2-clique of size 2.) Moreover, if G is disconnected and has components $G_1, \ldots, G_k$ then

$$\alpha_{=2}(G) = \max_{1 \leq i \leq k} \alpha_{=2}(G_i)$$

For these reasons we assume throughout that G is a non-complete, connected graph (although some of the algorithms also need to handle disconnected graphs temporarily).

Some classes of graphs. A *cactus* is a graph in which each edge occurs in at most one cycle. A (C_3, C_6)*-free graph* is a graph containing no triangle C_3 and no induced cycle of length 6. An *interval graph* is a graph for which there exists a family of intervals on the real line and a bijection between the vertices of the graph and the intervals of the family in such a way that two vertices are joined by an edge if and only if the intersection of the two corresponding intervals is non-empty. A *threshold graph* is a graph which can be constructed from the empty graph by a sequence of two operations: insertion of an isolated vertex, and insertion of a dominating vertex (i.e., a vertex adjacent to all the other vertices). A *cograph* is a graph that can be generated from the single-vertex graph by (repeated applications of) complementation and vertex-disjoint union. A *split graph* is a graph whose vertex set can be partitioned into two subsets, one inducing an independent set S and the other one inducing a clique K. We denote by $K_{p,m}$ the complete bipartite graph with p and m vertices in its vertex parts. The *line graph* of a graph G is the graph $L(G)$ whose vertices represent the edges of G, and two vertices of $L(G)$ are adjacent if and only if the corresponding two edges of G share a vertex. A graph is *outerplanar* if it has a crossing-free embedding in the plane such that all vertices are on the same face. A graph is k*-outerplanar* if for $k = 1$, G is outerplanar and for $k > 1$ the graph has a planar embedding such that if all vertices on the exterior face are deleted, the connected components of the remaining graph are all $(k-1)$-outerplanar. A graph G is *apex* if it contains a vertex v such that $G \setminus v$ is planar. A family of graphs on n vertices is δ*-dense* if it has at least $\frac{\delta n^2}{2}$ edges. It is *everywhere-*δ*-dense* if the minimum degree is at least δn. A family of graphs is *dense* (resp. *everywhere-dense*) if there is a constant $\delta > 0$ such that all members of this family are δ-dense (resp. everywhere-δ-dense).

3 Complexity jump from planar graphs to apex graphs

According to [8], Max Independent Set is known to be NP-hard in planar graphs, and thus also in apex graphs. On the other hand, we prove that Max Independent 2-Clique is polynomial-time solvable on planar graphs but NP-hard on apex graphs. This shows that inserting or removing a single vertex in a graph may dramatically change the complexity of Max Independent 2-Clique.

Theorem 1. Max Independent 2-Clique *is NP-hard on apex graphs.*

Proof. We establish a polynomial reduction from Max Independent Set on cubic planar graphs, which is proved NP-hard in [8], to Max Independent

2-Clique on apex graphs. Let $G = (V, E)$ be a cubic planar graph, an instance of Max Independent Set. The instance $G' = (V', E')$ of Max Independent 2-Clique is defined by inserting an additional vertex z that is adjacent to every vertex of V. It is easy to see that $\{z\}$ itself is a one-element non-extendable independent 2-clique, while the independent 2-cliques of G' not containing z are precisely the independent sets of G. □

This first theorem implies another interesting result:

Corollary 2. Max Independent 2-Clique *is NP-hard on the class of graphs of average degree at most 5.*

Proof. Cubic graphs on n vertices have $3n/2$ edges, thus the graph constructed in the proof of Theorem 1 is of order $n + 1$ and has $5n/2$ edges, yielding average degree less than 5. □

Now, in order to prove that Max Independent 2-Clique is polynomial-time solvable on planar graphs, we use a famous theorem introduced by Courcelle in [6] which states that any problem expressible in Monadic Second-Order Logic is linear-time solvable for graphs of bounded treewidth. This allows to show first the following:

