Dense Graph Partitioning on Sparse and Dense Graphs

Cristina Bazgan ✉ 🏠 🆔
Université Paris-Dauphine, PSL Research University, CNRS, UMR 7243, LAMSADE,
75016 Paris, France

Katrin Casel ✉ 🏠 🆔
Hasso Plattner Institut, Universität Potsdam, Germany

Pierre Cazals ✉ 🆔
Université Paris-Dauphine, PSL Research University, CNRS, UMR 7243, LAMSADE,
75016 Paris, France

──── **Abstract** ────

We consider the problem of partitioning a graph into a non-fixed number of non-overlapping subgraphs of maximum density. The density of a partition is the sum of the densities of the subgraphs, where the density of a subgraph is half its average degree, that is, the ratio of its number of edges and its number of vertices. This problem, called Dense Graph Partition, is known to be NP-hard on general graphs and polynomial-time solvable on trees, and polynomial-time 2-approximable.

In this paper we study the restriction of Dense Graph Partition to particular sparse and dense graph classes. In particular, we prove that it is NP-hard on dense bipartite graphs as well as on cubic graphs. On dense graphs on n vertices, it is polynomial-time solvable on graphs with minimum degree $n - 3$ and NP-hard on $(n - 4)$-regular graphs. We prove that it is polynomial-time $4/3$-approximable on cubic graphs and admits an efficient polynomial-time approximation scheme on graphs of minimum degree $n - t$ for any constant $t \geq 4$.

1 Introduction

The research around communities in social networks can be seen as a contribution to the well establish research of clustering and graph partitioning. Graph partitioning problems have been intensively studied with various measures in order to evaluate clustering quality, see e.g. [17, 18, 10, 6] for an overview. In the context of social networks, a 'community' is a collection of individuals who are relatively well connected compared to other parts of the social network graph . A 'community structure' then corresponds to a partition of the whole social network into communities.

We consider a classical definition of the density of a (sub)graph (see, for example, [12, 15, 8]) given by half its average degree, that is, the ratio between its number of edges and its number of vertices. For this definition of density, there are several papers on finding the densest subgraph. This problem was shown solvable in polynomial time by Goldberg [12] but if the size of the subgraph is a part on the input, the problem called k-DENSEST SUBGRAPH becomes NP-hard even restricted to bipartite or chordal graphs [7]. The approximability of k-DENSEST SUBGRAPH was also studied, see [14, 9, 4].

18th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2022).
Editors: Artur Czumaj and Qin Xin; Article No. 13; pp. 13:1–13:15

Leibniz International Proceedings in Informatics
LIPIcs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

In this paper, we study the problem MAX DENSE GRAPH PARTITION that models finding a community structure, that is, finding a dense *partition*. More precisely, given an undirected graph G, we aim to find a partition $\mathcal{P} = \{V_1, \ldots, V_k\}$, $k \geq 1$, of the vertices of G, such that sum of the densities of the subgraphs $G[V_i]$ is maximized. We denote the sum of the densities of the subgraphs $G[V_i]$ by $d(\mathcal{P})$, and call this the density of the partition $\mathcal{P}$.

Note that the general concept of a community structure does not put any restriction on the number of communities. We therefore address the problem MAX DENSE GRAPH PARTITION of finding a partition of maximum density, without fixing the number of classes of the partition. Indeed, when the number of classes is given, the problem is a generalization of a partition into k cliques. By not fixing the number of classes, MAX DENSE GRAPH PARTITION differs from partitioning into cliques: observe that while there exists a partition into exactly k sets of density $(n - k)/2$ if and only if the input graph can be partitioned into k cliques (see Lemma 2), there can be a partition into less than k sets with a density even higher than $(n - k)/2$ even if the input cannot be partitioned into k cliques. As an example, consider a complete graph of an even number n of vertices and turn four of the vertices into an independent set by removing all edges among them. The resulting graph cannot be partitioned into 3 cliques (at least one set contains two of the four independent vertices), but it has a partition into two sets of equal cardinality with density $(n - 2)/2 - 4/n$.

Darley et al. [8] studied MAX DENSE GRAPH PARTITION, and its complement MIN SPARSE GRAPH PARTITION. They defined the sparsity of a partition $\mathcal{P}$ as $F(\mathcal{P}) = \frac{|\mathcal{P}|}{2} + d(\mathcal{P})$ and the problem MIN SPARSE GRAPH PARTITION as finding a partition of a given undirected graph G such that the sparsity of the partition is minimized. Observe that MAX DENSE GRAPH PARTITION and MIN SPARSE GRAPH PARTITION are dual in the sense that solving the first one on a graph G is the same as solving the second one on the complement of G. In [8] it is shown that both problems are NP-complete, and that there is no constant factor approximation for MIN SPARSE GRAPH PARTITION unless $P = NP$. Moreover, a polynomial time algorithm for MAX DENSE GRAPH PARTITION on trees is given. We point out that their proof of NP-completeness is a polynomial-time reduction from k-COLORING. By construction, the same reduction when starting from 3-COLORING on graphs of degree at most 4 (proved NP-complete in [11]) yields as instance of MAX DENSE GRAPH PARTITION a graph on n vertices and of minimum degree greater than $n - 4n^{4/5}$. Thus it follows that MAX DENSE GRAPH PARTITION is NP-complete restricted to graphs of minimum degree $n - 4n^{4/5}$.

Aziz et al. [2] studied the problem FRACTIONAL HEDONIC GAME, and more particularly the MAX UTILITARIAN WELFARE problem as the simple symmetric version of the game defined as follows. Let N be a set of agents, the utility of $i \in N$ in a coalition $S \subseteq N$ is $u_i(S) = \frac{1}{|S|}\sum_{j \in S} u_i(j)$ where $u_i(j)$ is such that $u_i(j) \in \{0, 1\}$ for a simple game and $u_i(j) = u_j(i)$ for a symmetric one. For MAX UTILITARIAN WELFARE one tries to find a partition C of N into coalitions that maximizes $\sum_{S \in C} \sum_{i \in S} u_i(S)$. This game can be seen as a graph G where agents are vertices and there is an edge between two agents i and j if and only if $u_i(j) = 1$. In this context, $u_i(S) = \frac{1}{|S|}\sum_{j \in S} u_i(j) = \frac{1}{|S|} deg_{G[S]}(i)$. We deduce that $\sum_{S \in C} \sum_{i \in S} u_i(S) = \frac{1}{|S|} \sum_{S \in C} \sum_{i \in S} deg_{G[S]}(i) = \frac{1}{|S|} \sum_{S \in C} 2|E(S)| = 2 \cdot d(C)$. Hence, the problems MAX UTILITARIAN WELFARE and MAX DENSE GRAPH PARTITION are equivalent to within a constant, which means that the 2-approximation for the former given in [2] directly translates to the latter.

Our contributions.   The following overview summarises the results achieved in this paper concerning MAX DENSE GRAPH PARTITION (MDGP).

- MDGP is trivially solvable on graphs of maximum degree 2, we prove its NP-hardness for 3-regular (cubic) graphs.

- We establish that on bipartite complete graphs an optimal partition consists of one part, that is the whole graph. Moreover if the size of the two independent sets are relatively prime numbers then this optimal solution is unique. We use this result to show that MDGP is NP-hard on dense bipartite graphs.
- MDGP is trivial on complete graphs since the optimal solution is the whole graph as one part of the partition. Moreover, as we previously explained, it is NP-hard on graphs of minimum degree $n - 4n^{4/5}$. We show that for graphs of minimum degree $\geq n - 3$, the problem is solvable in polynomial time and any optimal solution has two parts. Moreover on $(n - 4)$-regular graphs, the problem becomes NP-hard.
- We further improve on the 2-approximation for MDGP [2] for specific sparse and dense graph classes. In particular, we show that MDGP admits a polynomial-time 4/3-approximation on cubic graphs. Moreover we establish a polynomial-time $\frac{n-1}{\delta+1}$-approximation, where δ is the minimum degree of the input graph (note that this improves on the ratio of 2 for all $\delta > \frac{n-3}{2}$). Also, we give an eptas (i.e. a $(1 + \varepsilon)$-approximation for any $\varepsilon > 0$) on graphs of minimum degree $n - t$ for any constant $t \geq 4$

Our paper is organized as follows. Notations and formal definitions are given in Section 2. The study of (dense) bipartite graphs is established in Section 3. Section 4 presents the results on cubic graphs. In Section 5 we study dense graphs. Some conclusions are given at the end of the paper.

