Discrete Applied Mathematics 270 (2019) 25–36

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Proportionally dense subgraph of maximum size: Complexity and approximation

Cristina Bazgan [a,1], Janka Chlebíková [b], Clément Dallard [b,*], Thomas Pontoizeau [a]

[a] Université Paris-Dauphine, Université PSL, CNRS, LAMSADE, 75016 Paris, France
[b] School of Computing, University of Portsmouth, Portsmouth, United Kingdom

ARTICLE INFO

Article history:
Received 5 November 2018
Received in revised form 3 July 2019
Accepted 8 July 2019
Available online 22 July 2019

Keywords:
Dense subgraph
Approximation
Complexity
Hamiltonian cubic graphs

ABSTRACT

We define a *proportionally dense subgraph* (PDS) as an induced subgraph of a graph with the property that each vertex in the PDS is adjacent to proportionally as many vertices in the subgraph as in the graph. We prove that the problem of finding a PDS of maximum size is APX-hard on split graphs, and NP-hard on bipartite graphs. We also show that deciding if a PDS is inclusion-wise maximal is co-NP-complete on bipartite graphs. Nevertheless, we present a simple polynomial-time $(2-\frac{2}{\Delta+1})$-approximation algorithm for the problem, where Δ is the maximum degree of the graph. Finally, we show that all Hamiltonian cubic graphs with n vertices (except two) have a PDS of size $\lfloor \frac{2n+1}{3} \rfloor$, which we prove to be an upper bound on the size of a PDS in cubic graphs.

1. Introduction

For a graph $G=(V,E)$, the density of a subgraph on a vertex set $S \subseteq V$ is commonly defined as $\frac{|E(S)|}{|S|}$, where $E(S)$ is the set of edges in the subgraph. The problem of finding a subgraph of maximum density can be solved in polynomial time using a max flow technique [8]. However, when the subgraph must contain *exactly* k vertices, the problem becomes NP-hard [3,7] and is known as the Densest k-subgraph problem. Two variants of the problem have also been studied where the number of vertices in the subgraph must be either *at least* k or *at most* k. The former is known to be NP-hard [10], but there exists a polynomial-time 2-approximation algorithm to solve it [2]. It was showed that any α-approximation for the *at most* k variant would imply a $\Theta(\alpha^2)$-approximation for the densest k-subgraph problem [1].

An induced subgraph on a vertex set $S \subset V$ is said to be *proportionally dense* if all of its vertices in S have *proportionally* as many neighbors in the subgraph as in the graph, and hence the condition $\frac{d_S(u)}{|S|-1} \geq \frac{d(u)}{|V|-1}$ holds for each vertex u in S. In this paper, we study the problem of finding a *proportionally dense subgraph* (PDS) with a maximum number of vertices. A *proportionally dense subgraph* grants more importance to the vertices than the standard definition of a dense subgraph, as all the vertices in a PDS must be 'satisfied', *i.e.* respect the above condition. This can be compared with *defensive alliances* in graphs, where the vertices in the alliance must have at least as many neighbors inside the alliance than outside it [11,14], without the notion of proportion of neighbors.

From a theoretical point of view, it is interesting to observe a problem that connects local and global properties of vertex subsets, interweaving the size of the subset and the number of neighbors. This interesting paradigm has rarely been seen in graph theory problems.

* Corresponding author.
E-mail addresses: bazgan@lamsade.dauphine.fr (C. Bazgan), janka.chlebikova@port.ac.uk (J. Chlebíková), clement.dallard@port.ac.uk (C. Dallard), thomas.pontoizeau@lamsade.dauphine.fr (T. Pontoizeau).

1 Institut Universitaire de France.

https://doi.org/10.1016/j.dam.2019.07.010
0166-218X/

The notion of proportionality of neighbors is closely related to community detection problems. Olsen [12] defined a *community structure* as a partition of the vertices of a graph into parts such that each vertex has a greater proportion of neighbors in its part than in any other part, each part being called a *community*. In the same paper, it was proved that any graph that is not a star contains a community structure that can be found in polynomial time (if there is no restriction on the number of communities), but that it is NP-complete to decide if a given subset of vertices can belong to a same community of a community structure. The special case where the community structure contains exactly two communities, namely a 2-*community structure*, has been studied in several classes of graphs: a 2-community structure always exists and can be found in polynomial time in trees, graphs with maximum degree 3, minimum degree $|V| - 3$, and complements of bipartite graphs [5]. Recently, the notion of 2-community structure has been studied under the name of 2-PDS partition [4]. In this paper, the authors described an infinite family of graphs without a 2-PDS partition, and a second infinite family of graphs without a connected 2-PDS partition (but with a disconnected one). These results answer some open questions originally introduced in [5]. However, the complexity of finding a 2-PDS partition remains unknown in general graphs, and for larger (fixed) number of PDSs. As there is equivalence between *proportionally dense subgraph* and *community* (with regard to the above definition), one may interpret the problem of finding a proportionally dense subgraph of maximum size as finding a community of maximum size. Hence, all the results presented in this paper can also be applied for community related problems.

Section 2 introduces the basic notations used in the paper. Section 3 presents various hardness results of the Max Proportionally Dense Subgraph problem. Section 4 gives positive results about the approximation of this problem. We prove that the problem can be solved in linear time on Hamiltonian cubic graphs in Section 5. Conclusion and open problems are given in Section 6.

2. Preliminaries

Throughout the paper, we assume that all graphs are simple, undirected and connected. For a graph $G = (V, E)$, we denote by $N(v)$ the set of neighbors of $v \in V$ and by $d(v)$ the degree of v, and thus $d(v) = |N(v)|$. Also, $\Delta(G)$ denotes the maximum degree of G (or simply Δ when no confusion arises).

In addition, given a subset of vertices $S \subset V$, we define $d_S(v) = |N(v) \cap S|$ and $\bar{S} := V \setminus S$; also, $G[S]$ represents the induced subgraph of S in G.

A *star* is a complete bipartite graph $K_{1,\ell}$ for any $\ell \geq 1$. A *split graph* is a graph in which the vertices can be partitioned into an independent set and a clique.

The Maximum Proportionally Dense Subgraph problem

Definition 1. Let $G = (V, E)$ be a graph and $S \subset V$, such that $2 \leq |S| < |V|$. We say that the induced subgraph $G[S]$ is a *proportionally dense subgraph* (PDS) if for each vertex $u \in S$,

$$\frac{d_S(u)}{|S|-1} \geq \frac{d(u)}{|V|-1}\text{, which is equivalent to } \frac{d_S(u)}{|S|-1} \geq \frac{d_{\bar{S}}(u)}{|\bar{S}|}. \tag{1}$$

We say a vertex u is *satisfied* (in $G[S]$) if it respects Eq. (1). The size of the proportionally dense subgraph $G[S]$ corresponds to the cardinality of S.

The proof of the above equivalence from Eq. (1) can be found in [5].

Max Proportionally Dense Subgraph (Max PDS)
Input: A graph G.
Output: A proportionally dense subgraph in G of maximum size.

A proportionally dense subgraph may be connected or not. We study both cases and talk about *connected PDS* in the former case. Notice that there exist graphs for which all proportionally dense subgraphs of maximum size are not connected, even if the graph is a cubic graph or a caterpillar. In the cubic graph illustrated in Fig. 1, the gray vertices represent a PDS of size 7, which is not connected. In fact, any connected induced subgraph on the set S with at least 6 vertices contains at least one vertex u of degree 1 in S, which is not satisfied since $\frac{d_S(u)}{|S|-1} \leq \frac{1}{6-1} < \frac{2}{4} \leq \frac{d_{\bar{S}}(u)}{|\bar{S}|}$. It can be checked that the maximum size for a PDS is 7 but only 5 for a connected PDS. Similarly, in the caterpillar in Fig. 1, any connected induced subgraph of size at least 12 has one vertex unsatisfied. The maximum size for a PDS is 12, while only 8 for a connected PDS.

3. Hardness results

In this section we prove several hardness results for Max PDS on split and bipartite graphs and further extend the results to prove that deciding if a PDS is inclusion-wise maximal is co-NP-complete.

We construct two polynomial-time reductions from Max Independent Set, which is known to be NP-hard [9].

Max Independent Set
Input: A graph G.
Output: A subset of pairwise non-adjacent vertices in G of maximum size.

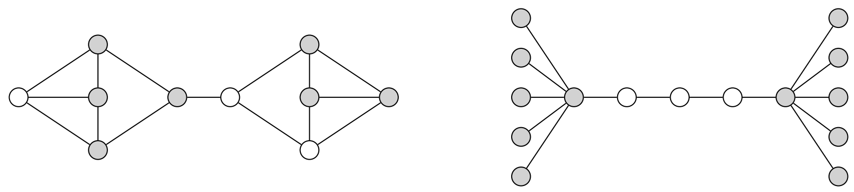

Fig. 1. Two graphs in which all PDS of maximum size are not connected. Gray vertices represent a PDS of maximum size in each graph.