Theorem 3. Max Independent 2-Clique *is linear-time solvable on graphs with bounded treewidth.*

Proof. We observe that the property 'Independent 2-Clique' is expressible in Monadic Second-Order Logic:

$$\mathit{I2C}(S) := \forall x \forall y (Sx \wedge Sy) \rightarrow (\neg edg(x, y) \wedge (\exists z, edg(x, z) \wedge edg(y, z)))$$

Since any problem expressible in Monadic Second-Order Logic is linear-time solvable for graphs of bounded treewidth (see [6]), $\alpha_{=2}$ can be determined in linear time in graphs of bounded treewidth. □

Based on this result, we prove the following result.

Theorem 4. Max Independent 2-Clique *is polynomial-time solvable on planar graphs.*

Proof. Let $G = (V, E)$ be a planar graph and $v \in V$ any vertex. Then all the other vertices in an independent 2-clique S containing v are at distance exactly 2 apart from v. Further, the 2-clique property for $S \setminus \{v\}$ is ensured by vertices within distance at most 3 from v. Thus, the vertices relevant for S to be an independent 2-clique induce a subgraph G' in G such that G' belongs to the class of '4-outerplanar' graphs. Graphs which are 4-outerplanar have treewidth at most 11 (more generally, all k-outerplanar graphs have treewidth at most $3k - 1$, due to [3]). Then, using Theorem 3, a polynomial-time algorithm for Max Independent 2-Clique in planar graphs consists in solving the problem for all subgraphs G' (which have treewidth at most 11) defined from each vertex v. □

4 Graph classes with polynomial-time algorithms

In the following we identify some graph classes on which Max Independent 2-Clique is computable in polynomial time, while Max Independent Set is not always polynomial-time solvable.

First, it is interesting to notice that, according to the next propositions, Max Independent 2-Clique is polynomial-time solvable on graphs of bounded degree and also on complements of graphs of bounded degree, while Max Independent Set is NP-hard on graphs of bounded degree [8] but polynomial-time solvable on their complements (using exhaustive search in the non-neighborhood of each vertex, which can be done in linear time).

Proposition 5. Max Independent 2-Clique *is linear-time solvable on graphs with bounded maximum degree, and also on graphs of minimum degree at least* $(n-d)$*, where* d *is constant.*

Now, notice that a natural way to find an independent 2-clique is to take an independent set included in the neighborhood of one vertex. Then, it is interesting to investigate the properties of a graph in which an independent 2-clique is not included in the neighborhood of one vertex. We show in Lemma 6 that such a graph necessarily contains a cycle of length 3 or 6, and cannot be a cactus if such an independent 2-clique has a certain size. Such properties allow to get an easy polynomial-time algorithm for Max Independent 2-Clique on (C_3, C_6)-free graphs, while Max Independent Set is NP-hard on this class of graphs (see [1]). From Theorem 4 we already know that Max Independent 2-Clique is linear-time solvable on cactus graphs, but the property of Lemma 6 allows to give a simpler algorithm for this class of graph.

Lemma 6. *Let* $G = (V, E)$ *be a graph. Suppose that there exists an independent 2-clique* S *not contained in the neighborhood of a single vertex. Then* G *contains an induced cycle of length* 3 *or* 6*. Moreover, if* $|S| \geq 4$*,* G *is not a cactus.*

This lemma implies the following theorem:

Theorem 7. *Any* (C_3, C_6)*-free graph* G *satisfies* $\alpha_{=2}(G) = \Delta(G)$ *and* Max Independent 2-Clique *is linear-time solvable on it.*

Recalling that a tree does not contain any cycle, for the classical class of trees we obtain the following:

Corollary 8. *Any tree* T *satisfies* $\alpha_{=2}(T) = \Delta(T)$ *and* Max Independent 2-Clique *is linear-time solvable on it.*

Finally, Lemma 6 allows to give a polynomial-time algorithm for Max Independent 2-Clique on cactus.