2 Preliminaries

In this paper we assume that all graphs are undirected, without loops or multiple edges, and not necessary connected. We use $G = (V, E)$ to denote an undirected graph with a set V of vertices and a set E of edges. We use $|V|$ to denote the number of vertices in G, i.e., the order of G, and we use $|E|$ to denote the number of edges in G, i.e., the size of G. We denote by $deg_G(v)$ the degree of $v \in V$ in G that is the number of edges incident to v and by $D_G(i)$ the set of vertices of degree i in G. The maximum degree of G, denoted by $\Delta(G)$, is the degree of the vertex with the greatest number of edges incident to it. The minimum degree of G, denoted by $\delta(G)$, is the degree of the vertex with the least number of edges incident to it. For any vertex $v \in V$, $N_G(v)$ is the set of neighbors of v in G and $N_G[v] = N_G(v) \cup \{v\}$. Moreover, $N_G(S) = \bigcup_{v \in S} N_G(v)$. For a graph $G = (V, E)$ and a subset $S \subseteq V$ we denote by $E(S)$ the set of the edges of G with both endpoints in S. For a given partition $\{A, B\}$ of V, we denote by $E(A, B) = \{uv \in E : u \in A, v \in B\}$. Further, $G[S]$ denotes the graph induced by S, defined as $G[S] = (S, E(S))$.

A triangle graph is the cycle graph C_3 or the complete graph K_3. A diamond graph has 4 vertices and 5 edges, it consists of a complete graph K_4 minus one edge. A graph is called cubic if all its vertices are of degree three. A graph is bipartite if its vertices can be partitioned into two sets A and B such that every edge connects a vertex in A to one in B. A complete bipartite graph is a special kind of bipartite graph where every vertex of A is connected to every vertex in B. A graph on n vertices is δ-dense if its minimum degree is at least δn. A set of instances is called dense if there is a constant $\delta > 0$ such that all instances in this set are δ-dense (this notion was introduced in [1] and called everywhere-dense).

The density $d(G)$ of a graph $G = (V, E)$ is the ratio between the number of edges and the number of vertices in G, that is, $d(G) = \frac{|E|}{|V|}$. Moreover, for $S \subseteq V$, $d(S) = d(G[S]) = \frac{|E(S)|}{|S|}$. We use $\mathcal{P}$ to denote a partition of the set V of vertices of G, that is, $\mathcal{P} = \{V_1, \ldots, V_k\}$, where $\cup_{i=1}^{k} V_i = V$, and $V_i \cap V_j = \emptyset$ for each $i, j \in \{1, \ldots, k\}$. Then the density of a partition $\mathcal{P}$ of G is defined as $d(\mathcal{P}) = \sum_{i=1}^{k} d(G[V_i])$, where $G[V_i]$ is the subgraph of G induced by the subset V_i of vertices, that is, $G[V_i] = (V_i, E_i)$, $E_i = \{\{u, v\} : \{u, v\} \in E \wedge u, v \in V_i\}$.

We study the problem of finding a partition $\mathcal{P} = \{V_1, \ldots, V_k\}$ of a given graph G, such that $k \geq 1$ and that, among all such partitions, $d(\mathcal{P})$ is maximized. We refer to this problem as MAX DENSE GRAPH PARTITION and we define its decision version as follows.

DENSE GRAPH PARTITION
Input: An undirected graph $G = (V, E)$, a positive rational number r.
Question: Is there a partition $\mathcal{P}$ such that $d(\mathcal{P}) \geq r$?

Given an optimization problem in NPO and an instance I of this problem, we denote by $|I|$ the size of I, by $opt(I)$ the optimum value of I, and by $val(I, S)$ the value of a feasible solution S of instance I. The performance ratio of S (or approximation factor) is $r(I, S) = \max\{\frac{val(I,S)}{opt(I)}, \frac{opt(I)}{val(I,S)}\} \geq 1$. For a function f, an algorithm is an $f(|I|)$-approximation, if for every instance I of the problem, it returns a solution S such that $r(I, S) \leq f(|I|)$. Moreover if the algorithm runs in polynomial time in $|I|$, then this algorithm gives a polynomial-time $f(|I|)$-approximation. We consider in this paper only polynomial time algorithms. When f is a constant α, the problem is polynomial-time α-approximable. When $f = 1 + \varepsilon$, for any $\varepsilon > 0$, the problem admits a polynomial-time approximation scheme. When the running time of an approximation scheme is of the form $O(g(1/\varepsilon)poly(|I|))$ the problem has an efficient polynomial-time approximation scheme (eptas).

Before we start studying specific graph classes, we observe the following helpful structural properties that hold for DENSE GRAPH PARTITION on general graphs.

▶ **Remark 1.** We can assume that for any optimal partition $\mathcal{P}$ and for any part $P \in \mathcal{P}$, $G[P]$ is connected, since otherwise turning each connected component into its own part does not decrease the density.

When discussing the density of a (sub)graph, it is often useful to think about how close this subgraph is to being a clique. We therefore call a pair of non-adjacent vertices in a (sub)graph a *missing edge*, and use the number of such missing edges to estimate the density of the (sub)graph. With such estimations, it is easy to show that the following intuition about favouring complete graphs as communities.

▶ **Lemma 2.** *Among all partitions of G into $t \geq 2$ parts, those where the parts correspond to complete graphs, if there exists such, have the largest density.*

Proof. Consider a partition of G into t parts $\{V_1, \ldots, V_t\}$ of size $n_1, \ldots, n_t$. If $G[V_i]$ has o_i missing edges for any $1 \leq i \leq t$, then the density of this partition is $\frac{n-t}{2} - \frac{o_1}{n_1} - \ldots - \frac{o_t}{n_t}$.

Consider a partition of G into t parts of size $n'_1, \ldots, n'_t$ such that each part induces a complete graph for any $1 \leq i \leq t$. Then the density of this partition is $\frac{n-t}{2}$ and thus it is larger than the density of any partition in t parts where at least one edge is missing inside $G[V_i]$ for some $1 \leq i \leq t$.   ◀

A direct consequence of this is the following.

▶ **Lemma 3.** *Let $G = (V, E)$ be a graph and $\mathcal{P}$ be any partition of V. Then $d(\mathcal{P}) \leq \frac{|V|}{2} - \frac{|\mathcal{P}|}{2}$.*

3   Dense Bipartite Graphs

In this section we show that MAX DENSE GRAPH PARTITION has a trivial solution on complete bipartite graphs. Moreover, using this result we show that the problem is NP-hard on dense bipartite graphs.

Considering all possibilities to partition a complete bipartite graph $G_{n,m}$ with the two subsets that are independent sets of size n and m one can derive the following.
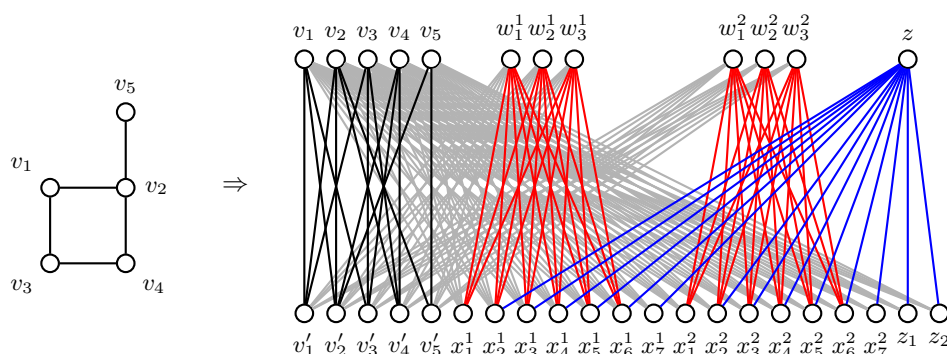
Figure 1 A graph G, instance of Dominating Set and the bipartite graph G' obtained from G, for $k = 2$ and $n = 5$.