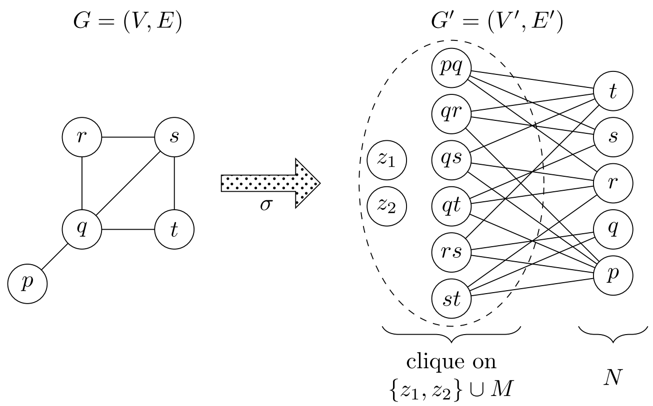

Fig. 2. The graph G' obtained from the graph G using the transformation σ.

3.1. Split graphs

We first describe a polynomial-time reduction, and then prove two intermediate results allowing us to easily prove the NP-hardness of Max PDS on split graphs.

Definition 2. Let $G = (V, E)$ be a graph not isomorphic to a star. We define the construction σ transforming the graph G into $G' := \sigma(G)$, where $G' = (V', E')$ is defined as follows:

- $V' := \{z_1, z_2\} \cup M \cup N$, where $N := V$, $M := \{uv : \{u, v\} \in E\}$ and z_1, z_2 are two additional vertices;
- for each $e \in M$ and each $u \in N$, the edge $\{e, u\} \in E'$ if and only if $u \notin e$;
- the set $M \cup \{z_1, z_2\}$ induces a clique in G'.

Obviously, the construction σ can be done in polynomial time. Notice that G' is a split graph, and is connected if and only if G is not isomorphic to a star. See Fig. 2 for an example.

Lemma 1. *Let $G = (V, E)$ be a graph not isomorphic to a star and let $G' = (V', E')$ be such that $G' = \sigma(G)$. Let $S \subset V'$ be a set of vertices such that $M \cup \{z_1, z_2\} \subseteq S$. Then a vertex $e \in M$ is satisfied in $G'[S]$ if and only if $d_S(e) \geq |S| - 2$.*

Proof. A vertex $e \in M$ has degree $d(e) = |V'| - 3$. Hence, if $d_S(e) < |S| - 2$, then $d_{\bar{S}}(e) = |\bar{S}|$ and e is not satisfied in $G'[S]$ as it does not respect Eq. (1). However, if $d_S(e) \geq |S| - 2$, then $d_{\bar{S}}(e) < |\bar{S}|$. Also, since G is connected, $|M| \geq |N| - 1$, and hence $|S| \geq |M| + 2 > |N| \geq |\bar{S}|$ and we have

$$|\bar{S}| \cdot d_S(e) \geq |\bar{S}| \cdot (|S| - 2) \geq (|\bar{S}| - 1) \cdot (|S| - 1) \geq (|S| - 1) \cdot d_{\bar{S}}(e)\,,$$

and thus e is satisfied in $G'[S]$. □

Lemma 2. *Let $G = (V, E)$ be a graph not isomorphic to a star and let $G' = (V', E')$ be such that $G' = \sigma(G)$. Let $S_1 \subset V'$ such that $G'[S_1]$ is a PDS. Then, there exists $S_2 \subset V'$ such that $G'[S_2]$ is a PDS, $|S_2| \geq |S_1|$ and $M \cup \{z_1, z_2\} \subseteq S_2$. Moreover, S_2 can be found in polynomial time.*

Proof. Firstly, we show that $N \nsubseteq S_1$.

- if $S_1 = N$, since $G'[N]$ is an independent set, then any vertex $u \in S_1$ has $d_{S_1}(u) = 0$ and $d_{\bar{S}_1}(u) > 0$; hence u does not satisfy Eq. (1) and $G'[S_1]$ is not a PDS;

- if $N \subset S_1$, then $\bar{S}_1$ is a subset of the clique $M \cup \{z_1, z_2\}$; it means any vertex $u \in S_1 \cap (M \cup \{z_1, z_2\})$ has $d_{\bar{S}_1}(u) = |\bar{S}_1|$ and $d_{S_1}(u) < |S_1| - 2$, and thus

$$|\bar{S}_1| \cdot d_{S_1}(u) < |\bar{S}_1| \cdot (|S_1| - 2) < |\bar{S}_1| \cdot (|S_1| - 1) = (|S_1| - 1) \cdot d_{\bar{S}_1}(u),$$

so u does not satisfy Eq. (1) and $G'[S_1]$ is not a PDS.

Now, let $S_2 := S_1 \cup M \cup \{z_1, z_2\}$ and $\bar{S}_2 := V' \setminus S_2$.

Observe that for any $f \in S_1 \cap M$, $d_{S_2}(f) - d_{S_1}(f) = |S_2| - |S_1| \geq 0$ and $d_{\bar{S}_2}(f) \leq d_{\bar{S}_1}(f)$. Thereby, we obtain $\frac{d_{S_2}(f)}{|S_2|-1} \geq \frac{d_{S_1}(f)}{|S_1|-1} \geq \frac{d_{\bar{S}_1}(f)}{|\bar{S}_1|} \geq \frac{d_{\bar{S}_2}(f)}{|\bar{S}_2|}$, so f is satisfied in S_2. Also, if a vertex in M is satisfied in S_2, then according to Lemma 1 it is also satisfied in any $S_2' \subseteq S_2$, as long as $M \cup \{z_1, z_2\} \subseteq S_2'$.

If there exists $e \in M \setminus S_1$ which is not satisfied in S_2, then following Lemma 1 it holds $d_{S_2}(e) < |S_2| - 2$. Thus, there exists a vertex $u \in S_2 \cap N$, non-adjacent to e, which we can transfer from S_2 to $\bar{S}_2$. Obviously, at most $|M \setminus S_1|$ transfers are needed to satisfy all the vertices in S_2, and thus $|S_2| \geq |S_1|$ holds true. Since $S_2 \cap N \subseteq S_1 \cap N$ and $N \not\subseteq S_1$, then $S_2 \neq V'$.

Note that $\bar{S}_2 \subseteq N$ and that each vertex $u \in S_1 \cap N$ is satisfied in S_2, since $d_{\bar{S}_2}(u) = 0$. Clearly, z_1 and z_2 are satisfied in S_2. Thus, $G'[S_2]$ is a PDS, and it can be found in polynomial time. □

Notice that Lemma 2 implies that there exists a PDS of maximum size in G' that is connected. Hence, the following result also holds when looking for a connected PDS.

Theorem 1. Max Proportionally Dense Subgraph *is NP-hard on split graphs.*

Proof. Let $G = (V, E)$ be a graph not isomorphic to a star, $G' = (V', E')$ be such that $G' = \sigma(G)$, and $k \in \{1, \ldots, |V| - 2\}$. Notice that since G is connected and not isomorphic to a star, then there is no independent set of size $|V| - 1$ in G. We claim that there is an independent set of size at least k in G if and only if there is a PDS of size at least $|M| + 2 + k$ in G'.

Let R be an independent set of G of size at least k. In G', we define $S := M \cup \{z_1, z_2\} \cup R$ and $\bar{S} := V' \setminus S$. First, note that $R \subseteq N$ thus $\bar{S} = N \setminus R$. The vertices in $S \cap N \cup \{z_2, z_2\}$ are obviously satisfied in $G'[S]$ as they only have neighbors in S. Hence, if there exist unsatisfied vertices, then they must be from the set M. Choose a vertex $e \in M$. Since R is an independent set of G, then for each edge $e = \{u, v\} \in E$ at most one of the vertices u and v belongs to R. Hence, the vertex $e \in M$ is not adjacent to at most one vertex in S, and thus $d_S(e) \geq |S| - 2$. According to Lemma 1, the vertex e is satisfied in $G[S]$. Thus, $G[S]$ is a PDS of size at least $|M| + 2 + k$.

Let $S \subset V'$ be of size at least $|M| + 2 + k$ such that $G'[S]$ is a PDS. According to Lemma 2, there exists $S' \subset V'$ such that $G'[S']$ is a PDS, $|S'| \geq |S|$ and $\{z_1, z_2\} \cup M \subseteq S'$. We claim that $R' := S' \cap N$ is an independent set of G of size at least k. Obviously $|R'| \geq k$. Moreover, Lemma 1 states that for all satisfied vertices $e \in M$, $d_{S'}(e) \geq |S'| - 2$. Hence, for each vertex $e \in M$ there is at most one vertex $u \in S'$ that is not adjacent to e. Since the vertices $e \in M$ and $u \in N$ are not adjacent in G', it implies that $u \in e$ in G, and therefore the edge $e \in E$ has at most one endpoint $u \in R'$ in the graph G. Thus, R' is an independent set of size at least k. □

Proposition 1. *It is NP-hard to approximate* Max Proportionally Dense Subgraph *within 1.0026028 on split graphs, and hence the problem is APX-hard (even on split graphs).*

Proof. Let I be an instance of Max Independent Set on a cubic graph $G = (V, E)$. It is known that it is NP-hard to decide whether $opt(I) < \frac{12\tau+11+2\varepsilon}{24\tau+28} \cdot |V|$ or $opt(I) > \frac{12\tau+12-2\varepsilon}{24\tau+28} \cdot |V|$, for any $\varepsilon > 0$, where $\tau \leq 6.9$ [6].