Proposition 9. Max Independent 2-Clique *is linear-time solvable on cactus graphs.*

We focus in the following part of this section on classes of graphs on which both Max Independent 2-Clique and Max Independent Set are polynomial-time solvable. We first investigate a subclass of split graphs, namely threshold graphs. It follows from the definitions that a threshold graph $G = (V, E)$ is a split graph with the following property: the vertices of the independent set S can be ordered as $v_1, \ldots, v_p$ such that $N_G(v_1) \subseteq N_G(v_2) \subseteq \ldots \subseteq N_G(v_p)$. We denote by $u_1, \ldots, u_q$ the vertices of the clique K, and we suppose that $d_G(u_1) \leq d_G(u_2) \leq \ldots \leq d_G(u_q)$. Without loss of generality, we assume that there is no isolated vertex in G. Note that a threshold graph can be recognized in linear time (see [10]).

Proposition 10. Max Independent 2-Clique *is linear-time solvable on threshold graphs. Moreover, in every threshold graph G without isolated vertices we have $\alpha_{=2}(G) = \alpha(G)$.*

Proof. Let $G = (V, E)$ be a threshold graph with the previous decomposition into S and K. Let $N_G(v_p) = \{u_r, u_{r+1}, \ldots, u_q\}$, for some $r \geq 1$. Then a maximum independent 2-clique in G is S if $K \setminus N_G(v_p) = \emptyset$, and otherwise it is $S \cup \{z\}$ with any $z \in K \setminus N_G(v_p)$, since in both cases the common neighbor of all these vertices is u_q. Since Max Independent Set can be solved in linear time in threshold graphs [7], Max Independent 2-Clique can be solved in linear time. □

The previous result can be extended in two directions, for interval graphs and for cographs.

Using the results of Booth and Lueker [4] it can be tested in linear time whether a graph G is an interval graph; and if it is, then an interval representation $I_1, \ldots, I_n$ of G can also be generated.

Proposition 11. Max Independent 2-Clique *is polynomial-time solvable on interval graphs.*

Proof. Consider any $G = (V, E)$ and let $I_1, \ldots, I_n$ be an interval representation of G. In order to determine $\alpha_{=2}(G)$, first notice that all vertices of an independent 2-clique S of G must have a common neighbor. Indeed, if I and I' are the leftmost and the rightmost intervals of S then any of their common neighbors intersects all intervals located between them, and therefore is a common neighbor of all members of S. Then, for every vertex I, we compute a maximum independent set in the subgraph induced by the neighborhood of I. An optimal solution is such an independent set with maximum size. Since Max Independent Set is polynomial-time solvable on interval graphs [9], the result follows. □

We consider now the class of cographs, that contains all threshold graphs.

Proposition 12. Max Independent 2-Clique *is polynomial-time solvable on cographs.*

Notice that since Max Independent Set is linear-time solvable on chordal graphs [7], it is also polynomial-time solvable on interval graphs and threshold graphs. Moreover, Max Independent Set is also polynomial-time solvable on cographs by bottom-up tree computation [5].

5 NP-hardness and non-approximability

Using the reduction from the proof of Theorem 1, we can conclude:

- Max Independent 2-Clique is NP-hard on dense (resp. everywhere dense) graphs, since Max Independent Set is NP-hard on dense (resp. everywhere dense) graphs. Moreover, Max Independent 2-Clique is not $n^{1-\varepsilon}$-approximable for any $\varepsilon > 0$, if P $\neq$ NP, on everywhere dense graphs (and respectively dense graphs) since the same result holds for Max Independent Set on everywhere dense graphs (and respectively dense graphs). In order to get this last result, we use the same inaproximability result for Max Independent Set on general graphs [15] and a reduction preserving approximation from general graphs to everywhere dense graphs (that consists of adding a clique of the same size as the size of the graph and joining every vertex from the original graph to all vertices in this clique).
- Max Independent 2-Clique is NP-hard on K_4-free graphs, since Max Independent Set is NP-hard on K_3-free graphs [1].