▶ **Lemma 4.** *The density $d(G_{n,m})$ of a complete bipartite graph $G_{n,m}$ is greater than or equal to the density $d(\mathcal{P})$ of any partition $\mathcal{P}$ of $G_{n,m}$.*

It follows that an optimal solution of any complete bipartite graph is the whole graph. The calculations used to prove Lemma 4 inductively also give the following.

▶ **Corollary 5.** *For any complete bipartite graph $G = (A, B, E)$ with $|A| = n$ and $|B| = m$, a partition $\mathcal{P} = \{V_1, \ldots, V_k\}$ of $A \cup B$ satisfies $d(\mathcal{P}) = \frac{nm}{n+m}$ if and only if $G[V_i] = G_{n_i,m_i}$ with $n_i \neq 0$ and $m_i \neq 0$ and $\frac{n_i}{m_i} = \frac{n}{m}$ for all $i \in \{1, \ldots, k\}$.*

Consequently, for any complete bipartite graph $G_{n,m}$, if n and m are relatively prime the only optimal solution of $G_{n,m}$ is the whole graph. Otherwise, several optimal solutions exist and are characterized exactly by Corollary 5.

▶ **Theorem 6.** *Dense Graph Partition is NP-hard on dense bipartite graphs.*

Proof. We give a reduction from Dominating Set. Let $G = (V, E)$ with $V = \{v_1, \ldots, v_n\}$ and an integer $k \geq 1$ be an instance of Dominating Set. Assume without loss of generality that G is connected. We first construct a bipartite graph $G' = (V_1, V_2, E')$, that is not dense, and show how solving Dense Graph Partition on it solves Dominating Set on G. In a second step, we show how to make G' dense maintaining the reduction.

We construct $G' = (V_1, V_2, E')$ as follows:

▬ $V_1 = V \cup \{w_i^j : 1 \leq i \leq n - k, \ 1 \leq j \leq k\} \cup \{z\}$

▬ $V_2 = V' \cup \{x_r^j : 1 \leq r \leq N, \ 1 \leq j \leq k\} \cup \{z_i : 1 \leq i \leq N - n\}$ where $V' = \{v_1', \ldots, v_n'\}$ and $N \in \mathbb{N}$ is chosen as follows. Let $c \in \mathbb{N}$ be the smallest integer such that $c(n - k + 1) - 1 > n$ (note that $1 \leq c \leq n$) and define $N = c(n - k + 1) - 1$. For this choice of N it follows that the greatest common divisor of N and $n - k + 1$ is 1, and $n < N \leq 2n$.

▬ $E' = E_d \cup E_{wx} \cup E_c \cup E_z$ with

$E_d = \{\{v_i, v_j'\} : \{v_i, v_j\} \in E\} \cup \{\{v_i, v_i'\} : 1 \leq i \leq n\}$,

$E_{wx} = \{\{w_i^j, x_r^j\} : 1 \leq i \leq n - k, \ 1 \leq r \leq N - 1, 1 \leq j \leq k\}$,

$E_c = \{\{w_i^j, v_s'\} : 1 \leq i \leq n - k, 1 \leq j \leq k, \ 1 \leq s \leq n\} \cup \{\{v_s, x_r^j\} : 1 \leq s \leq n, \ 1 \leq r \leq N, \ 1 \leq j \leq k\}$ and

$E_z = \{\{z, z_j\} : 1 \leq j \leq N - n\} \cup \{\{z, x_r^j\} : 2 \leq r \leq N, 1 \leq j \leq k\} \cup \{\{v_i, z_j\} : 1 \leq i \leq n, \ 1 \leq j \leq N - n\}$

Notice that G' is a bipartite graph with $|V_1| = n + 1 + k(n - k)$ and $|V_2| = (k + 1)N$.

We show that there exits a dominating set of cardinality at most k in G if and only if there exists a partition $\mathcal{P}$ of G' with $d(\mathcal{P}) = (k + 1)d(G_{n-k+1,N})$.

Suppose there exists a dominating set D in G with $|D| = k$. Let $D = \{v_{i_1}, \ldots, v_{i_k}\}$ and $N'(v_{i_j}) = N_G[v_{i_j}] \setminus (D \cup N_G(\{v_{i_1}, \ldots, v_{i_{j-1}}\}))$. Define the partition $\mathcal{P} = \{P_1, \ldots, P_{k+1}\}$ by: $P_j = \{v_{i_j}\} \cup \{v'_r \colon v_r \in N'(v_{i_j})\} \cup \{w^j_r \colon 1 \le r \le n - k\} \cup \{x^j_r \colon 1 \le r \le N - |N'(v_{i_j})|\}$ for $1 \le j \le k$ and $P_{k+1} = V_1 \cup V_2 \setminus (\cup_{j=1}^k P_j)$. With this definition, $\mathcal{P}$ is clearly a partition of $V_1 \cup V_2$, and each part P_j contains $n - k + 1$ vertices from V_1 and N vertices from V_2 for each $1 \le j \le k + 1$. Further, each P_j induces a complete bipartite graph $G_{n-k+1,N}$: All vertices w^j_r and x^j_r are connected to each other, and to all vertices in V_2 and V_1, respectively, by construction. Further, v_{i_j} is connected in G' to all vertices in $N'(v_{i_j})$; note here that in G' we connected v_i to its "copy" v'_i for all $1 \le i \le n$, which models the case that v_{i_j} dominates itself. For P_{k+1}, note that z is adjacent to all x^j_i-vertices, and each z_i is adjacent to all vertices in V. Since D is a dominating set, each vertex from V' is contained in some $N'(v_{i_j})$, thus $V_2 \setminus (\cup_{j=1}^k P_j)$ only contains x^j_i-vertices. Also, the P_j contain all w^j_i vertices and hence $V_1 \setminus (\cup_{j=1}^k P_j)$ only contains vertices from V.

Conversely, let $\mathcal{P}$ be a partition of G' of density $(k+1)d(G_{n-k+1,N})$. Thus, Corollary 5 implies that the vertices for each set $P \in \mathcal{P}$ induce a complete bipartite graph $G_{r,s}$ such that $\frac{r}{s} = \frac{|V_1|}{|V_2|} = \frac{k(n-k)+n+1}{(k+1)N} = \frac{n-k+1}{N}$. Since the greatest common divisor of $n - k + 1$ and N is one, this yields $r \ge n - k + 1$ and $s \ge N$ and especially $\mathcal{P}$ can contain at most $k + 1$ sets.

For all w^j_i and w^t_ℓ, if $j \ne t$, w^j_i and w^t_ℓ have n common neighbors, and since $n < N$ there is no part $P \in \mathcal{P}$ such that $w^j_i, w^t_\ell \in P$. Moreover, for all i, j, w^j_i and z have $N - 1$ common neighbors so they also cannot be in the same $P \in \mathcal{P}$. Hence, there are exactly $k + 1$ parts in $\mathcal{P}$ that are complete bipartite graphs $G_{n-k+1,N}$.

For all $1 \le j \le k$, denote by P_j the set containing the vertices w^j_i for all $1 \le i \le n - k$ and P_z the set containing z. To reach cardinality exactly $n - k + 1$, $P_j \cap V_1$ has to contain exactly one vertex from V for each $1 \le j \le k$. Further, since for any i, v'_i is not adjacent to z, $V' \subseteq \cup_{j=1}^k P_j$. As each $P \in \mathcal{P}$ induces a complete bipartite graph in G', $D = V \cap \cup_{j=1}^k P_j$ is a set of size k, such that each vertex in V' is adjacent to at least one vertex in D, so we deduce that D is a dominating set of size k in G.

We extend the construction of the proof to create from G' a dense bipartite graph $G'' = (V'', E'')$ by adding four sets of vertices $V^u_1, V^d_1, V^u_2, V^d_2$ with $|V^u_1| = |V^d_1| = kn|V_1| = kn(k(n-k)+n+1)$ and $|V^u_2| = |V^d_2| = kn|V_2| = knN(k+1)$. Further, we add edges to turn the pairs (V^u_1, V^u_2), (V^d_1, V^d_2), (V^u_1, V_1), and (V^d_2, V_2) each into complete bipartite graphs. Observe that with this construction G'' has $|V''| = (2kn+1)(k(n-k)+n+1) + (2kn+1)N(k+1) < 10k^2n^2$ vertices and that all vertices have degree at least $kn|V_1| \ge \frac{1}{2}k^2n^2 \in \Theta(|V''|)$. (Note that if $k \ge \frac{n}{2}$, G is a trivial yes-instance for DOMINATING SET.)