We construct an instance I' of Max PDS defined on the graph $G' = (V', E')$ such that $G' = \sigma(G)$. Note that $M \subset V'$ is of size $|E|$, that is $|M| = |E| = \frac{3|V|}{2}$ since G is cubic. From Theorem 1, we know that $opt(I') = |M| + 2 + opt(I)$. Consequently, it is NP-hard to decide whether $opt(I') < |M| + 2 + \frac{12\tau+11+2\varepsilon}{24\tau+28} \cdot |V| = \frac{48\tau+53+2\varepsilon}{24\tau+28} \cdot |V| + 2$ or $opt(I') > |M| + 2 + \frac{12\tau+12-2\varepsilon}{24\tau+28} = \frac{48\tau+54-2\varepsilon}{24\tau+28} \cdot |V| + 2$. We obtain that it is NP-hard to approximate Max PDS within 1.0026028. □

3.2. Bipartite graphs

In the following, we modify the previous construction in order to prove the NP-hardness of Max PDS on bipartite graph. The reduction will also be used to show the NP-hardness of an "extension version" of the problem, implying the co-NP-completeness of deciding if a PDS is inclusion-wise maximal.

Definition 3. Let $G = (V, E)$ be a graph not isomorphic to a star, and an integer k such that $1 \leq k < |V| - 1$. We define the construction β transforming the graph G into $G' := \beta(G, k)$, where $G' = (V', E')$ is defined as follows:

- $V' := L \cup M \cup N$, where $N := V$, $M := \{uv : \{u, v\} \in E\}$ and L contains $|L| := |M| \cdot (|V| - k - 1) - k + 1$ additional vertices;
- for each $e \in M$ and each $u \in N$, the edge $\{e, u\} \in E'$ if and only if $u \notin e$;
- for each $e \in M$ and each $v \in L$, the edge $\{e, v\} \in E'$.

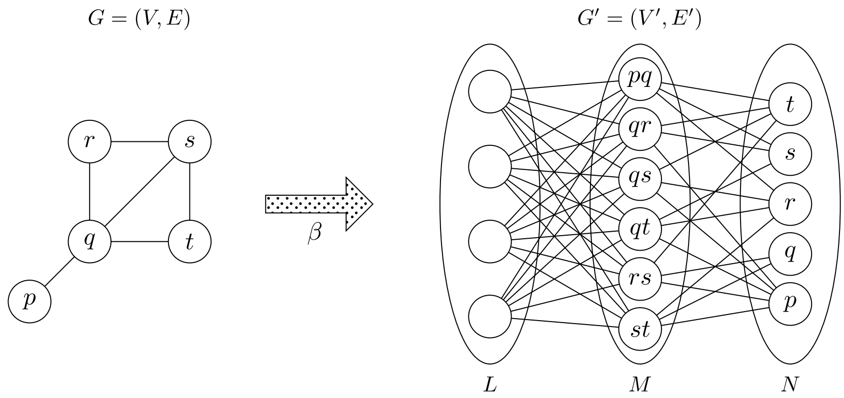

Fig. 3. The graph G' obtained from G using the transformation β and $k = 3$.

Obviously, the construction β can be done in polynomial time. Clearly, G' is connected if and only if the input graph is not isomorphic to a star. Also, notice that G' is a bipartite graph as there are edges only between M and $L \cup N$. See Fig. 3 for an example.

We now prove intermediate results, which help concluding that Max PDS is NP-complete on bipartite graphs.

Lemma 3. *Let m, n and k be integers such that $1 \leq k < n-1 \leq m$ and $\ell := m \cdot (n-k-1) - k + 1$. Then $\frac{\ell+k-1}{\ell+m+k-1} = \frac{n-k-1}{n-k}$.*

Proof. $(n-k)\cdot(\ell+k-1) = (n-k-1)\cdot(\ell+k-1)+\ell+k-1 = (n-k-1)\cdot(\ell+k+1)+m\cdot(n-k-1) = (n-k-1)\cdot(\ell+m+k+1)$. $\square$

Lemma 4. *Let $G = (V, E)$ be a graph not isomorphic to a star, k an integer such that $1 \leq k < |V| - 1$ and $G' = (V', E')$ be such that $G' = \beta(G, k)$. Let $S \subset V'$ be such that $|S| \geq |L| + |M| + k$. Then a vertex $f \in M$ is satisfied in $G'[S]$ if and only if $d_{\bar{S}}(f) < |\bar{S}|$.*

Proof. If $d_{\bar{S}}(f) = |\bar{S}|$, f is obviously not satisfied. If $d_{\bar{S}}(f) < |\bar{S}|$, then notice that $d(f) = |L| + |N| - 2 = |V'| - |M| - 2$. Therefore, $d_S(f) = d(f) - d_{\bar{S}}(f) \geq |V'| - |M| - 2 - |\bar{S}| + 1 = |S| - |M| - 1$. Also, $|\bar{S}| \leq |N| - k$. Consequently, according to Lemma 3,

$$\frac{d_S(f)}{|S|-1} = \frac{|S|-|M|-1}{|S|-1} \geq \frac{|L|+k-1}{|L|+|M|+k-1} = \frac{|N|-k-1}{|N|-k} \geq \frac{d_{\bar{S}}(f)}{|\bar{S}|}. \quad \square$$

Lemma 5. *Let $G = (V, E)$ be a graph not isomorphic to a star, k an integer, $1 \leq k < |V| - 1$, and let $G' = (V', E')$ be such that $G' = \beta(G, k)$. Let $S_1 \subset V'$ such that $G'[S_1]$ is a PDS and $|S_1| \geq |L| + |M| + k$. Then, there exists $S_2 \subset V'$ such that $G'[S_2]$ is a PDS, $|S_2| \geq |S_1|$ and $L \cup M \subseteq S_2$. Moreover, S_2 can be found in polynomial time.*

Proof. First, we prove that $M \subset S_1$. As $|S_1| \geq |L| + |M| + k > |M| + |N|$, then $S_1 \cap L \neq \emptyset$. Take a vertex $z \in S_1 \cap L$ and notice that since $d(z) = |M|$, then $d_{\bar{S}_1}(z) = |M \setminus S_1|$. The vertex z is satisfied in $G'[S_1]$ if and only if

$$\frac{|M| - d_{\bar{S}_1}(z)}{|L|+|M|+k-1} \geq \frac{d_{S_1}(z)}{|S_1|-1} \geq \frac{d_{\bar{S}_1}(z)}{|\bar{S}_1|} \geq \frac{d_{\bar{S}_1}(z)}{|N|-k}.$$

This implies that

$$\begin{aligned} & |M| \cdot (|N|-k) - d_{\bar{S}_1}(z) \cdot (|N|-k) \geq d_{\bar{S}_1}(z) \cdot (|L|+|M|+k-1) \\ \iff & |M| \cdot (|N|-k) - d_{\bar{S}_1}(z) \cdot (|N|-k) \geq d_{\bar{S}_1}(z) \cdot |M| \cdot (|N|-k) \\ & \iff |M| \cdot (|N|-k) \geq d_{\bar{S}_1}(z) \cdot (|M|+1) \cdot (|N|-k) \\ & \iff 0 \geq d_{\bar{S}_1}(z). \end{aligned}$$

Thus, we have $d_{\bar{S}_1}(z) = 0$ and conclude that $M \subset S_1$.

Let $S_2 := S_1 \cup L \cup M$ and $f \in M$. As f is satisfied in $G'[S_1]$, according to Lemma 4, we have $d_{\bar{S}_1}(f) < |\bar{S}_1|$. Since f is connected to all the vertices in L, necessarily $d_{\bar{S}_2}(f) < |\bar{S}_2|$ and f remains satisfied in $G'[S_2]$. Obviously, the vertices in L are satisfied in $G'[S_2]$ since all their neighbors are in M. This is also true for the vertices in $N \cap S_2$. $\square$

Notice that Lemma 5 implies that there exists a PDS of maximum size that is connected in G'. Hence, the following result also holds when looking for a connected PDS.

Theorem 2. Max Proportionally Dense Subgraph *is NP-hard on bipartite graphs.*

Proof. Let $G = (V, E)$ be a graph not isomorphic to a star, $k \in \{1, \dots, |V| - 1\}$. Notice that since G is connected and not isomorphic to a star, then there is no independent set of size $|V| - 1$ in G. Let $G' = (V', E')$ such that $G' = \beta(G, k)$. We claim that there is an independent set of size at least k in G if and only if there is a PDS of size at least $|L| + |M| + k$ in G'.

