Sparse halves in dense triangle-free graphs

Sergey Norin[*1] and Liana Yepremyan[†2]

[1]Department of Mathematics and Statistics, McGill University
[2]School of Computer Science, McGill University

Abstract

Erdős [3] conjectured that every triangle-free graph G on n vertices contains a set of $\lfloor n/2 \rfloor$ vertices that spans at most $n^2/50$ edges. Krivelevich proved the conjecture for graphs with minimum degree at least $\frac{2}{5}n$ [9]. In [8] Keevash and Sudakov improved this result to graphs with average degree at least $\frac{2}{5}n$. We strengthen these results by showing that the conjecture holds for graphs with minimum degree at least $\frac{5}{14}n$ and for graphs with average degree at least $\left(\frac{2}{5} - \gamma\right)n$ for some absolute $\gamma > 0$. Moreover, we show that the conjecture is true for graphs which are close to the Petersen graph in edit distance.

Keywords: Triangle-free graph, sparse half, minimum degree, Petersen graph, edit distance, blowup

1.   Introduction

In this paper we consider the edge distribution in triangle-free graphs. A fundamental result in extremal graph theory, Turán's theorem implies that if a graph G on n vertices has more than $(1 - 1/r)\binom{n}{2}$ edges, then G has K_{r+1} as a subgraph. One can consider the following generalization of this problem that was first studied by Erdős, Faudree, Rousseau and Schelp [5] in 1990. Fix $\alpha \in (0, 1]$, and suppose that every set of αn vertices in G induces more than βn^2 edges. What is the smallest $\beta := \beta(\alpha, r)$ that forces G to contain K_{r+1}?

In particular, one of Erdős's old and favorite conjectures says that $\beta(\frac{1}{2}, 2) = \frac{1}{50}$ [3]; he first proposed this in 1975 and offered a \$250 prize for its solution later, in [4]. It is easy to check that the bound $1/50$ is tight. It is achieved on the uniform *blowup* of C_5 which is obtained from the 5-cycle by replacing each vertex i by an independent set V_i of size $n/5$ (for simplicity assume n is divisible by 5) and each edge ij by a complete bipartite graph joining V_i and V_j. The blowup of the Petersen graph also achieves this bound tightly, see Figure 1.

Conjecture 1.1 (Erdős). *Any triangle-free graph G on n vertices contains a set of $\lfloor n/2 \rfloor$ vertices that spans at most $n^2/50$ edges.*
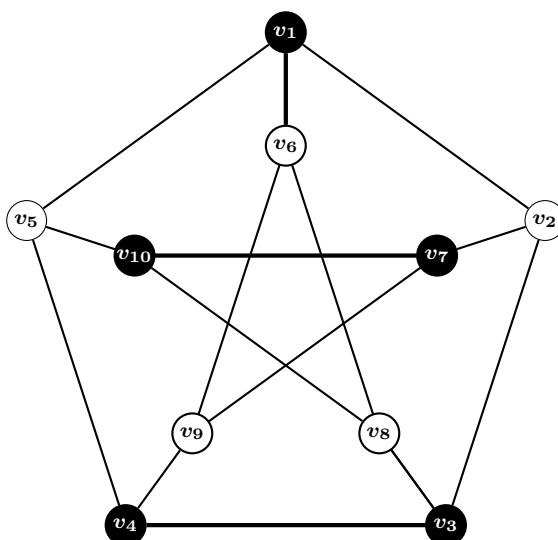
Figure 1: A sparse half in the uniform blowup of Petersen graph.

In [9] Krivelevich proved that the conjecture holds if $1/50$ is replaced by $1/36$. He also showed that it is true for triangle-free graphs with minimum degree $\frac{2}{5}n$. In Section 3 we improve this result by proving the following theorem.

Theorem 1.1. *Every triangle-free graph on n vertices with minimum degree at least $\frac{5}{14}n$ contains a set of $\lfloor n/2 \rfloor$ vertices that spans at most $n^2/50$ edges.*

Our proof of Theorem 1.1 is mainly based on the structural characterization of the graphs with minimum degree at least $\frac{5}{14}n$, established by Jin, Chen and Koh in [2, 7]. We also use some averaging arguments similar to the ones used in [8, 9].