We claim that there exists a partition $\mathcal{P}'$ of G'' with $d(\mathcal{P}') = (k+1)d(G_{n-k+1,N}) + 2kn(k+1)d(G_{n-k+1,N})$ if and only if there exists a dominating set of size k for G. Corollary 5 again implies that this density for G'' can only be achieved by a partition into complete bipartite graphs $G_{r,s}$ with $\frac{r}{s} = \frac{(2kn+1)(k(n-k)+n+1)}{(2kn+1)N(k+1)} = \frac{n-k+1}{N}$. The vertices in V^d_1 are only adjacent to vertices in V^d_2, and the vertices in V^u_2 are only adjacent to vertices in V^u_1. Clustering these in a ratio $\frac{r}{s}$ results in clusters containing exactly all newly added vertices, and this can be done with just two sets in total. What remains is to cluster the graph G' into complete bipartite graphs $G_{r,s}$ such that $\frac{r}{s} = \frac{|V_1|}{|V_2|} = \frac{k(n-k)+n+1}{(k+1)N} = \frac{n-k+1}{N}$ as before.   ◄

4   Cubic Graphs

In this section, we study DENSE GRAPH PARTITION on cubic graphs, show that it remains NP-complete on this restricted graph class, but also give a polynomial time $\frac{4}{3}$-approximation for its optimization variant MAX DENSE GRAPH PARTITION. We start with some general observations on the structure of communities in cubic graphs.

▶ **Definition 7.** *For $P \subseteq V$, the utility of a vertex $v \in P$ is defined by $u_P(v) = \frac{d(P)}{|P|}$, and the utility of P is defined by $u(P) = u_P(v)$ for any $v \in P$. For a partition $\mathcal{P} = \{V_1, \ldots, V_k\}$, the utility of a vertex v in $\mathcal{P}$ is defined by $u_{\mathcal{P}}(v) = u_{V_i}(v)$ with i such that $v \in V_i$.*

Considering these definitions, we can remark that:

- For any subset $P \subseteq V$, and $v, w \in P$, $u_P(v) = u_P(w)$.
- If $P = \{v\}$ then $u_P(v) = 0$.
- For any partition $\mathcal{P}$ of G, $\sum\limits_{V_i \in \mathcal{P}} d(V_i) = \sum\limits_{v \in V} u_{\mathcal{P}}(v)$.

▶ **Lemma 8.** *Let $G = (V, E)$ be a cubic graph without connected components that induce a K_4. For any partition $\mathcal{P}$ of G the following holds:*

- *$u_{\mathcal{P}}(v) \leq \frac{1}{3}$ for all vertices $v \in V$*
- *if $P \in \mathcal{P}$ is not a triangle, diamond or Case 1 in Figure 2 then $u(P) \leq \frac{1}{4}$*

Proof. Let $\mathcal{P}$ be a partition of G, $P \in \mathcal{P}$ and $v \in P$. Since G is cubic, $d(P) \leq \frac{3|P|}{2|P|} = \frac{3}{2}$. Then $u_{\mathcal{P}}(v) \leq \frac{3}{2|P|}$. If $|P| \geq 6$, $u_{\mathcal{P}}(v) \leq \frac{3}{2 \cdot 6} = \frac{1}{4}$. For $|P| = 5$ it follows that $u_{\mathcal{P}}(v) \leq \frac{7}{25} < \frac{1}{3}$, since a cubic graph on 5 vertices cannot have more than 7 edges. Also, since there exists no K_4 in G, the only graph on 5 vertices with 7 edges is Case 1 in Figure 2, and all other graphs on 5 vertices have 6 or less edges which yields a utility of at most $\frac{6}{25} < \frac{1}{4}$.

Case analysis on the graphs of size 4 or less yields that the largest utility is achieved for P being a triangle, which gives $u_{\mathcal{P}}(v) = \frac{1}{3}$. Further, if P is not a triangle or a diamond, case analysis on the graphs of size 4 or less shows that $u_{\mathcal{P}}(v)$ is maximized when P is an induced matching and its value is $\frac{1}{4}$. ◀

Considering all cases to distribute vertices of a diamond, as illustrated in Figure 2 can be used to show the following.

▶ **Lemma 9.** *Let G be a cubic graph without connected components that induce a K_4, and let v_1, v_2, v_3, v_4 be vertices in G that induce a diamond. Then $u_{\mathcal{P}}(v_1) + u_{\mathcal{P}}(v_2) + u_{\mathcal{P}}(v_3) + u_{\mathcal{P}}(v_4) \leq \frac{5}{4}$ for any partition $\mathcal{P}$ for G.*

▶ **Lemma 10.** *Let G be a cubic graph on n vertices without connected components that induce a K_4, and let D be the set of diamonds in G and T the set of triangles in G that do not belong to a diamond. For any partition $\mathcal{P}$, $d(\mathcal{P}) \leq \frac{5}{4}|D| + |T| + \frac{1}{4}(n - 3|T| - 4|D|)$.*

Proof. By Lemma 8, the only vertices with utility more than $\frac{1}{4}$ are those that are in triangles, diamonds, or the unique neighbors of diamonds (in the sense of vertex v_5 in Case 1 of Figure 2), and we know that the sum of the utilities of the vertices constituting a triangle

Case 1   Case 2   Case 3
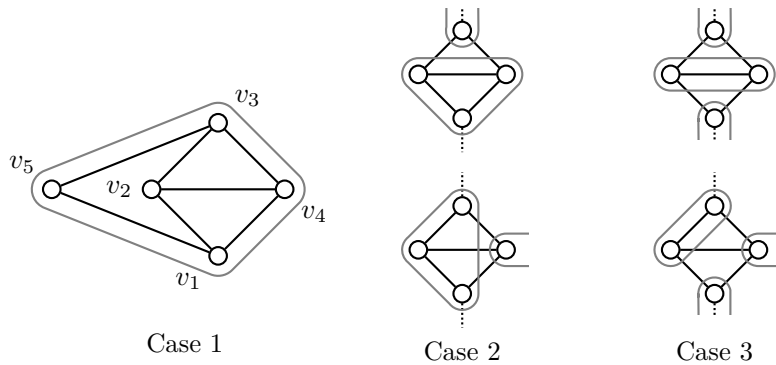
Figure 2 Different cases of Lemma 9.
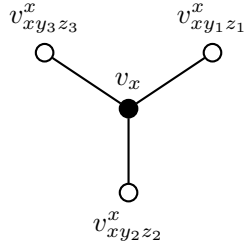
Figure 3 Subgraph containing one vertex of type 1, v_x, and its neighbors in G.
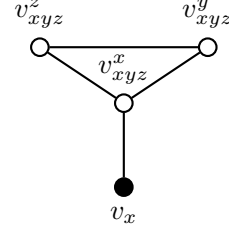
Figure 4 Subgraph containing one vertex of type 1, v_x, and three of type 2.

is at most $3 \cdot \frac{1}{3} = 1$. By Lemma 9, we further know that the sum of the utilities of the vertices constituting a diamond is at most $\frac{5}{4}$. The unique neighbors of diamonds have a utility of more than $\frac{1}{4}$ if and only if they are in a part isomorphic to Case 1 of Figure 2, which has a density of $\frac{7}{5} < \frac{5}{4} + \frac{1}{4}$. Thus, if S is the set of unique neighbors of diamonds, then the sum of the utilities of the vertices in the diamonds in D and the vertices in S is at most $\frac{5}{4}|D| + \frac{1}{4}|S|$. All remaining vertices have a utility of at most $\frac{1}{4}$ by Lemma 8. We deduce that $d(G) \leq \frac{5}{4}|D| + \frac{1}{4}|S| + |T| + \frac{1}{4}(n - 3|T| - 4|D| - |S|) = \frac{5}{4}|D| + |T| + \frac{1}{4}(n - 3|T| - 4|D|)$. ◄

We will reduce from the following problem, known to be NP-hard by [13]:

RESTRICTED EXACT COVER BY 3-SETS (RX3C)
Input: A set X of elements with $|X| = 3q$ and a collection C of 3-element subsets of X where each element appears in exactly 3 sets.
Question: Does C contain an exact cover for X, i.e. a subcollection $C' \subseteq C$ such that every element occurs in exactly one member of C' ?