Let R be an independent set of G of size at least k. In G', we define $S := L \cup M \cup R$ and $\bar{S} := V' \setminus S$. First, note that $R \subseteq N$ thus $\bar{S} = N \setminus R$. The vertices in $L \cup R$ are obviously satisfied in $G'[S]$ as all their neighbors are in S. Hence, if there exist vertices not satisfied in $G'[S]$, then they must belong to the set M. Consider a vertex $e \in M$. Since R is an independent set of G, then for each edge $e = \{u, v\} \in E$ at most one of the vertices u and v belongs to R, and, therefore, at least one belongs to $\bar{S}$. Therefore, the vertex $e \in M$ is not adjacent to at least one vertex in $\bar{S}$, and thus $d_{\bar{S}}(f) < |\bar{S}|$. According to Lemma 4, e is satisfied in $G[S]$. Thus, $G[S]$ is a PDS of size at least $|L| + |M| + k$.

Let $S \subset V'$ be of size at least $|L| + |M| + k$ such that $G'[S]$ is a PDS. According to Lemma 5, there exists $S' \subset V'$ such that $G'[S']$ is a PDS, $|S'| \geq |S|$ and $L \cup M \subseteq S'$. We claim that $R' := S' \cap N$ is an independent set of G of size at least k. Obviously $|R'| \geq k$. Lemma 4 states that for all satisfied vertices $e \in M$, $d_{\bar{S}'}(e) < |\bar{S}'|$. Therefore, as $d_N(e) = |N| - 2$ and $\bar{S}' \subseteq N$, there is at most one vertex $u \in S' \cap N$ not adjacent to e. From the construction σ, if there is no edge between the vertices $e \in M$ and $u \in N$ in G', then $u \in e$ in G. Hence, the edge $e \in E$ in G has at most one vertex $u \in R'$. Thus, R' is an independent set of size at least k. □

Below, we prove that deciding if a subset of vertices can be extended into a larger subset which induces a PDS is NP-complete. We obtain as a corollary that deciding if a PDS is inclusion-wise maximal is co-NP-complete.

PDS Extension
Input: A graph $G = (V, E)$, $U \subset V$.
Question: Is there a vertex subset $S \subset V$ such that $U \subset S$ and $G[S]$ is a proportionally dense subgraph?

To prove that PDS Extension is NP-complete, we use again the construction β as defined in Definition 3.

Lemma 6. *Let $G = (V, E)$ be a graph not isomorphic to a star, k an integer, $1 \leq k < |V| - 1$, and $G' = (V', E')$ be such that $G' = \beta(G, k)$. Let $S \subset V'$ be such that $L \cup M \subset S$ and $G'[S]$ is a PDS. Then $|S| \geq |L| + |M| + k$.*

Proof. Let $u \in S \cap N$, and notice that $d_S(u) < |M|$, so there exists a vertex in M which is not connected to u. Let $f \in M$ be such a vertex. Note that $d_S(f) \leq |S| - |M| - 1$ and $d_{\bar{S}}(f) \geq |\bar{S}| - 1$, as f is not connected to u. Let $k' := |N \setminus \bar{S}| = |N| - |\bar{S}|$. We claim that $k' \geq k$. Suppose by contradiction that $k' < k$. Then $\frac{|L|+k'-1}{|L|+|M|+k'-1} < \frac{|L|+k-1}{|L|+|M|+k-1}$ and $\frac{|N|-k-1}{|N|-k} < \frac{|N|-k'-1}{|N|-k'}$. According to Lemma 3, we conclude that $\frac{|L|+k'-1}{|L|+|M|+k'-1} < \frac{|N|-k'-1}{|N|-k'}$. Therefore,

$$\frac{d_S(f)}{|S| - 1} \leq \frac{|L| + k' - 1}{|L| + |M| + k' - 1} < \frac{|N| - k' - 1}{|N| - k'} \leq \frac{d_{\bar{S}}(f)}{|\bar{S}|},$$

which contradicts that f is satisfied, and thus that $G'[S]$ is a PDS. We conclude that $|S| = |L| + |M| + k' \geq |L| + |M| + k$. □

Theorem 3. PDS Extension *is NP-complete on bipartite graphs.*

Proof. Obviously, PDS Extension is in NP. Let $G = (V, E)$ be a graph not isomorphic to a star, $k \in \{1, \dots, |V| - 1\}$. Notice that since G is connected and not isomorphic to a star, then there is no independent set of size $|V| - 1$ in G. Let $G' = (V', E')$ such that $G' = \beta(G, k)$. We claim that there is an independent set of size at least k in G if and only if there is PDS of size of size at least $|L| + |M| + k$ in G'.

Assume there exists an independent set of size k in G. Then, there exists $S \subset V'$ of size $|S| \geq |L| + |M| + k$ such that $G'[S]$ is a PDS, and $L \cup M \subset S$ (see proof of Theorem 2).

According to Lemma 6, if there exists $S \subset V'$ such that $G'[S]$ is a PDS and $L \cup M \subset S$, then $|S| \geq |L| + |M| + k$. Therefore, there exists an independent set of size at least k in G (see proof of Theorem 2).

We conclude that deciding if there exists $S \subset V'$ such that $L \cup M \subset S$ and $G'[S]$ is a PDS is NP-complete, and thus that PDS Extension is NP-complete on bipartite graphs. □

Notice that the set $L \cup M$ is connected, and thus if it can be extended into a PDS, then the PDS is connected. Hence, it is NP-complete to decide whether a vertex subset (inducing a connected subgraph) can be extended into a connected PDS. Furthermore, the set $L \cup M$ can induce a PDS or not, depending on the values of k and $|V|$. Indeed, $G'[L \cup M]$ is a PDS if and only if $\frac{|L|}{|L|+|M|-1} \geq \frac{|N|-2}{|N|}$, which implies $k \leq \frac{n}{2}$. Therefore, we conclude that deciding if a PDS is inclusion-wise maximal is co-NP-complete.

Corollary 1. *Let $G = (V, E)$ be a graph and $S \subset V$ such that $G[S]$ a proportionally dense subgraph. Deciding if S is inclusion-wise maximal is co-NP-complete on bipartite graphs.*

4. Approximation

In this section we show that there exists a polynomial-time 2-approximation algorithm for Max Proportionally Dense Subgraph, which establishes the APX-completeness of the problem. When the maximum degree Δ of the graph is bounded, the approximation ratio can be further improved to $(2 - \frac{2}{\Delta+1})$ using a better upper bound on the size of a PDS.

Lemma 7. *Let $G = (V, E)$ be a graph and $S \subset V$ such that $G[S]$ is not a proportional dense subgraph. If $|S| = \lceil \frac{|V|}{2} \rceil$, then there exists $u \in S$ such that $d_S(u) < d_{\bar{S}}(u)$. Moreover, if $|V|$ is even and $|S| = \frac{|V|}{2} + 1$, then there exists $u \in S$ such that $d_S(u) \leq d_{\bar{S}}(u)$.*

Proof. Let $S \subset V$ be a subset such that $G[S]$ is not a PDS. Then, there exists a vertex $u \in S$ such that Eq. (1) is not satisfied in $G[S]$, and therefore $|\bar{S}| \cdot d_S(u) < (|S| - 1) \cdot d_{\bar{S}}(u)$ $(*)$.

- If $|S| = \lceil \frac{|V|}{2} \rceil$, the inequality $(*)$ implies $\lfloor \frac{|V|}{2} \rfloor \cdot d_S(u) < (\lceil \frac{|V|}{2} \rceil - 1) \cdot d_{\bar{S}}(u) \leq \lfloor \frac{|V|}{2} \rfloor \cdot d_{\bar{S}}(u)$, and hence $d_S(u) < d_{\bar{S}}(u)$.
- If $|S| = \frac{|V|}{2} + 1$ ($|V|$ even), assume by contradiction that for each vertex $v \in S$ it holds $d_S(v) > d_{\bar{S}}(v)$. In particular, the inequality $(*)$ implies $(\frac{|V|}{2} - 1) \cdot (d_{\bar{S}}(u) + 1) < \frac{|V|}{2} \cdot d_{\bar{S}}(u)$, which is true if and only if $d_{\bar{S}}(u) \geq \frac{|V|}{2}$. Thus, $d(u) = d_S(u) + d_{\bar{S}}(u) > |V| - 1$, a contradiction. $\square$

Theorem 4. *For any graph $G = (V, E)$, a proportionally dense subgraph of size $\lceil \frac{|V|}{2} \rceil$ or $\lceil \frac{|V|}{2} \rceil + 1$ can be constructed in $\mathcal{O}(|V| \cdot |E|)$ time.*

Proof. First, we show that Algorithm 1 terminates and returns a PDS of size $\lceil \frac{|V|}{2} \rceil$ or $\lceil \frac{|V|}{2} \rceil + 1$.