Keevash and Sudakov [8] improved Krivelevich's result of [9] showing that the conjecture holds for graphs with average degree $\frac{2}{5}n$. In Section 5 we extend their result as follows.

Theorem 1.2. *There exists $\gamma > 0$ such that every triangle-free graph on n vertices with at least $(\frac{1}{5} - \gamma)n^2$ edges contains a set of $\lfloor n/2 \rfloor$ vertices that spans at most $n^2/50$ edges.*

Finally, we study the validity of the conjecture in the neighborhood of the known extremal examples, in the following sense. In the uniform blowup of the Petersen graph every set of $\lfloor n/2 \rfloor$ has at least $n^2/50$ edges (see Figure 1). To the best of our knowledge, the uniform blowups of the Petersen graph and C_5, as defined in the beginning of the introduction, are the only known examples for which Conjecture 1.1 is tight. In Section 4 we develop a set of tools which allow us to prove the Conjecture 1.1 for the classes of graphs which are close to a fixed graph in edit distance. In Section 6 we use these tools to verify the conjecture for graphs which are close to the Petersen graph, while Theorem 1.2 shows that it also holds for graphs close to the 5-cycle. These results can

---

*snorine@gmail.com

†liana.yepremyan@mail.mcgill.ca

be considered as a proof of a local version of Conjecture 1.1, in the spirit of recent results of Lovász [10] and Razborov [12, 13]. Lovász [10] proved Sidorenko's conjecture locally in the neighborhood of the conjectured extremal example with respect to the cut-norm (and consequently, with respect to the edit distance). Razborov proved that the Caccetta-Häggkvist conjecture [12] and the Turán's $(3,4)$-problem [13] hold locally in the neighborhood of families of the known examples in a slightly different sense, as he additionally forbade certain induced subconfigurations.

2.   Notation and Preliminary Results

For a graph G, we let $\mathsf{v}(G) := |V(G)|$ and $\mathsf{e}(G) := |E(G)|$. Denote by $N_G(v)$ the neighborhood of a vertex $v \in V(G)$ and by $d_G(v)$ the degree of v (i.e. $d_G(v) := |N_G(v)|$). Whenever there is no ambiguity, we will use $N(v)$ and $d(v)$ for shorter notation. The maximum and the minimum degrees are denoted by $\Delta(G)$ and $\delta(G)$, respectively.

A graph G is called *triangle-free* if it does not contain a triangle, it is *maximal triangle-free* if adding any edge creates a triangle. We say that G *contains a sparse half* if there exists a set of $\lfloor n/2 \rfloor$ vertices in G that spans at most $n^2/50$ edges. Conjecture 1.1 says that there must exist a sparse half in every triangle-free graph.

We say that $\omega : V(G) \to (0,1)$ is a *weight function on G* if $\sum_{v \in V(G)} \omega(v) = 1$. A pair (G, ω), where ω is a weight function on G is called a *weighted graph*. The weight $\omega(e)$ of an edge $e = (u,v)$ in (G, ω) is defined as $\omega(u) \cdot \omega(v)$. For a set X of vertices or edges in G let $\omega(X) := \sum_{x \in X} \omega(x)$. The *degree* of a vertex v in a weighted graph (G, ω) is defined as $\omega(N(v))$. The *minimum degree* of the weighted graph (G, ω) is denoted by $\delta(G, \omega)$.

We call a real function $\mathsf{s} : V(G) \to \mathbb{R}^+$ *a half* of G if $\mathsf{s}(v) \le \omega(v)$ for every $v \in V(G)$, and $\mathsf{s}(V(G)) := \sum_{v \in V(G)} \mathsf{s}(v) = 1/2$. For every edge $e = (u,v)$, we define $\mathsf{s}(e) := \mathsf{s}(u)\mathsf{s}(v)$. Let s be a half of G, if $\mathsf{s}(E(G)) := \sum_{e \in E(G)} \mathsf{s}(e) \le \frac{1}{50}$ then s is called a *sparse half*.