We now investigate graph classes in which Max Independent 2-Clique is NP-hard while Max Independent Set is polynomial-time solvable. We first consider a graph class containing threshold graphs, namely the class of split graphs, for which Max Independent 2-Clique becomes NP-hard (and even not $n^{1-\varepsilon}$-approximable). Since Max Independent Set is polynomial-time solvable on chordal graphs [7], it is also polynomial-time solvable on split graphs.

Theorem 13. *On split graphs,* Max Independent 2-Clique *is NP-hard and it is not $n^{1-\varepsilon}$-approximable in polynomial time unless $P = NP$.*

We prove now that Max Independent 2-Clique is NP-hard (and even not $n^{1/2-\epsilon}$-approximable) on bipartite graphs while Max Independent Set is polynomial-time solvable since the number of vertices in a maximum independent set equals the number of edges in a minimum edge covering.

Theorem 14. *On bipartite graphs,* Max Independent 2-Clique *is NP-hard and is not $n^{1/2-\varepsilon}$-approximable in polynomial time, unless $P = NP$.*

Proof. First we prove the NP-hardness. Max Independent Set is known to be NP-hard on 3-regular graphs [8], so Max Clique is also NP-hard on $(n-4)$-regular graphs (where n is the number of vertices), by considering its complement. We reduce Max Clique on $(n-4)$-regular graphs to Max Independent 2-Clique on bipartite graphs. Let $G = (V, E)$ be an $(n-4)$-regular graph. We construct an instance of $G' = (V', E')$ of Max Independent 2-Clique on bipartite graphs as follows (see Figure 2).

Let V_1, V_2, V_3, V_4 be four copies of V. Let E_1 be a set of $|E|$ vertices corresponding to the edges in E, and define $V' := V_1 \cup V_2 \cup V_3 \cup V_4 \cup E_1$. Let there exist an edge in E' between a vertex v in V_i, $i \in \{1, 2, 3, 4\}$ and a vertex e in E_1 if and only if the corresponding vertex v in V is incident with the corresponding edge e in E.

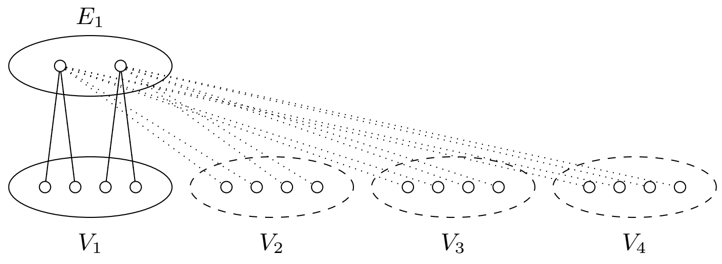

Fig. 2. The bipartite graph G', an instance of Max Independent 2-Clique

Now we show that G contains a clique of size at least k if and only if G' contains an independent 2-clique of size at least $4k$.

Given a clique $C \subseteq V$ of size at least k in G, the union of the four copies of C in G' is an independent 2-clique of size at least $4k$.

For the other direction, notice first that the value of a maximum independent set in a 3-regular graph is at least $\lceil \frac{n}{4} \rceil$. Then, the value of a maximum clique in an $(n-4)$-regular graph is also at least $\lceil \frac{n}{4} \rceil$. Thus the size of a maximum independent 2-clique in G' is at least n.