The following definition gives the construction to reduce RX3C to DENSE GRAPH PARTITION.

▶ **Definition 11.** *Let $I = (X, C)$ be an instance of RX3C. We define the construction σ transforming I into the graph $G = (V, E) := \sigma(I)$ as follows (see Figures 3 and 4):*
- *for each element $x \in X$, add the vertex v_x to V (called vertices of type 1 or black vertices).*
- *for each subset of the collection $\{x, y, z\} \in C$, add the vertices v_{xyz}^x, v_{xyz}^y, v_{xyz}^z to V (called vertices of type 2 or white vertices).*
- *add the edges $\{v_{xyz}^x, v_{xyz}^y\}$, $\{v_{xyz}^x, v_{xyz}^z\}$ and $\{v_{xyz}^y, v_{xyz}^z\}$ to E*
- *add the edges $\{v_{xyz}^x, v_x\}$, $\{v_{xyz}^y, v_y\}$ and $\{v_{xyz}^z, v_z\}$ to E*

Notice that G is a cubic graph on $|X|$ vertices of type 1 and $3|X|$ vertices of type 2.

Case distinction on the subgraphs in $\sigma(I)$ can be used to show:

▶ **Lemma 12.** *For $G = (V, E) = \sigma(I)$ and any $P \subseteq V$, it holds that $u(P) \geq \frac{1}{4}$ if and only if $G[P]$ is isomorphic to one of the following three graphs:*
- *a triangle where all the vertices are of type 2 and then $u(P) = \frac{1}{3}$.*
- *an edge with at least one vertex of type 2, and then $u(P) = \frac{1}{4}$.*
- *the subgraph described in Figure 4 and then $u(P) = \frac{1}{4}$.*

▶ Remark 13. The case-analysis in the proof of Lemma 12 also shows that for any subset $P \subseteq V$ of the vertices of the graph $\sigma(I)$, if v is of type 2 then $u_S(v) \leq \frac{1}{3}$, otherwise $u_S(v) \leq \frac{1}{4}$.

▶ **Theorem 14.** *DENSE GRAPH PARTITION is NP-complete on cubic graphs.*

Proof. Let $I = (X, C)$ be an RX3C instance. We claim that I is a yes-instance of RX3C if and only if $I' = (G, \frac{7|X|}{6})$ with $G = \sigma(I)$ is a yes-instance of DENSE GRAPH PARTITION.

Let $C' \subseteq C$ be an exact cover for X of size $\frac{|X|}{3}$. Consider the partition $\mathcal{P}$ built from the following $\frac{5|X|}{3}$ parts: for any $c \in C'$, $c = \{x, y, z\}$ add $\{v_x, v_{xyz}^x\}$, $\{v_y, v_{xyz}^y\}$, $\{v_z, v_{xyz}^z\}$ and for any $c \notin C'$, $c = \{x, y, z\}$ add $\{v_{xyz}^x, v_{xyz}^y, v_{xyz}^z\}$. Since C' is an exact cover, $\mathcal{P}$ is a partition for G and its density is $\frac{3}{2} \cdot \frac{|X|}{3} + \frac{2}{3}|X| = \frac{7}{6}|X|$.

Let $\mathcal{P}'$ be a partition of G of density $d(\mathcal{P}') = \frac{7}{6}|X|$. Firstly, we show that $\mathcal{P}'$ has necessarily the following shape: $\frac{2|X|}{3}$ parts of size 3 containing only vertices of type 2 forming a triangle and $|X|$ parts of size 2 containing one vertex of type 1 and one of type 2 adjacent in G. From Remark 1, we can assume that all parts induce connected subgraphs.

We first show that $d(\mathcal{P}') = \frac{7|X|}{6}$ implies that there are at least $\frac{2|X|}{3}$ parts in $\mathcal{P}'$ which are triangles. Assume by contradiction that $\mathcal{P}'$ has $\frac{2|X|}{3} - \ell$ triangles, with $\ell > 0$. Since G has $4|X|$ vertices, there are $2|X| + 3\ell$ vertices that do not belong to a part in $\mathcal{P}'$ that corresponds to a triangle. By Lemma 12 the utility of these vertices is at most $\frac{1}{4}$ which yields:

$$d(\mathcal{P}') \leq \frac{2|X|}{3} - \ell + (2|X| + 3\ell) \cdot \frac{1}{4} = \frac{7|X|}{6} - \frac{\ell}{4} < \frac{7|X|}{6} .$$

This contradicts the choice of $\mathcal{P}'$ such that $d(\mathcal{P}') = \frac{7|X|}{6}$, hence $\mathcal{P}'$ has at least $\frac{2|X|}{3}$ triangles.

Assume by contradiction that $\mathcal{P}'$ has $\frac{2|X|}{3} + \ell$ triangles, with $\ell > 0$. Since there are $3|X|$ vertices of type 2 and among these vertices $3 \cdot (\frac{2|X|}{3} + \ell)$ belong to a triangle, $|X| - 3\ell$ vertices of type 2 do not belong to a triangle. Each neighbor of a vertex v_x of type 1 is of type 2, so if the utility of v_x is positive, there exists a vertex of type 2, v_{xyz}^x, neighbor of v_x, that is in the same part as v_x and v_{xyz}^x does not belong to a triangle. Moreover, as all type 1 vertices have no common neighbors, for each type 1 vertex with positive utility, there is a type 2 vertex that is not in a triangle. Since there are at most $|X| - 3\ell$ type 2 vertices that do not belong to a triangle, there are at most $|X| - 3\ell$ type 1 vertices with positive utility, thus:

$$d(\mathcal{P}') \leq \frac{2|X|}{3} + \ell + \frac{|X| - 3\ell}{4} + \frac{|X| - 3\ell}{4} \leq \frac{7|X|}{6} - \frac{\ell}{2} < \frac{7|X|}{6} .$$

This again contradicts the choice of $\mathcal{P}'$ with $d(\mathcal{P}') = \frac{7|X|}{6}$, thus $\mathcal{P}'$ has exactly $\frac{2|X|}{3}$ triangles.

We claim that $d(\mathcal{P}') = \frac{7|X|}{6}$ implies that all type 1 vertices are in a part that is a matching with a type 2 vertex. There are $|X|$ type 1 vertices and $|X|$ type 2 vertices that are not in some triangle in $\mathcal{P}'$. Since there are exactly $\frac{2|X|}{3}$ parts in $\mathcal{P}'$ forming a triangle and the utility of each other vertex is smaller than or equal to $\frac{1}{4}$, to reach a density of $\frac{7|X|}{6}$ it is necessary that each of the $2|X|$ vertices outside the parts that are triangles has a utility of exactly $\frac{1}{4}$. To reach this utility, by Lemma 12 there are two possibilities, the graph described in Figure 4 and an edge. Since there are exactly $|X|$ vertices of type 1 and $|X|$ vertices of type 2 outside the triangles in $\mathcal{P}'$, and vertices of type 1 only have neighbors of type 2, the only possibility is if each type 1 vertex is matched with one type 2 vertex.

Consider now the following subcollection $C'' \subseteq C$: for each triple $v_{xyz}^x, v_{xyz}^y, v_{xyz}^z$ that does not belong to a triangle, we add the set $\{x, y, z\}$ to C''. The subcollection C'' is a cover since each type 1 vertex is a neighbor of one of these vertices and it is an exact cover since there are exactly $\frac{|X|}{3}$ 3-element subsets that do not belong to a triangle. ◀

Our observations about the maximum utility of certain vertices can also be used to show the following positive result.