There is a natural way of associating a weighted graph to a graph G of order n; that is, we can assign to each vertex $v \in V(G)$ weight equal to $\frac{1}{n}$. For each graph G, the corresponding weighted graph defined as above, is called *uniformly weighted G*; we denote it by (G, ω_u).

Lemma 2.1. *If (G, ω_u) has a sparse half, then so does G.*

Proof. We claim that, if (G, ω_u) has a sparse half, then it has a sparse half s such that $\mathsf{s}(v) = 0$ or $\mathsf{s}(v) = \frac{1}{n}$ for all $v \in V(G)$ except for possibly one vertex.

Indeed, let us choose a sparse half s of G such that the number of vertices $v \in V(G)$ such that either $\mathsf{s}(v) = 0$ or $\mathsf{s}(v) = \frac{1}{n}$ is maximum. We show that there exists at most one vertex u such that $0 < \mathsf{s}(u) < \frac{1}{n}$.

Suppose not and there exist $u, v \in V(G)$ such that $0 < \mathsf{s}(u), \mathsf{s}(v) < \frac{1}{n}$. We define a new half $\mathsf{s}' : V(G) \to \mathbb{R}^+$. Let s' be the same as s on all vertices of G, except u and v. Without loss of generality, suppose $\mathsf{s}(N(u)) \le \mathsf{s}(N(v))$. Let $\delta = \min\{\mathsf{s}(u), \frac{1}{n} - \mathsf{s}(v)\}$ and define $\mathsf{s}'(u) = \mathsf{s}(u) - \delta$ and $\mathsf{s}'(v) = \mathsf{s}(v) + \delta$. We will show that s' is a sparse half.

Suppose that u and v are adjacent, then

$$\mathsf{s}'(E(G)) = \mathsf{s}(E(G)) - \sum_{\substack{x \in N(u) \\ x \neq v}} \delta\mathsf{s}(x) + \sum_{\substack{y \in N(v) \\ y \neq u}} \delta\mathsf{s}(y) + \mathsf{s}'(u)\mathsf{s}'(v) - \mathsf{s}(u)\mathsf{s}(v)$$

$$= \mathsf{s}(E(G)) - \sum_{\substack{x \in N(u) \\ x \neq v}} \delta\mathsf{s}(x) + \sum_{\substack{y \in N(v) \\ y \neq u}} \delta\mathsf{s}(y) - \delta\mathsf{s}(v) + \delta\mathsf{s}(u) - \delta^2$$

$$= \mathsf{s}(E(G)) - \delta\left(\mathsf{s}(N(u)) - \mathsf{s}(N(v))\right) - \delta^2 < \mathsf{s}(E(G)).$$

The calculation in the case when u and v are non-adjacent is similar.

It follows that s' contradicts the choice of s. Hence for all vertices $v \in V(G)$ except maybe one vertex either $\mathsf{s}(v) = 0$ or $\mathsf{s}(v) = \frac{1}{n}$. Let

$$S = \{v \in V(G) \mid \mathsf{s}(v) = 1/n\}.$$

It follows from the previous observation that $|S| \geq \lfloor n/2 \rfloor$. It is easy to see that $\mathsf{e}(G[S]) \leq n^2\mathsf{s}(E(G)) \leq n^2/50$, where $G[S]$ is defined to be the subgraph of G induced by the vertex set S. This shows that S ia a sparse half in G, as desired. $\square$

Lemma 2.1 allows us to work with weighted graphs, which proves to be convenient. We prove that every weighted triangle-free graph with minimum degree at least $5/14$ contains a sparse half. Our proof uses structural characterization of these graphs found by Jin, Chen and Koh in [2, 7]. To state their result we need a few additional definitions.