We consider now a solution C' of Max Independent 2-Clique in G' with at least $4k \geq n$ vertices (this restriction is always possible because of the previous comment). Notice that C' cannot contain both a vertex from E_1 and a vertex from $V' \setminus E_1$ since the distance between any two vertices of C' must be 2. A solution which is a subset of E_1 would mean pairwise intersecting edges in G, hence would have size at most $\max(3, n-4) < n$. Therefore C' must be a subset of $V' \setminus E_1$. Notice that for any $i \in \{1,2,3,4\}$, $C' \cap V_i$ must be a copy of a clique in G. Then C' is a union of copies of four cliques in G, and $|C'| \geq 4k$. Let C_0 be the copy of largest size, which thus has $|C_0| \geq k$. Then C_0 is the copy of a clique C of G of size at least k.

For the proof of non-approximability, we construct an E-reduction (see [11]) from Max Clique. Let $I = (V, E)$ be an instance of Max Clique. Consider a reduction similar to the one for the proof of NP-hardness, except that we now consider $\ell = |V|$ copies $V_1, \ldots, V_\ell$ instead of four copies of V; adjacencies are defined in the same way as before. We denote by $I' = (V', E')$ the corresponding instance of Max Independent 2-Clique from the reduction. As in previous proof, starting with a clique of size $opt(I)$, we can construct an independent 2-clique of size $\ell \cdot opt(I)$ in G' and thus $opt(I') \geq \ell \cdot opt(I)$. Let S' be any independent 2-clique in I' of size at least ℓ (it always exists, take e.g. the ℓ copies of the same vertex, one copy in each V_i). As before, S' cannot contain both a vertex of E_1 and a vertex from $V \setminus E_1$ since two vertices of S' must have distance 2 in G', and S' cannot contain only vertices from E_1 since any independent 2-clique included in E_1 is of size at most $\max(3, \Delta(G)) < \ell - 1$. Moreover, each subset $V_i \cap S'$ corresponds to a clique in G. Let S be the subset $V_i \cap S'$ of largest size. We have $|S| \geq \frac{|S'|}{\ell}$ and then $opt(I) \geq |S| \geq \frac{|S'|}{\ell} = \frac{opt(I')}{\ell}$ when S' is an optimal solution. Using that $opt(I') \geq \ell \cdot opt(I)$ we get $opt(I') = \ell \cdot opt(I)$ and we obtain:

$$\epsilon(I,S) = \frac{opt(I)}{|S|} - 1 \leq \frac{\ell \cdot opt(I')}{\ell \cdot |S'|} - 1 = \epsilon(I',S')$$

Since we clearly have $opt(I') \leq p(|I|) \cdot opt(I)$ with a polynomial p, the reduction is an E-reduction. Then, since Max Clique is not $\ell^{1-\varepsilon}$-approximable unless $P = NP$ [15], the same property holds for Max Independent 2-Clique. Thus Max Independent 2-Clique is not $n^{1/2-\varepsilon}$ approximable where $n = |V'|$ since $n = \ell^2 + |E|$. □

Finally we prove that Max Independent 2-Clique is NP-hard (and even APX-hard) on line graphs, while Max Independent Set is polynomial-time solvable since it consists in a maximum matching in the original graph.

Theorem 15. *On line graphs,* Max Independent 2-Clique *is NP-hard and even APX-hard.*

Proof. First we prove the NP-hardness. We establish a reduction from the Max Clique problem on graphs of minimum degree at least $n-4$. Consider an instance $G = (V,E)$ of Max Clique with $|V| = n$. We construct a graph $G' = (V',E')$ (see Figure 3) as follows. Let $G_0 = (V_0,E_0)$ be a copy of G. Let V' be $V_0 \cup A \cup B \cup C$ where A, B, C are three sets of n vertices. Then, let $E' = E_0 \cup E_1 \cup E_2 \cup E_3 \cup E_4$ such that E_1 is a perfect matching between V_0 and A, E_2 is the set of all possible edges (i.e., a complete bipartite graph) between the vertices of A and the vertices of B, E_3 is a perfect matching between B and C, and E_4 is the set of all possible edges between any two vertices of C (a complete subgraph). The line graph of G', denoted by $L(G')$, is an instance of Max Independent 2-Clique. Notice that an independent 2-clique in $L(G')$ corresponds to a set of edges in G' such that, for each pair of edges $\{e_1,e_2\}$ in the set, e_1 and e_2 are not adjacent but are joined by an edge. We show that G contains a clique of size at least k if and only if $L(G')$ contains an independent 2-clique of size at least $k+n$.