Algorithm 1: Find a proportional dense subgraph of size $\lceil \frac{|V|}{2} \rceil$ or $\lceil \frac{|V|}{2} \rceil + 1$.

Input: $G = (V, E)$ a graph.
Output: $S \subset V$ such that $G[S]$ is a PDS.
1 Let $S \subset V$ with $|S| = \lceil \frac{|V|}{2} \rceil$;
2 **while** *$G[S]$ is not a PDS* **do**
3   Let $u \in S$ such that $d_{\bar{S}}(u) - d_S(u)$ is maximum;
4     $S := \bar{S} \cup \{u\}$;
5 **return** S;

- **Case 1:** $|V|$ **is odd.** Notice that at the end of each loop, the set S is modified without changing its size $|S| = \frac{|V|+1}{2} = \lceil \frac{|V|}{2} \rceil$. If $G[S]$ is not a PDS, then according to Lemma 7 there exists an unsatisfied vertex $v \in S$ for which $d_S(v) < d_{\bar{S}}(v)$. Therefore, the vertex u chosen within the loop has the property $d_{\bar{S}}(u) - d_S(u) > 0$. Thus, the size of the cut between S and $\bar{S}$ decreases after each loop and the algorithm terminates.
- **Case 2:** $|V|$ **is even.** Notice that Algorithm 1 starts with $|S| = \frac{|V|}{2}$. If $G[S]$ is not a PDS, then due to Lemma 7, there exists a vertex $v \in S$ such that $d_S(v) < d_{\bar{S}}(v)$. The selection of the vertex $u \in S$ inside the loop ensures that the size of the cut between S and $\bar{S}$ strictly decreases at the end of the loop. Now, observe that after the first loop, $|S| = \frac{|V|}{2} + 1$. If $G[S]$ is not a PDS, according to Lemma 7, there exists a vertex $v \in S$ such that $d_S(v) \leq d_{\bar{S}}(v)$. Therefore, the vertex u inside the loop has $d_S(u) \leq d_{\bar{S}}(u)$. Obviously, after the second loop, $|S| = \frac{|V|}{2}$. Since after each loop $|S|$ alternates between $\frac{|V|}{2}$ and $\frac{|V|}{2} + 1$, the cut between S and $\bar{S}$ strictly decreases every two loops, and the algorithm terminates.

It is easy to see that the while-loop is called at most $\mathcal{O}(|E|)$ times. Now, we prove how one can obtain a $\mathcal{O}(|V| \cdot |E|)$ running time by computing Lines 2 to 4 in $\mathcal{O}(|V|)$ time.

Preprocessing. Once S has been defined at Line 1, compute and store the following *properties* for each vertex $u \in V$: $d_S(u)$, $d_{\bar{S}}(u)$, and whether u belongs to S or $\bar{S}$. The computation of these properties for all the vertices can be done in $\mathcal{O}(|E|)$ time. While computing the properties, one can also choose a vertex $u \in S$ that maximizes $d_{\bar{S}}(u) - d_S(u)$ (as in Line 3).

Main loop. If $d_{\bar{S}}(u) - d_S(u) > 0$, then S is not a PDS. However, if $d_{\bar{S}}(u) - d_S(u) = 0$, then S is a PDS if and only if $|S| < \frac{|V|}{2} + 1$ (so we decide Line 2 in constant time). Therefore, if S is not a *PDS*, set $S := \bar{S} \cup \{u\}$ (as in Line 4), update the properties of all the vertices and select $u \in S$ maximizing $d_{\bar{S}}(u) - d_S(u)$ (as in Line 3) in $\mathcal{O}(|V|)$. Then, repeat from the beginning of the main loop. $\square$

Corollary 2. Max Proportionally Dense Subgraph *is polynomial-time* 2*-approximable.*

Proof. For any graph $G = (V, E)$, Algorithm 1 yields a PDS of size at least $\lceil \frac{|V|}{2} \rceil$ and since any PDS has size at most $|V| - 1$, we obtain a 2-approximation algorithm. $\square$

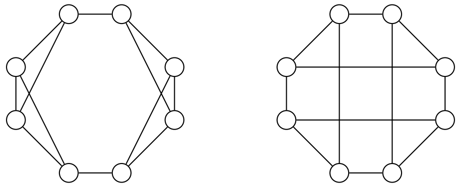

Fig. 4. Two Hamiltonian cubic graphs with 8 vertices without PDS of size $\lfloor \frac{2\times 8+1}{3} \rfloor = 5$.

We proved the APX-hardness of Max PDS in Proposition 1, and hence we conclude the APX-completeness of the problem.

Corollary 3. Max Proportionally Dense Subgraph *is APX-complete.*

In the following we show how the approximation ratio can be improved with regard to the maximum degree Δ of the graph.

Lemma 8. *Let $G = (V, E)$ be a graph and $S \subset V$ such that $G[S]$ is a proportionally dense subgraph. Then $|S| \leq \lfloor \frac{|V|\cdot(\Delta(G)-1)+1}{\Delta(G)} \rfloor$.*

Proof. Let v be a vertex of S with at least one neighbor in $\bar{S} = V \setminus S$ (such a vertex exists since G is connected). Since $G[S]$ is a PDS, v fulfills the proportion condition, that is $\frac{\Delta(G)-1}{|S|-1} \geq \frac{d_S(v)}{|S|-1} \geq \frac{d_{\bar{S}}(v)}{|\bar{S}|} \geq \frac{1}{|V|-|S|}$ which implies that $|S| \leq \frac{|V|\cdot(\Delta(G)-1)+1}{\Delta(G)}$, and hence $|S| \leq \lfloor \frac{|V|\cdot(\Delta(G)-1)+1}{\Delta(G)} \rfloor$. $\square$

Proposition 2. Max Proportionally Dense Subgraph *is polynomial-time $(2 - \frac{2}{\Delta+1})$-approximable.*

Proof. Let $G = (V, E)$ be a graph, S be a set returned by Algorithm 1 and $opt(G)$ denote the size of a PDS of maximum size in G. According to Lemma 8 we have $opt(G) \leq \frac{|V|\cdot(\Delta(G)-1)+1}{\Delta(G)}$. Therefore, since $|V| \geq \Delta + 1$ and $opt(G) \geq |S|$, we obtain

$$\frac{opt(G)}{|S|} \leq \frac{2\cdot opt(G)}{|V|} \leq \frac{2\cdot(|V|\cdot(\Delta-1)+1)}{|V|\cdot\Delta} \leq \frac{2\cdot((\Delta+1)\cdot(\Delta-1)+1)}{(\Delta+1)\cdot\Delta} = 2 - \frac{2}{\Delta+1}. \quad \square$$

Algorithm 1 shows that the decision version associated with Max PDS is in FPT when parameterized by its natural parameter k (*i.e.* the size of a PDS). Indeed, if the parameter $k \leq \lceil \frac{|V|}{2} \rceil$, then a PDS of size greater than k can be found in polynomial time using Algorithm 1. On the other hand, if $k > \lceil \frac{|V|}{2} \rceil$, then we have $|V| < 2k$ and an exhaustive search can be done in $O(2^{2k})$ operations.

5. Hamiltonian cubic graphs

In this section we prove that all Hamiltonian cubic graphs of order n, except two graphs (see Fig. 4), have a proportionally dense subgraph of the maximum possible size $\lfloor \frac{2n+1}{3} \rfloor$ (see Lemma 8 for an upper bound on a PDS size). Furthermore, we show that such a PDS can be found in linear time if a Hamiltonian cycle is given in the input. Note that almost all cubic graphs are Hamiltonian, as proved in [13].

We represent a Hamiltonian cubic graph of order n as a cycle with the vertices labeled in such a way that $(0, 1, \ldots, n-1)$ is a Hamiltonian cycle and a set of edges between non-successive vertices in the Hamiltonian cycle. We always refer to this cycle when we say *the* Hamiltonian cycle of a graph. To avoid tedious notations, we use $i \in \mathbb{N}$ (with $0 \in \mathbb{N}$) to refer to the vertex labeled by i mod n.

Definition 4. Let $G = (V, E)$ be a Hamiltonian cubic graph, $u \in V$. Let P be a set of successive vertices in the Hamiltonian cycle labeled with $u, u+1, \ldots, u-k-1$, with k such that $|V| - 2 \geq k \geq 2$. The set P is called a *shift* if the first and the last vertices of the sequence, u and $u-k-1$, are such that $d_P(u) = d_P(u-k-1) = 2$.