▶ **Theorem 15.** *MAX DENSE GRAPH PARTITION is polynomial-time $\frac{4}{3}$-approximable on cubic graphs.*

Proof. Let $I = G$ be a cubic graph, instance of MAX DENSE GRAPH PARTITION. If G contains connected components isomorphic to K_4, create a part for each such component, as this is the optimum way to partition these sets. Then, let D be the set of all diamonds in G, and T the set of all triangles that do not belong to a diamond. Diamonds (resp. triangles) can be found in polynomial time simply by enumerating all 4-tuples (resp. 3-tuples) of vertices and checking if they induce a diamond (resp. triangle) as subgraph. Let G' be the graph obtained from G after removing the vertices of D and T. Compute a maximum matching M of G', and let G'' be the graph obtained from G' after removing the vertices of M. Since M is a maximal matching, the vertices in G'' form an independent set.

We show in the following that $|V(G'')| \leq \frac{|V(G)|}{4}$. For each $v \in V$ we associate a function $t(v)$ and initialize it with $t(v) = 1$. When removing the diamonds and triangles from G in order to get G' we update the function t as follows:

- For every diamond $\{u_1, u_2, u_3, u_4\} \subseteq V$ that is deleted from V, let u_1 and u_3 be the vertices with neighbors outside of the diamond (if these vertices still exist) and let v_1 and v_3 be these neighbors (with the possibility that $v_1 = v_3$). We update the function t : $t(v_1) := t(v_1) + t(u_1) + t(u_2)$ and $t(v_3) := t(v_3) + t(u_3) + t(u_4)$ (thus $t(v_1) := t(v_1) + t(u_1) + t(u_2) + t(u_3) + t(u_4)$ if $v_1 = v_3$). If v_1 or v_3 were already deleted, we delete their associated t function.
- For every triangle $\{u_1, u_2, u_3\} \subseteq V$ that is deleted from V, let v_1 (resp. v_2 and v_3) be the neighbor of u_1 (resp. u_2 and u_3) outside of the triangle (if these vertices exist). We update the function t : $t(v_1) := t(v_1) + t(u_1)$, $t(v_2) := t(v_2) + t(u_2)$ and $t(v_3) = t(v_3) + t(u_3)$. If v_1, v_2 or v_3 do not exist, we delete their associated t function.

Observe that after updating t for any $v \in V(G')$, if $v \in D_{G'}(3)$ then $t(v) \geq 1$, if $v \in D_{G'}(2)$ then $t(v) \geq 2$, if $v \in D_{G'}(1)$ then $t(v) \geq 3$ and if $v \in D_{G'}(0)$ then $t(v) \geq 4$. In order to justify this, observe that the t function associated to vertices in $V(G')$ cannot decrease. If a vertex v is of degree $3 - i$ in G', $1 \leq i \leq 3$, then there are at least i adjacent edges to distinct vertices in triangles or diamonds that were removed from G and increase $t(v)$. Each time when a neighbor of a vertex v from a diamond or a triangle is removed then $t(v)$ increases by at least one. Then, in G', each vertex v of degree $3 - i$ has $t(v) \geq i + 1$.

Let n'_i be the number of vertices of degree i in G'. By the previous remark, we have

$$\sum_{v \in V(G')} t(v_i) \geq 4n'_0 + 3n'_1 + 2n'_2 + n'_3 \tag{1}$$

Since G' is a subcubic triangle-free graph and M a maximum matching in G', using a result of Munaro [16], we get

$$|V(M)| \geq \frac{9}{10}n'_3 + \frac{3}{5}n'_2 + \frac{3}{10}n'_1 \tag{2}$$

We show now that $4|V(G'')| \leq \sum_{v \in V(G')} t(v_i)$. In fact, combining $|V(G')| = n'_0 + n'_1 + n'_2 + n'_3$ with inequality (2) gives $|V(G'')| \leq n'_0 + \frac{7}{10}n'_1 + \frac{2}{5}n'_2 + \frac{1}{10}n'_3$. Thus, $4|V(G'')| \leq 4n'_0 + \frac{28}{10}n'_1 + \frac{8}{5}n'_2 + \frac{4}{10}n'_3 \leq 4n'_0 + 3n'_1 + 2n'_2 + n'_3 \leq \sum_{v \in V(G')} t(v_i)$ using inequality (1). Then $4|V(G'')| \leq \sum_{v \in V(G')} t(v_i)$ and since $|V(G)| \geq \sum_{v \in V(G')} t(v_i)$ we get $|V(G'')| \leq \frac{1}{4}V(G)$.

Consider the partition $\mathcal{P} = D \cup T \cup M \cup V(G'')$ in the sense that $\mathcal{P}$ contains a set for each diamond in D, one set for each triangle in T, one set for each edge in the matching M and one set for each vertex in $V(G'')$. Then $d(\mathcal{P}) = \frac{5}{4}|D| + |T| + \frac{1}{2}|M| \geq \frac{5}{4}|D| + |T| + \frac{1}{4}(n - 3|T| - 4|D| - \frac{n}{4})$ since $|V(G'')| \leq \frac{1}{4}V(G)$. By Lemma 10 it follows that:

$$\frac{opt(I)}{d(\mathcal{P})} \leq \frac{\frac{5}{4}|D| + |T| + \frac{1}{4}(n - 3|T| - 4|D|)}{\frac{5}{4}|D| + |T| + \frac{1}{4}(n - 3|T| - 4|D| - \frac{n}{4})} = \frac{\frac{1}{4}|D| + \frac{1}{4}|T| + \frac{n}{4}}{\frac{1}{4}|D| + \frac{1}{4}|T| + \frac{3n}{16}} = 1 + \frac{n}{4|D| + 4|T| + 3n} \leq 1 + \frac{1}{3}. \qquad \blacktriangleleft$$

5   Dense Graphs

In this section we consider graphs $G = (V, E)$ on n vertices such that G can be viewed as $G = \overline{H}$ where H is a graph of small maximum degree. Note that the edges of H are exactly the *missing edges* of G. We first consider graphs $G = (V, E)$ on n vertices such that $\delta(G) \geq n - 3$, that is $G = \overline{H}$ where H has $\Delta(H) = 2$ and has $q \leq n$ edges and show that MAX DENSE GRAPH PARTITION is solvable in polynomial time on these graphs.

▶ **Lemma 16.** *For any graph G on n vertices such that $\delta(G) \geq n - 3$, its density $d(G)$ is greater than or equal to the density of any partition $\mathcal{P}$ of G into $t \geq 3$ parts.*

Proof. The density of G is given by $d(G) = \frac{\frac{n(n-1)}{2} - q}{n} = \frac{n-1}{2} - \frac{q}{n}$. From Lemma 2, among all partitions of G into $t \geq 3$ parts, those where the parts correspond to complete graphs have the largest density. The density of such a partition into t parts of size $n_1, \ldots, n_t$ is $\frac{n-t}{2}$. Thus, the density of G is at least as large as the density of this last partition since $t \geq 3$ and $q \leq n$ (note here that a graph with minimum degree $n - 3$ has at most n missing edges).   ◄

Observe that in the proof of the previous lemma when $q = n$ and $t = 3$, the density of a partition in 3 parts corresponding to complete subgraphs and the density of the entire graph are the same. This previous lemma implies that for any graph G such that $\delta(G) \geq n - 3$, there exists a partition into one or two parts of maximum density.