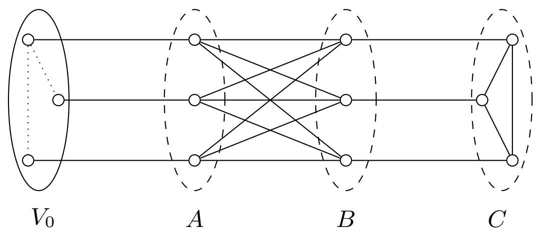

Fig. 3. The graph G' for which the corresponding line graph $L(G')$ is an instance of Max Independent 2-Clique

Consider a clique S of size k in G, and let S_0 be its copy in G'. We define a set of edges S' of size at least $k+n$ in G' as follows. For any vertex $v \in S_0$, add in S' its adjacent edge in E_1. Moreover add the entire E_3 to S'. We show now that any pair of edges in S' have an adjacent edge in common. Two edges of $S' \cap E_1$ have a common adjacent edge in E_0 since the subgraph induced by S_0 is a clique. Similarly, two edges of E_3 have a common adjacent edge in E_4. Moreover, an edge of $S' \cap E_1$ and an edge of E_3 have a common adjacent edge in E_2 since the subgraph induced by $A \cup B$ is $K_{n,n}$. Then, the corresponding set of vertices in $L(G')$ is an independent 2-clique of size $k+n$.

In the other direction, consider an independent 2-clique in $L(G')$ of size $k+n$. Notice that it is always possible to take the set of vertices in $L(G')$ corresponding to E_3 in G' and two edges in E_1 whose vertices in V_0 are neighbors in G', hence we can suppose that $k \geq 2$. Let S' be the set of all corresponding edges in G'. Suppose first that there is exactly one edge from E_0 in S'. Then, there are at most $n-2$ edges from E_1 in S', and there are at most 2 edges from E_2 in S', due to the constraints of an independent 2-clique. There cannot be edges from $E_3 \cup E_4$ in S' since they would not be joined to the edge of $E_0 \cap S'$ by any edge. Then, S' contains at most $n+1$ edges in S', which contradicts $k \geq 2$. Suppose now that there are at least two edges from E_0 in S'. Name two of them $e_{0,1}$ and $e_{0,2}$. Then, there are at most $n-4$ edges from E_1 in S' but there is no edge from E_2 in S'. Indeed, an edge e_2 from E_2 in S' can be joined by an edge to at most one of $e_{0,1}$ and $e_{0,2}$. Then the size of S' does not exceed n, which contradicts $k \geq 2$. Thus, we can assume that there is no edge from E_0 in S'. Similarly, there is no edge from E_4 in S'. Now, notice that $|S' \cap (E_2 \cup E_3)| \leq n$ since if $S' \cap (E_2 \cup E_3)$ contained $n+1$ edges then at least two of these edges would have a common endpoint. Consequently, $|S' \cap E_1| \geq k$. Moreover, any two edges from $S' \cap E_1$ must have a common adjacent edge in E_0 since they cannot have a common adjacent edge in E_2. Then, the subgraph of G induced by the set of vertices in V_0 which are the endpoints of the edges in $S' \cap E_1$ must be a clique whose size is at least k.