Notice that a shift P contains $|V| - k$ vertices. Also, any vertex of P has at least two neighbors in P. Consequently, if $k \geq \lceil \frac{|V|-1}{3} \rceil$, then $|P| \leq \lfloor \frac{2\cdot|V|+1}{3} \rfloor$, and the following holds for any $u \in P$:

$$\frac{d_P(u)}{|P|-1} \geq \frac{2}{|V|-k-1} \geq \frac{1}{k} \geq \frac{d_{\bar{P}}(u)}{|\bar{P}|}.$$

Thus, $G[P]$ is a PDS. If $k = \lceil \frac{|V|-1}{3} \rceil$, then $G[P]$ is a PDS of the maximum possible size $\lfloor \frac{2\cdot|V|+1}{3} \rfloor$ (see Lemma 8) and we call P a *good shift*. On the other hand, if $k = \lceil \frac{|V|-1}{3} \rceil - 1$, then the size of P is one vertex larger than the size of the maximum possible PDS, and thus $G[P]$ is not a PDS. Such a shift is called an *almost good shift*.

In the following, we prove that either G contains a good shift or we can find an almost good shift P and a vertex $v \in P$ such that $G[P \setminus \{v\}]$ is a proportionally dense subgraph of the maximum possible size $\lfloor \frac{2\cdot|V|+1}{3} \rfloor$.

Definition 5. Let $G = (V, E)$ be a Hamiltonian cubic graph. For each $v \in V$, we denote by $c(v)$ the non-successive neighbor of v in the Hamiltonian cycle. Additionally, we define the subsets of vertices L and R in the following way for $k := \lceil \frac{|V|-1}{3} \rceil$:

- $L := \{u \in V : c(u) \in \{u-k, u-k+1, \dots, u-2\}\}$;
- $R := \{u \in V : c(u) \in \{u+2, u+3, \dots, u+k\}\}$.

For a Hamiltonian cubic graph $G = (V, E)$ and $u \in V$, notice that $u \in L$ if and only if $c(u) \in R$, and symmetrically $u \in R$ if and only if $c(u) \in L$. This particularly implies that $|L| = |R| \leq \frac{|V|}{2}$. Moreover, notice that for a vertex $u \in L$, the set $P := \{u, u+1, \dots, u-k-1\}$ cannot be a good shift, since $d_P(u) = 1$. In the same way, if $u \in R$, the set $P := \{u+k+1, u+k+2, \dots, u-1, u\}$ cannot be a good shift, since $d_P(u) = 1$. These observations are summed up in the following lemma.

Lemma 9. *Let $G = (V, E)$ be a Hamiltonian cubic graph, $k := \lceil \frac{|V|-1}{3} \rceil$ and $u \in V$. If $u \notin L$ and $(u-(k+1)) \notin R$, then the set $\{u, u+1, \dots, u-(k+1)-1, u-(k+1)\}$ is a good shift. Symmetrically, if $u \notin R$ and $(u+k+1) \notin L$, then the set $\{u+k+1, u+k+2, \dots, u-1, u\}$ is a good shift.*

Proof. The proof is straightforward. Since $u \notin L$ and $(u-(k+1)) \notin R$, we have $d_P(u) = d_P(u-(k+1)) = 2$, where $P := \{u, u+1, \dots, u-(k+1)\}$. The other case is similar. $\square$

An important consequence of Lemma 9 is that if G is a Hamiltonian cubic graph with no good shift, then we can define subsets of vertices that must be either in L or in R. To define such subsets we introduce the following notation.

Definition 6. Let $G = (V, E)$ be a Hamiltonian cubic graph and $u \in V$. We define the vertex subset $\langle u \rangle := \{v \in V : v \equiv u \pmod{(k+1)}\}$ where $k := \lceil \frac{|V|-1}{3} \rceil$.

Corollary 4. *Let $G = (V, E)$ be a Hamiltonian cubic graph with no good shift and $u \in V$:*

- *if $u \notin R$ then $\langle u \rangle \subseteq L$,*
- *if $u \notin L$, then $\langle u \rangle \subseteq R$,*
- *$|L| = |R| = \frac{|V|}{2}$.*

Proof. First, notice that for any integer $\delta \geq 1$, $u - \delta\cdot(k+1) \equiv u - \delta\cdot(k+1) + |V|\cdot\delta\cdot(k+1) \pmod{|V|} \equiv u + \delta\cdot(|V|-1)\cdot(k+1) \pmod{|V|}$. Moreover, $u \equiv u + |V|\cdot(k+1) \pmod{|V|}$. Thus, we have $\{u - \delta\cdot(k+1) : \delta \geq 1, \delta \in \mathbb{N}\} = \{u + \delta\cdot(k+1) : \delta \geq 1, \delta \in \mathbb{N}\} = \langle u \rangle$.

Now, if $u \notin R$, then, with our assumption that G has no good shift and Lemma 9, we derive that $\langle u \rangle = \{u + \delta\cdot(k+1) : \delta \geq 1, \delta \in \mathbb{N}\} \subseteq L$. Symmetrically, if $u \notin L$, then $\{u - \delta\cdot(k+1) : \delta \geq 1, \delta \in \mathbb{N}\} \subseteq R$.

This implies that for any vertex $u \in V$, either $u \in L$ or $u \in R$. Finally, since $u \in L$ if and only if $c(u) \in R$ and $u \in R$ if and only if $c(u) \in L$, then it is obvious that $|L| = |R| = \frac{|V|}{2}$. $\square$

Let $G = (V, E)$ be a Hamiltonian cubic graph with no good shift and $d := \gcd(k+1, |V|)$, where $\gcd(k+1, |V|)$ is the greatest common divisor of $(k+1)$ and $|V|$. We show that V can be partitioned into d subsets of vertices $\langle 0 \rangle$, $\langle 1 \rangle$, …, $\langle d-1 \rangle$. This partition will be useful to find an *almost good shift* P and a vertex to remove from P in order to obtain a PDS in G. This result comes from a basic property of the cyclic group $\mathbb{Z}/n\mathbb{Z}$ that we recall in the following lemma.

Lemma 10. *Let $\alpha \geq 1$ and $\beta \geq 1$ be positive integers, and $d := \gcd(\alpha, \beta)$. If all integers are considered* mod α*, then $\{0, 1, \dots, \alpha-1\} = \cup_{i\in\{0,1,\dots,d-1\}} \langle i \rangle$ where $\langle i \rangle := \{l : l \equiv i \pmod{\beta} \text{ and } l \in \{0, 1, \dots, \alpha-1\}\}$. Moreover, for any $i, j \in \{0, 1, \dots, d-1\}$ with $i \neq j$, $\langle i \rangle \cap \langle j \rangle = \emptyset$.*

Proof. First, we prove that for any $u \geq d$, $u \in \langle i \rangle$ for some $i \in \{0, 1, \dots, d-1\}$. Let $u \geq d$. Then there exist two integers a, b with $b \leq d-1$, such that $u = a\cdot d + b$. Moreover, there exist two integers c, f such that $c\cdot\beta + f\cdot\alpha = d$ since $d = \gcd(\alpha, \beta)$. Then, $u = a\cdot c\cdot\beta + a\cdot f\cdot\alpha + b \equiv b + a\cdot c\cdot\beta \pmod{\alpha}$. Thus, $u \in \langle b \rangle$ with $b \leq d-1$. This proves that any integer is in a set $\langle i \rangle$ for some $i \leq d-1$, *i.e.* $\{0, 1, \dots, \alpha-1\} = \cup_{i\in\{0,1,\dots,d-1\}} \langle i \rangle$.

To prove the second part of the statement, we first show that $\alpha = |\langle u \rangle|\cdot d$ for any $u \in \{0, 1, \dots, d-1\}$. Let $u \in \{0, 1, \dots, d-1\}$ and $p \geq 1$ be the smallest integer such that $u + p\cdot\beta \equiv u \pmod{\alpha}$. Notice that $|\langle u \rangle| = p$ and let us show that $\alpha = p\cdot d$. Let α', β' be two integers such that $\alpha = \alpha'\cdot d$, $\beta = \beta'\cdot d$ and $\gcd(\alpha', \beta') = 1$. We prove that $\alpha' = p$ by verifying that α' divides p and p divides α'. First, notice that $u + \alpha'\cdot\beta = u + \alpha'\cdot k'\cdot d = u + \alpha\cdot\beta' \equiv u \pmod{\alpha}$.

Thus, p divides α'. On the other hand, recall that $u + p \cdot \beta \equiv u \pmod{\alpha}$ and notice that $u + p \cdot \beta = u + p \cdot \beta' \cdot d$, then $p \cdot \beta' \cdot d \equiv 0 \pmod{\alpha}$. This implies that α divides $p \cdot \beta' \cdot d$, and thus α' divides $p \cdot \beta'$. Since $\gcd(\alpha', \beta') = 1$, α' divides p. Now, notice that two sets $\langle i \rangle$, $\langle j \rangle$ for some integers i, j are either equal or disjoint. Since for any $u \in \{0, 1, \dots, \alpha - 1\}$ we have $|\langle u \rangle| = \frac{\alpha}{d}$, then obviously all sets $\langle i \rangle$, $i \in \{0, 1, \dots, d-1\}$ are disjoints. $\square$

In the following lemma we summarize the possible values of $\gcd(n, k+1)$ for some specific values of n and k.