▶ **Lemma 17.** *For any graph G on n vertices such that $\delta(G) \geq n - 3$, in any partition for G into two parts, the sum of missing edges in the two parts is at least o, where o is the number of odd cycles in $\overline{G}$.*

Proof. Let C be an odd cycle in $\overline{G}$. Since C is not bipartite, there is no partition $\{V_1, V_2\}$ of V such that all the edges of C have one endpoint in V_1 and one endpoint in V_2. Hence, for any partition $\{V_1, V_2\}$ at least one of the missing edges from C is inside $G[V_1] \cup G[V_2]$.   ◄

▶ **Lemma 18.** *Among all partitions into 2 parts of fixed size containing x missing edges, the one containing all missing edges in the largest part has the best density.*

Proof. Consider two partitions $\{V_1, V_2\}$ and $\{V_1', V_2'\}$ such that $|V_1| = |V_1'| = n_1$ and $|V_2| = |V_2'| = n_2$ with $n_1 \leq n_2$ and $G[V_1]$ (resp. $G[V_2]$) containing x_1 (resp. x_2) missing edges and $G[V_1']$ (resp. $G[V_2']$) containing 0 (resp. $x = x_1 + x_2$) missing edges. Then:
$d(\{V_1, V_2\}) = \frac{n-2}{2} - \frac{x_1}{n_1} - \frac{x_2}{n_2}$, and
$d(\{V_1', V_2'\}) = \frac{n-2}{2} - \frac{x}{n_2}$.
Since $x = x_1 + x_2$ and $n_1 \leq n_2$, it follows that $d(\{V_1, V_2\}) \leq d(\{V_1', V_2'\})$.   ◄

▶ **Lemma 19.** *Among all partitions into 2 parts containing 0 (resp. x) missing edges in the smaller (resp. larger) part, the one with a maximum number of vertices in the largest part has the best density.*
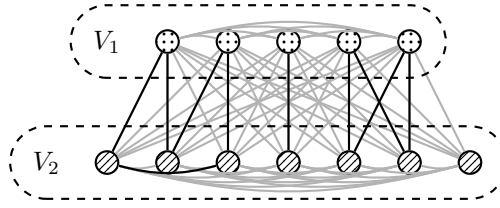
Proof. Consider two partitions $\{V_1, V_2\}$ and $\{V_1', V_2'\}$ such that $|V_1| = n_1$, $|V_2| = n_2$ with $n_1 \leq n_2$ and $|V_1'| = n_1'$, $|V_2'| = n_2'$ with $n_1' \leq n_2'$ and $G[V_1]$ (resp. $G[V_2]$) containing 0 (resp. x) missing edges and $G[V_1']$ (resp. $G[V_2']$) containing 0 (resp. x) missing edges. Moreover suppose $n_2 \leq n_2'$. The densities for these partitions are:
$d(\{V_1, V_2\}) = \frac{n-2}{2} - \frac{x}{n_2}$, and
$d(\{V_1', V_2'\}) = \frac{n-2}{2} - \frac{x}{n_2'}$.
Since $n_2 \leq n_2'$, it follows that $d(\{V_1, V_2\}) \leq d(\{V_1', V_2'\})$.   ◄

■ **Figure 5** Construction of V_1 and V_2 in Theorem 20.

▶ **Theorem 20.** MAX DENSE GRAPH PARTITION *is solvable in polynomial time on graphs G with n vertices with $\delta(G) \geq n - 3$.*

Proof. Let G be a graph of minimum degree $n - 3$. We first define a partition $\{V_1, V_2\}$ of the vertices of G by giving vertices color 1 or 2, in the sense that V_1 (resp. V_2) contains vertices of color 1 (resp. 2). An example is given in Figure 5. We assign color 2 to each vertex of degree $n - 1$. Since the minimum degree in G is $n - 3$, the graph H of missing edges is a collection of paths and cycles. We color the vertices on paths or cycles with an even number of vertices alternating by 1 and 2. For vertices on paths or cycles with an odd number of vertices we also color them alternating by 1 and 2, always starting with color 2. Thus cycles of odd size have two adjacent vertices of color 2.

Let o be the number of odd cycles in H. The partition $\{V_1, V_2\}$ defined by our 2-coloring contains o missing edges in V_2 and $|V_2|$ is maximized among all such partitions. Its density is equal to $\frac{n-2}{2} - \frac{o}{n_2}$, where $n_2 = |V_2|$. Denote by d_{n-1} the number of vertices of G of degree $n - 1$ and by p_o the number of paths with an odd number of vertices (even length) among the missing edges. The sets V_1 and V_2 contain the same number of vertices of degree $n - 2$ that are extremities of a path with an even number of vertices in H. The set V_2 contains p_o more vertices of degree $n - 2$, that are extremities of a path with an odd number of vertices, than V_1. The set V_2 contains o more vertices of degree $n - 3$ than V_1. Thus $n_1 = \frac{1}{2}(n - d_{n-1} - p_o - o)$ and $n_2 = \frac{1}{2}(n + d_{n-1} + p_o + o)$. We claim that there is no partition into two parts that has a higher density.

By Lemma 17, any partition into two sets contains at least o missing edges inside the two parts. By construction we have maximized the number of vertices in the part with the missing edges among all partitions with the minimum number o of missing edges, i.e., there is no partition into two parts $\{V_1', V_2'\}$ with o missing edges all contained in V_2' and $|V_2'| > |V_2|$. Hence, by Lemmas 18 and 19, it remains to show that any partition $\{V_1', V_2'\}$ with $o + x > o$ missing edges for some $x > 0$ has a smaller density than $\{V_1, V_2\}$.

Let $\{V_1', V_2'\}$ be a partition with $o + x > o$ missing edges for some $x > 0$ and assume w.l.o.g. that $|V_1'| \leq |V_2'|$. By definition of the partition $\{V_1, V_2\}$, it follows that $|E(H)| = 2n_1 - p_o + o$ (note that all non-edges have to either be among the o missing edges in the partition or in the cut between V_1 and V_2). In the partition $\{V_1', V_2'\}$, it follows that $|E(H)| \leq 2|V_1'| - r_1 + (o + x)$, where r_1 is the number of vertices in V_1' adjacent to only one edge in H. In the cut between V_1' and V_2', each vertex in V_1' is adjacent to at most two such edges. Combining these two bounds on $|E(H)|$ yields

$$2n_1 - p_o \leq 2|V_1'| - r_1 + x. \tag{3}$$

We claim that $r_1 \geq p_o - x$. To see this, observe that every path of odd length either results in a vertex in V_1' adjacent to only one edge in $E(H)$ (r_1) or in a missing edge. Also, every cycle of odd length creates at least one missing edge. Thus the number of missing edges $o + x$ for $\{V_1', V_2'\}$ is at least $p_o - r_1 + o$. Reordering this yields the claimed

$$r_1 \geq p_o - x. \tag{4}$$
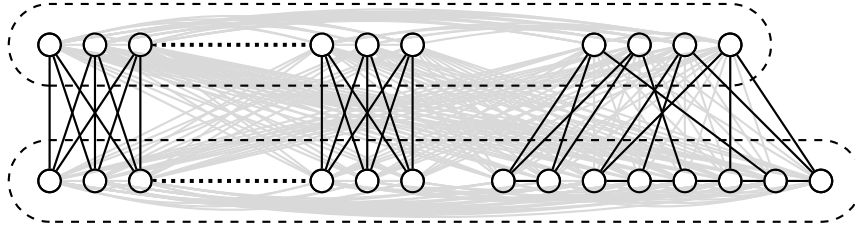
Figure 6 The construction of G' in Definition 21.

Inequalities (3) and (4) yield $2n_1 - p_o \leq 2|V_1'| - p_o + 2x$ and thus $n_1 - |V_1'| \leq x$. Since $\{V_1', V_2'\}$ is a partition it follows that $|V_2'| = n - |V_1'| \leq n - n_1 + x = n_2 + x$.

By Lemmas 18 and 19, the best case of missing edges for $\{V_1', V_2'\}$ is that they all are in the larger part V_2', hence the density of $\{V_1', V_2'\}$ is at most $\frac{n-2}{2} - \frac{o+x}{|V_2'|}$. With $|V_2'| \leq n_2 + x$, we can bound $d(V_1', V_2') \leq \frac{n-2}{2} - \frac{o+x}{n_2+x}$. Since H is of degree at most 2, we know that there cannot be more missing edges than vertices in a part, thus in particular $o \leq n_2$. This last observation allows to bound $d(V_1', V_2') \leq \frac{n-2}{2} - \frac{o+x}{n_2+x} \leq \frac{n-2}{2} - \frac{o}{n_2} = d(V_1, V_2)$, thus the density of $\{V_1', V_2'\}$ is not larger than the density of $\{V_1, V_2\}$.   ◀

We now consider graphs $G = (V, E)$ on n vertices that are $(n-4)$-regular, that is $G = \overline{H}$ where H is a cubic graph. We show that DENSE GRAPH PARTITION is NP-hard on $(n-4)$-regular graphs, by showing a reduction from MIN UNCUT on cubic graphs, that is the complement of MAX CUT. This last problem on cubic graphs was proved NP-hard and even not polynomial-time 1.003-approximable, unless P=NP [3].