A mapping $\varphi : V(G) \to V(H)$ for graphs G and H is a *homomorphism from G to H*, if it is edge-preserving, that is, for any pair of adjacent vertices $u, v \in V(G)$, $\varphi(u)$ and $\varphi(v)$ are adjacent in H. We say that G is *of H-type* if there exists a homomorphism from G to H. Let φ be a surjective homomorphism and ω be a weight function on G. We define a weight function ω_φ on H in the following way. For every vertex $v \in V(H)$, let $\omega_\varphi(v) := \omega(\varphi^{-1}(v))$. The next lemma shows that a sparse half in a homomorphic image of a graph G can be lifted to a sparse half in the graph G.

Lemma 2.2. *Let G, H be graphs and $\varphi : V(G) \to V(H)$ be a surjective homomorphism. Then for any weight function ω, if (H, ω_φ) has a sparse half, then so does (G, ω).*

Proof. Let s_H be a sparse half on (H, ω_φ). Define

$$\mathsf{s}_G(v) := \frac{\omega(v)}{\omega_\varphi(\varphi(v))}\mathsf{s}_H(\varphi(v))$$

for every $v \in V(G)$. It is easy to check that s_G is a sparse half of G. $\square$

Now let us introduce a family of graphs that plays a key role in our results. For an integer $d \geq 1$, let F_d be the graph with

$$V(F_d) = \{v_1, v_2, ..., v_{3d-1}\},$$

such that the vertex v_j has neighbors $v_{j+d}, \ldots, v_{j+2d-1}$ (all indices larger than $3d - 1$ are taken modulo $3d - 1$). It is easy to see that $F_1 = K_2$, $F_2 = C_5$ and F_3 is a Möbius
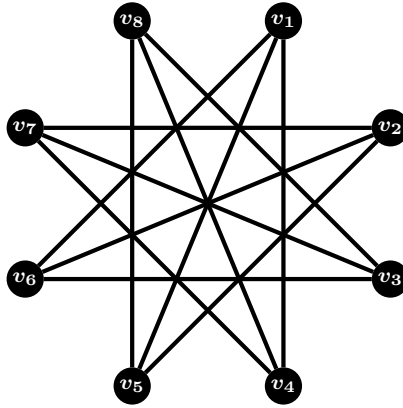
Figure 2: The Möbius ladder F_3

ladder (see Figure 2). Clearly, if a graph is of F_d-type, it is of F_k-type, for all $k > d$. These graphs were first introduced by Woodall [14]. In 1973 he conjectured that if the binding number of a graph G is at least $3/2$ then G contains a triangle. Graphs F_d are in the family of graphs constructed by Woodall to explain that $3/2$ in his conjecture is the least possible. Let us state some general properties of these graphs that we will use implicitly or explicitly throughout the paper. These can be easily checked.

Fact 2.3. *For every $d \geq 1$, F_d is triangle-free, 3-colorable and the only maximum independent sets are vertex neighborhoods, in particular, $\alpha(F_d) = d$.*

In 1974 Andrasfai, Erdős and Sos [1] showed that every triangle-free graph G of order n with $\delta(G) > 2n/5$ is bipartite (in other words, F_1-type). In a similar spirit, Häggkvist [6] proved that every triangle-free graph G of order n with $\delta(G) > 3n/8$ is of F_2-type (i.e. C_5-type). In 1995 Jin [7] improved Häggkvist's result as following.

Theorem 2.4 (Jin, [7]). *Every triangle-free graph G of order n with minimum degree $\delta(G) > 10n/29$ is of F_9-type.*

Note that the results of Andrasfai, Erdős and Sos [1], Häggkvist [6] and Jin [7] are sharp. In [2] the authors extended all the results mentioned above and proved the following stronger statement.

Theorem 2.5 (Chen, Jin, Koh, [2]). *Let $d \geq 1$, G be a triangle-free graph of order n with $\chi(G) \leq 3$ and $\delta(G) > \left\lfloor \frac{(d+1)n}{3d+2} \right\rfloor$, then G is of F_d-type.*

For our purposes, we use the following statement which follows from Theorem 2.4, Theorem 2.5 and Fact 2.3.