Lemma 11. *Let n be an even integer, $n \geq 4$. Then:*

- *if $n = 3k - 1$, then $\gcd(n, k+1) \in \{2, 4\}$,*
- *if $n = 3k$, then $\gcd(n, k+1) \in \{1, 3\}$,*
- *if $n = 3k + 1$, then $\gcd(n, k+1) = 2$.*

Proof. Consider the case $n = 3k - 1$, then $d := \gcd(k+1, 3k-1) = \gcd(k+1, 3k-1-2(k+1)) = \gcd(k+1, k-3) = \gcd(4, k-3)$. As n is even, then k is odd and $d \in \{2, 4\}$. The other cases can be proved using the same reasoning. $\square$

Firstly, we show that if $|V| = 3k$, then there is always a good shift in G.

Corollary 5. *Let G be a Hamiltonian cubic graph with $3k$ vertices, $k \geq 2$. Then G has a good shift.*

Proof. Suppose by contradiction that there is no good shift in $G = (V, E)$. Notice that if $|V| = 3k$, then $k = \lceil \frac{|V|-1}{3} \rceil$. Let $d := \gcd(k+1, |V|)$. From Lemma 11 we get $d \in \{1, 3\}$. According to Corollary 4, $|L| = |R| = \frac{|V|}{2}$. If $d = 1$, then $V = \langle 0 \rangle$ (Lemma 10), and hence $V = L$ or $V = R$, which is impossible. If $d = 3$, then $|V| = \langle 0 \rangle \cup \langle 1 \rangle \cup \langle 2 \rangle$ (Lemma 10). According to Corollary 4, $\langle i \rangle \subseteq L$ or $\langle i \rangle \subseteq R$ for any $i \in \{0, 1, 2\}$, and thus $|R| \neq |L|$, which is not possible. $\square$

From Lemmas 10 and 11, if a Hamiltonian cubic graph $G = (V, E)$ has no good shift, then V can be written as $V = \langle 0 \rangle \cup \langle 1 \rangle \cup \langle 2 \rangle \cup \langle 3 \rangle$ (we may have $\langle 0 \rangle = \langle 2 \rangle$ and $\langle 1 \rangle = \langle 3 \rangle$). Hence, those graphs can be split into two categories:

- *type RLRL:* for any vertices $i, i+1$ with $i \in V$, we have $i \in L$ and $i+1 \in R$, or $i \in R$ and $i+1 \in L$. In this case, we always assume without loss of generality that $R = \langle 0 \rangle \cup \langle 2 \rangle$ and $L = \langle 1 \rangle \cup \langle 3 \rangle$.
- *type RRLL:* there exist two vertices $i, i+1$ with $i \in V$ such that $i, i+1 \in L$ or $i, i+1 \in R$. In this case, we always assume without loss of generality that $R = \langle 0 \rangle \cup \langle 1 \rangle$ and $L = \langle 2 \rangle \cup \langle 3 \rangle$.

Now, we show that if a Hamiltonian cubic graph G has no good shift, then there exists an almost good shift P in G (Lemma 12) and a vertex $v \in P$ such that $G[P \setminus \{v\}]$ is a PDS (Lemma 13 and Theorem 5).

Lemma 12. *Any Hamiltonian cubic graph with no good shift has an almost good shift.*

Proof. Let $G = (V, E)$ be a Hamiltonian cubic graph with no good shift, $k = \lceil \frac{|V|-1}{3} \rceil$ and $d := \gcd(k+1, |V|)$. Since G has no good shift, according to Lemma 11 and Corollary 5, $d \in \{2, 4\}$ and $|V| = 3k - 1$ or $|V| = 3k + 1$. From Corollary 4, we know that each vertex in V belongs to either L or R.

- Case 1: *G is of type RLRL.* Let $P := \{0, 1, \dots, -k\}$. Since $|V|$ is even, then $|P|$ is even. Therefore, since two vertices $i, i+1 \in P$ do not both belong to L or R, then the vertex $-k$ belongs to L. Then the set P fulfills the requirements.
- Case 2: *G is of type RRLL.* Consider the set $P := \{1, 2, \dots, -k+1\}$. According to Lemma 11, since $d = 4$, $|V| = 3k - 1$. Hence, $-k + 1 = 2 - (k+1) \in \langle 2 \rangle$. Thus, $-k + 1 \in L$ and P fulfills the requirements. $\square$

Recall that the graphs H_1 and H_2 from Fig. 4 have no proportionally dense subgraph of the maximum possible size. In Theorem 5, we show that these are the only cubic Hamiltonian graphs with this property.

Before proving the main theorem, we first deal with small graphs ($|V| < 20$) that are particular cases that need to be treated independently.

Lemma 13. *Let $G = (V, E)$ be a Hamiltonian cubic graph not isomorphic to H_1 or H_2 with $|V| < 20$. Then there exists a PDS of size $\lfloor \frac{2 \cdot |V|+1}{3} \rfloor$ in G.*

Proof. Let $k = \lceil \frac{|V|-1}{3} \rceil$. Since G is cubic, its number of vertices is even. From Lemma 11, $\gcd(k+1, |V|) \in \{1, 2, 3, 4\}$. If $\gcd(k+1, |V|) \in \{1, 3\}$, then there exists a good shift from Corollary 5. We then suppose that $\gcd(k+1, |V|) \in \{2, 4\}$. The following cases remain:

- If $|V| = 4$, then G is the complete graph K_4, and any set of 3 vertices induces a PDS of size $\lfloor \frac{2 \cdot 4+1}{3} \rfloor$.
- If $|V| = 8$, we claim that G must have a good shift. By contradiction, suppose that G has no good shift. If G is of *type RRLL* then G is isomorphic to H_1, and if G is of *type RLRL* then G is isomorphic to H_2, which is impossible since we assumed that G is not isomorphic to H_1 or H_2.

- If $|V| = 10$ and G has no good shift, since $\gcd(k+1, |V|) = 2$, G is necessarily of *type RLRL* and $c(0) = 3$, $c(1) = 8$, $c(2) = 5$, $c(4) = 7$, $c(6) = 9$. In this case, $V \setminus \{0, 6, 9\}$ induces a PDS of size $\lfloor \frac{2\cdot 10+1}{3} \rfloor$.
- If $|V| = 14$, if G has no good shift, since $\gcd(k+1, |V|) = 2$, then G is necessarily of *type RLRL*. Following Lemma 12, let $P := \{0, 1, \dots, 9\}$ be an almost good shift and:
  - If $c(6) \neq 9$, notice that $c(7), c(5) \in P$ (since $5, 7 \in L$) and $c(6) \in V \setminus P$. Thus, $G[P \setminus \{6\}]$ is a PDS of size $\lfloor \frac{2\cdot 14+1}{3} \rfloor$. If $c(3) \neq 0$, the case is symmetrical.
  - If $c(3) = 0$ and $c(6) = 9$, notice that $c(3) \in P$, $c(5) \in P$ and $d_P(c(4)) = 3$ since $c(4) \neq 9$. Thus, $G[P \setminus \{4\}]$ is a PDS of size $\lfloor \frac{2\cdot 14+1}{3} \rfloor$.
- If $|V| = 16$, if G has no good shift, since $\gcd(k+1, |V|) = 2$, G is necessarily of *type RLRL*. Following Lemma 12, let $P := (0, 1, \dots, -k)$ be an almost good shift. Since $0 \in R$, we have either $c(0) = 3$ or $c(0) = 5$. In each case, the graph is completely determined due to the constraints. In the first case, $P \setminus \{4\}$ induces a PDS of size $\lfloor \frac{2\cdot 16+1}{3} \rfloor$. In the second case, $P \setminus \{3\}$ induces a PDS of size $\lfloor \frac{2\cdot 16+1}{3} \rfloor$.

In each case, if G is not isomorphic to H_1 or H_2, then either G has a good shift which is a PDS of size $\lfloor \frac{2\cdot|V|+1}{3} \rfloor$, or we give a PDS of such size. $\square$

Theorem 5. *Let $G = (V, E)$ be a Hamiltonian cubic graph not isomorphic to H_1 or H_2. Then there exists a connected PDS of size $\lfloor \frac{2\cdot|V|+1}{3} \rfloor$ in G.*

Proof. If $|V| < 20$, then there is a PDS of size $\lfloor \frac{2\cdot|V|+1}{3} \rfloor$ in G from Lemma 13. Now we suppose that $|V| \geq 20$, which implies that $k := \lceil \frac{|V|-1}{3} \rceil \geq 7$.

From Lemma 11, $\gcd(k+1, |V|) \in \{1, 2, 3, 4\}$. If $\gcd(k+1, |V|) \in \{1, 3\}$, then there exists a good shift (Corollary 5).