MIN UNCUT
Input: A graph $G = (V, E)$, an integer k.
Question: Does G contain a partition of V into two parts A, B such that the number of edges with both endpoints in the same part is at most k?

▶ **Definition 21.** *Let $I = (G, k)$ be an instance of* MIN UNCUT *where $G = (V, E)$ is a cubic graph. We define the construction σ transforming the graph G into the graph $G' := (V', E') = \sigma(G)$ (see Figure 6) as follows:*
- *let $G_0 = (V_0, E_0)$ be the union of $\frac{n^2-n}{6}$ copies of $K_{3,3}$ (see remark below). Thus G_0 is a cubic bipartite graph with $n^2 - n$ vertices and V_0 is the union of two independent sets L, R such that $|L| = |R|$.*
- *let $G_1 = (V \cup V_0, E \cup E_0)$.*
- *let $G' = \overline{G_1}$.*

▶ **Remark 22.** Note that we can assume that the number of vertices of a cubic graph G is a multiple of 6. Since G is cubic, n is a multiple of 2. If n is not a multiple of 3, we consider the instance I_{triple} defined as follows: G_{triple} is the union of 3 copies of G and $k_{triple} = 3k$, and thus in the new instance I_{triple} the graph has $3n$ vertices. Note that the number of edges with both endpoints in the same part is $3k$ in G_{triple} if and only if it is k in G.

Let $n = |V|$, $m = |E|$, $n' = |V'|$ and $m' = |E'|$. Observe that $n' = n^2$, and G' is a $(n'-4)$-regular graph. With this construction one can show the following.

▶ **Theorem 23.** DENSE GRAPH PARTITION *is NP-complete on $(n-4)$-regular graphs with n vertices.*

At the end of this section we show that a partition into a bounded number of cliques provides a good approximation for graphs of large minimum degree.

▶ **Theorem 24.** Dense Graph Partition *is polynomial-time $\frac{n-1}{\delta(G)+1}$-approximable on graphs G with n vertices.*

Proof. Let G be a graph on n vertices with minimum degree $\delta = \delta(G)$, instance of Max Dense Graph Partition. If $\delta \geq n - 3$, we can give an optimum solution in polynomial time by Theorem 20. So assume $\delta \leq n - 4$. By Lemma 3, any partition $\mathcal{P}$ for the vertices of G satisfies $d(\mathcal{P}) \leq \frac{n-1}{2}$. Using Brooks' theorem [5], $\overline{G}$ is $(n - \delta - 1)$-colorable, and further, such a coloring can be computed in polynomial time. (Note that $\delta \leq n - 4$ implies that $\overline{G}$ is not a complete graph or a circle, the two exceptions in Brooks' theorem where one more color is needed.) Using such a coloring, G can be partitioned into $n - \delta - 1$ cliques. Then the density of this partition is $\frac{n-(n-\delta-1)}{2} = \frac{\delta+1}{2}$. Comparing this value with the upper bound of $\frac{n-1}{2}$ on the optimum shows that this partition into $n - \delta - 1$ cliques gives a polynomial-time $\frac{n-1}{\delta+1}$-approximation for Dense Graph Partition. ◀

Notice that if $\delta(G) > \frac{n-3}{2}$, the ratio given in Theorem 24 improves upon the current best ratio of 2 for Dense Graph Partition on general graphs. This approximation can further be used to show the following.

▶ **Theorem 25.** *There is an efficient polynomial-time approximation scheme for Max Dense Graph Partition on graphs G with n vertices and $\delta(G) = n - t$ for a constant $t \geq 4$.*

Proof. Let $I = G$ be a graph on n vertices and $\delta(G) = n - t$, instance of Max Dense Graph Partition. We establish in the following an eptas. Given $\varepsilon > 0$, consider two cases.

If $n \geq t - 1 + \frac{t-2}{\varepsilon}$, then let $\mathcal{P}$ be a partition that corresponds to a $(t-1)$-coloring of $\overline{G}$ such that each part is a clique in G as in the proof of Theorem 24. Then $d(\mathcal{P}) = \frac{n-t+1}{2} \geq \frac{n+1-\frac{n\varepsilon+\varepsilon+2}{1+\varepsilon}}{2} \geq \frac{n-1}{2(1+\varepsilon)} \geq \frac{opt(I)}{1+\varepsilon}$, where the last inequality $opt(I) \leq \frac{n-1}{2}$ comes from Lemma 3.

Otherwise, that is $n < t - 1 + \frac{t-2}{\varepsilon}$, enumerate all the partitions of G and consider the best one. Since the number of partitions of G is the Bell number of order $|V| = n$, B_n, and $B_n \leq n^n$, we get an optimal solution in time $(1/\varepsilon)^{O(1/\varepsilon)}$. ◀

6   Conclusion

In order to have a better understanding of the complexity of Max Dense Graph Partition it would be nice to study it on other graph classes. It was proved to be polynomial-time solvable on trees, but the complexity on graphs of bounded treewidth remains open. Moreover no result exists on split graphs. Concerning approximation, no lower bound was established, it would be nice to improve the 2-approximation algorithm or to show that no polynomial-time approximation scheme exist on general instances.

References

1   Sanjeev Arora, David R. Karger, and Marek Karpinski. Polynomial time approximation schemes for dense instances of np-hard problems. *Journal of Computer and System Sciences*, 58(1):193–210, 1999.

2   Haris Aziz, Serge Gaspers, Joachim Gudmundsson, Julián Mestre, and Hanjo Taubig. Welfare maximization in fractional hedonic games. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence, IJCAI 2015*, pages 468–474. AAAI Press, 2015.

3 Piotr Berman and Marek Karpinski. On some tighter inapproximability results (extended abstract). In Jiří Wiedermann, Peter van Emde Boas, and Mogens Nielsen, editors, *Proceedings of the 26th International Colloquium on Automata, Languages and Programming, ICALP 1999*, volume 1644 of *LNCS*, pages 200–209. Springer, 1999.

4 Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities: an $O(n^{1/4})$ approximation for densest k-subgraph. In Leonard J. Schulman, editor, *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 201–210. ACM, 2010.

5 Rowland Leonard Brooks. On colouring the nodes of a network. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 37:2, pages 194–197. Cambridge University Press, 1941.

6 Aydin Buluç, Henning Meyerhenke, Ilya Safro, Peter Sanders, and Christian Schulz. Recent advances in graph partitioning. In Lasse Kliemann and Peter Sanders, editors, *Algorithm Engineering - Selected Results and Surveys*, volume 9220 of *Lecture Notes in Computer Science*, pages 117–158. Springer, 2016.

7 Derek G. Corneil and Yehoshua Perl. Clustering and domination in perfect graphs. *Discrete Applied Mathematics*, 9(1):27–39, 1984.

8 Julien Darlay, Nadia Brauner, and Julien Moncel. Dense and sparse graph partition. *Discrete Applied Mathematics*, 160(16-17):2389–2396, 2012.

9 Uriel Feige, Guy Kortsarz, and David Peleg. The dense k-subgraph problem. *Algorithmica*, 29(3):410–421, 2001.

10 Santo Fortunato. Community detection in graphs. *Physics Reports*, 486:75–174, 2010.

11 M. R. Garey, D. S. Johnson, and L. Stockmeyer. Some simplified NP-complete graph problems. *Theoretical Computer Science*, 1(3):237–267, 1976.

12 Andrew V. Goldberg. *Finding a maximum density subgraph*. University of California Berkeley, 1984.

13 Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

14 Subhash Khot. Ruling out PTAS for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM J. Comput.*, 36(4):1025–1071, 2006.

15 Samir Khuller and Barna Saha. On finding dense subgraphs. In *Proceedings of 36th International Colloquium on Automata, Languages and Programming, ICALP 2009, Part I*, volume 5555 of *LNCS*, pages 597–608, 2009.

16 Andrea Munaro. On line graphs of subcubic triangle-free graphs. *Discrete Mathematics*, 340(6):1210–1226, 2017.

17 Mark E. J. Newman. Detecting community structure in networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(2):321–330, 2004.

18 Satu Elisa Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27–64, 2007.