For the proof of APX-hardness, we prove that the reduction above is an L-reduction (see [14]). We proved in the NP-hardness part that any independent 2-clique in I' has a size at most $2n$. Then $opt(I') \leq 2n = 8 \cdot \frac{n}{4} \leq 8 \cdot opt(I)$ follows since $opt(I) \geq \frac{n}{4}$ in graphs of degree at least $n-4$. Moreover, starting with a clique of size $opt(I)$, we can construct an independent 2-clique of size $opt(I)+n$ and therefore $opt(I') \geq n + opt(I)$. Let S' be an independent 2-clique in I' of size at least $n+2$ (we proved in the NP-hardness part that it always exists and that such a set must be included in $E_1 \cup E_2 \cup E_3$). Let S be the set of vertices in V_0 which are incident with edges in $E_1 \cap S'$. We have $|S'| - |S| \leq n$ which implies $n + |S| \geq |S'|$. Then we obtain $opt(I) - |S| \leq opt(I') - n - |S| = opt(I') - (n+|S|) \leq opt(I') - |S'|$. Since Max Independent Set is APX-hard on the class of graphs of maximum degree 3 [2], Max Clique is also APX-hard on the class of graphs of minimum degree at least $n-4$. Thus, Max Independent 2-Clique is APX-hard on line graphs. □

6 Conclusion

Even if Max Independent 2-Clique and Max Independent Set are similar problems, their complexity can be very different depending on the graph class we try to solve the problem in. We showed that Max Independent 2-Clique is NP-hard on apex, dense and everywhere dense, K_4-free, split, bipartite and line graphs while it is polynomial-time solvable on bounded treewidth, planar, bounded degree (and complement of bounded degree), (C_3, C_6)-free, interval graphs and on cographs. Many further types of graphs may be of interest, concerning separation of graph classes in which the problem is NP-hard from the ones where the problem is solvable in polynomial-time.

References

1. V. E. Alekseev. On the local restrictions effect on the complexity of finding the graph independence number. *Combinatorial-Algebraic Methods in Applied Mathematics*, pages 3–13, 1983.
2. P. Alimonti and V. Kann. Hardness of approximating problems on cubic graphs. In *Proceedings of the 3rd Italian Conference on Algorithms and Complexity (CIAC 1997)*, pages 288–298, 1997.
3. H. L. Bodlaender. A partial k-arboretum of graphs with bounded treewidth. *Theoretical Computer Science*, 209:1–45, 1998.
4. K. S. Booth and G. S. Lueker. Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms. *Journal of Computer and System Sciences*, 13(3):335–379, 1976.
5. L. S. Burlingham, D.G. Corneil, and H. Lerchs. Complement reducible graphs. *Discrete Applied Mathematics*, pages 163 – 174, 1981.
6. B. Courcelle. The monadic second-order logic of graphs iii : tree-decompositions, minors and complexity issues. *RAIRO - Informatique Théorique et Applications*, 26:257–286, 1992.
7. A. Frank. Some polynomial algorithms for certain graphs and hypergraphs. *Congressus Numerantium No. XV*, pages 3–13, 1976.
8. M. R. Garey, D. S. Johnson, and L. Stockmeyer. Some simplified NP-complete problems. *Theoretical Computer Science*, 1:237–267, 1976.
9. U. I. Gupta, D. T. Lee, and J. Y.-T. Leung. Efficient algorithms for interval graphs and circular-arc graphs. *Networks*, 12(4):459–467, 1982.
10. P. Heggernes and D. Kratsch. Linear-time certifying recognition algorithms and forbidden induced subgraphs. *Nordic Journal of Computing*, 14:87–108, 2007.
11. S. Khanna, R. Motwani, M. Sudan, and U. V. Vazirani. On syntactic versus computational views of approximability. *SIAM Journal on Computing*, 28(1):164–191, 1998.
12. R. D. Luce. Connectivity and generalized cliques in sociometric group structure. *Psychometrika*, 15:169–190, 1950.
13. R. J. Mokken. Cliques, clubs and clans. *Quality and Quantity*, 13(2):161–173, 1979.
14. C. H. Papadimitriou and M. Yannakakis. Optimization, approximation, and complexity classes. *Journal of Computer and System Sciences*, 43(3):425–440, 1991.
15. D. Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, 3(1):103–128, 2007.