We suppose that $\gcd(k+1, |V|) \in \{2, 4\}$. If G contains a good shift, then the proof is done. Notice that in such case, the PDS is obviously connected. Now, we assume that G has no good shift. We prove that given an almost good shift P, there exists a vertex $u^* \in P$ such that $G[P \setminus \{u^*\}]$ is a PDS. Observe that such vertex u^* exists if and only if $c(u^*-1), c(u^*+1) \in P$, and either $c(u^*) \in V \setminus P$ or $d_P(c(u^*)) = 3$.

- If G is of *type RLRL*, then $R = \langle 0\rangle \cup \langle 2\rangle$ and $L = \langle 1\rangle \cup \langle 3\rangle$. According to Lemma 12, the set $P := \{0, 1, 2, \dots, -k\}$ is an almost good shift and $0 \in R$, $1 \in L$. Since $2 \in R$ and $4 \in R$, then $c(2) \in P$ and $c(4) \in P$. If $c(3) \neq 0$, then $c(3) \in V \setminus P$ since $3 \in L$. Thus, $G[P \setminus \{3\}]$ is a PDS of size $\lfloor \frac{2\cdot|V|+1}{3} \rfloor$. Symmetrically, if $c(-k-3) \neq -k$, then $c(-k-3) \in V \setminus P$ since $3 \in R$. Thus, $G[P \setminus \{-k-3\}]$ is a PDS of size $\lfloor \frac{2\cdot|V|+1}{3} \rfloor$. On the other hand, if $c(3) = 0$ and $c(-k-3) = -k$, then $c(k-1) \neq -k$ and $c(k-1) \in P$. Moreover, since $k-3 \in R$ then $c(k-3) \in P$. Therefore, $c(k-2) \in V \setminus P$ or $d_P(c(k-2)) = 3$ (since $k \geq 7$, $k-2 \neq 3$ and $c(k-2) \neq 0$). Thus, $G[P \setminus \{k-2\}]$ is a PDS of size $\lfloor \frac{2\cdot|V|+1}{3} \rfloor$. Notice that the resulting PDS is connected. Indeed, let v be the vertex we removed from the path $\{0, 1, \dots, -k\}$. It is easy to see that, either $c(v-1) \in \{v+1, v+2, \dots, -k\}$, or $c(v+1) \in \{0, 1, \dots, v-1\}$ since the graph is of type *RLRL*, and thus the PDS is connected.
- If G is of *type RRLL*, then $R = \langle 0\rangle \cup \langle 1\rangle$ and $L = \langle 2\rangle \cup \langle 3\rangle$. According to Lemma 12, the set $P := \{1, 2, \dots, -k+1\}$ is an almost good shift and $1 \in R$, $2 \in L$, $-k \in R$, $-k+1 \in L$. Since $k+1 \in \langle 0\rangle$ and $k+2 \in \langle 1\rangle$, we necessarily have $k-1, k \in L$ and $k+1, k+2 \in R$. In this case, notice that since $k \geq 7$, $\{k-3, k-2, k-1\} \in P$. Moreover, $k-3, k-2 \in R$, which implies $c(k-3), c(k-2) \in P$. We show that either $c(k-1) \in P$ or $c(k) \in P$. Suppose that $c(k) \notin P$. Then since $k \in L$, we have $c(k) = 0$. Since $k-1 \in L$, we have $c(k-1) \in \{-1, 0, 1, \dots, k-3\}$. Since $0 = c(k)$ and $-1 \in L$, then $c(k-1) \neq -1$ and $c(k-1) \neq 0$. Thus, $c(k-1) \in \{1, 2, \dots, k-3\} \subset P$. Thus, either $c(k-1) \in P$ or $c(k) \in P$. Now, if $c(k-1) \in P$, then since $c(k-3) \in P$, the set $G[P \setminus \{k-2\}]$ is a PDS of size $\lfloor \frac{2\cdot|V|+1}{3} \rfloor$. Else, $c(k) \in P$ and then since $c(k-2) \in P$, the set $G[P \setminus \{k-1\}]$ is a PDS of size $\lfloor \frac{2\cdot|V|+1}{3} \rfloor$. Notice that the resulting PDS is connected. Indeed, let v be the vertex we removed from the almost good path $\{1, 2, \dots, -k+1\}$. Again, it is easy to verify that either $v = k-2$, and then $c(k-3) \in \{k-1, k, \dots, -k+1\}$, or $v = k-1$, and then $c(k) \in \{1, 2, \dots, k-2\}$ since the graph is of type *RRLL*. Thus the PDS is connected. $\square$

According to Lemma 8, a PDS in a cubic graph of order n contains at most $\lfloor \frac{2n+1}{3} \rfloor$ vertices. Thus, we obtain the following corollary.

Corollary 6. *Let G be a Hamiltonian cubic graph with a given Hamiltonian cycle. Then a connected proportional dense subgraph of maximum size in G can be found in linear time.*

6. Conclusion and open problems

We prove that Max Proportionally Dense Subgraph is APX-hard even on split graphs, and NP-hard on bipartite graphs, whether the PDS is required to be connected or not. Furthermore, the problem is proved to be $(2 - \frac{2}{\Delta+1})$-approximable, where Δ is the maximum degree of the graph. We also show that deciding if a PDS is inclusion-wise maximal is co-NP-complete, even on bipartite graphs. Nevertheless, Max PDS can be solved in linear time on Hamiltonian cubic graphs if a Hamiltonian cycle is given.

However, the complexity of finding a PDS of maximum size in cubic graphs remains unknown. More specifically, the question whether a PDS of size $\lfloor \frac{2n+1}{3} \rfloor$ always exists in a cubic graph is still open (except for the two graphs given in Fig. 4). Also, Algorithm 1 returns a PDS of size $\lceil \frac{n}{2} \rceil$ or $\lceil \frac{n}{2} \rceil + 1$ (in linear time), but the PDS may not be connected. An interesting open question is whether there is always a connected PDS of size at least $\lceil \frac{n}{2} \rceil$. Finally, the parameterized complexity of finding a PDS of size at least $\lceil \frac{n}{2} \rceil + k$ is unknown.

References

[1] R. Andersen, K. Chellapilla, Finding dense subgraphs with size bounds, in: 6th International Workshop on Algorithms and Models for the Web-Graph (WAW), in: LNCS, vol. 5427, Springer, 2009, pp. 25–37, http://dx.doi.org/10.1007/978-3-540-95995-3_3.

[2] R. Anderson, Finding large and small dense subgraphs, 2007, arXiv:cs/0702032.

[3] Y. Asahiro, R. Hassin, K. Iwama, Complexity of finding dense subgraphs, Discrete Appl. Math. 121 (1–3) (2002) 15–26, http://dx.doi.org/10.1016/S0166-218X(01)00243-8.

[4] C. Bazgan, J. Chlebíková, C. Dallard, Graphs without a partition into two proportionally dense subgraphs, 2018, arXiv:cs/1806.10963.

[5] C. Bazgan, J. Chlebíková, T. Pontoizeau, Structural and algorithmic properties of 2-community structures, Algorithmica 80 (6) (2018) 1890–1908, http://dx.doi.org/10.1007/s00453-017-0283-7.

[6] M. Chlebík, J. Chlebíková, Complexity of approximating bounded variants of optimization problems, Theoret. Comput. Sci. 354 (2006) 320–338, http://dx.doi.org/10.1016/j.tcs.2005.11.029.

[7] U. Feige, D. Peleg, G. Kortsarz, The dense k-subgraph problem, Algorithmica 29 (3) (2001) 410–421, http://dx.doi.org/10.1007/s004530010050.

[8] A.V. Goldberg, Finding a Maximum Density Subgraph, Technical Report UCB/CSD-84-171, EECS Department, University of California, Berkeley, 1984.

[9] R.M. Karp, Reducibility among combinatorial problems, in: Complexity of Computer Computations, Springer, 1972, pp. 85–103, http://dx.doi.org/10.1007/978-1-4684-2001-2_9.

[10] S. Khuller, B. Saha, On finding dense subgraphs, in: 36th International Colloquium on Automata, Languages and Programming (ICALP), in: LNCS, vol. 5555, Springer-Verlag, 2009, pp. 597–608, http://dx.doi.org/10.1007/978-3-642-02927-1_50.

[11] P. Kristiansen, S.M. Hedetniemi, S.T. Hedetniemi, Alliances in graphs, J. Combin. Math. Combin. Comput. 48 (2004) 157–177.

[12] M. Olsen, A general view on computing communities, Math. Social Sci. 66 (3) (2013) 331–336, http://dx.doi.org/10.1016/j.mathsocsci.2013.07.002.

[13] R.W. Robinson, N.C. Wormald, Almost all cubic graphs are Hamiltonian, Random Struct. Algorithms 3 (2) (1992) 117–125, http://dx.doi.org/10.1002/rsa.3240030202.

[14] J.A. Rodríguez-Velázquez, I.G. Yero, J.M. Sigarreta, Defensive k-alliances in graphs, Appl. Math. Lett. 22 (1) (2009) 96–100, http://dx.doi.org/10.1016/j.aml.2008.02.012.