Corollary 2.6 (Chen, Jin, Koh [2], [7]). *Every triangle-free graph G of order n with $\delta(G) \geq \frac{5}{14}n$ is of F_5-type.*

To use Lemma 2.1 and Corollary 2.6 together we need to make sure that the homomorphism in the statement of Corollary 2.6 is surjective. Fortunately, this is not difficult if we allow ourselves to change the target graph.

5

Lemma 2.7. *Let φ be a homomorphism from graph G to F_d, $d \geq 2$. Then either φ is surjective or there exists a homomorphism φ' from G to F_{d-1}.*

Proof. Suppose φ is a homomorphism from graph G to F_d that is not surjective. Let $V_i = \varphi^{-1}(v_i)$ for all $i = 1, 2, \ldots, 3d - 1$. Without loss of generality suppose V_1 is empty. Define a mapping $\varphi' : V(G) \to V(F_{d-1})$ as follows,

$$\varphi'(v) = \begin{cases} v_{i-1}, & \text{if } v \in V_i \text{ and } 2 \leq i \leq d - 1, \\ v_{d-1}, & \text{if } v \in V_d \cup V_{d+1}, \\ v_{i-2}, & \text{if } v \in V_i \text{ and } d + 2 \leq i \leq 2d - 2, \\ v_{2d-3}, & \text{if } v \in V_{2d-1} \cup V_{2d}, \\ v_{i-3}, & \text{if } v \in V_i \text{ and } 2d + 1 \leq i \leq 3d - 1. \end{cases}$$

It is easy to check that φ' is a homomorphism from G to F_{d-1}. $\qquad\square$

In the next section we show that for $1 \leq d \leq 4$ the weighted graph (F_d, ω) with minimum degree at least $5/14$, has a sparse half for any weight function ω. In particular, if $\varphi : V(G) \to F_d$ is a surjective homomorphism, then the weighted graph (F_d, ω_φ) has a sparse half. By Lemma 2.1 this implies that graph G has a sparse half. Hence Theorem 1.1 will follow from the results of the next section.

3. Weighted triangle-free graphs with minimum degree $\geq \frac{5}{14}$

Theorem 3.1. *For all $1 \leq d \leq 5$, if (F_d, ω) is a weighted graph with minimum degree at least $5/14$, then it has a sparse half.*

Proof. The argument is separated into cases based on the value of d.

d=1: Suppose $V(F_1) = \{v_1, v_2\}$ then since $\omega(v_1) + \omega(v_2) = 1$, either $\omega(v_1) \geq \frac{1}{2}$ or $\omega(v_2) \geq \frac{1}{2}$, therefore v_1 or v_2 supports a sparse half in (F_1, ω).

d=2: Let $V(F_2) = \{v_1, v_2, \ldots, v_5\}$. If any two consecutive vertices together have total weight at least $1/2$, then they induce an independent set, which means that they support a sparse half.

Now suppose that no two consecutive vertices have total weight at least $1/2$, then any three consecutive vertices have total weight at least $1/2$. We define the following halves s_i for each $i = 1, 2, \ldots, 5$ on the vertices of the graph and prove that there is at least one sparse half among them.

$$\mathsf{s}_i(v) = \begin{cases} \omega(v), & \text{if } v = v_i \text{ or } v = v_{i+1}, \\ \frac{1}{2} - (\omega(v_i) + \omega(v_{i+1})), & \text{if } v = v_{i+2}, \\ 0, & \text{otherwise.} \end{cases}$$
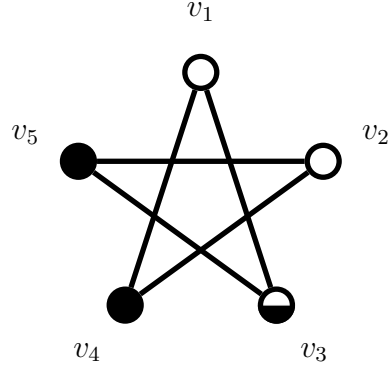
Figure 3: A sparse half in uniformly weighted C_5

Note that
$$s_i(E(G)) = \omega(v_i)\left(1/2 - (\omega(v_i) + \omega(v_{i+1}))\right). \tag{1}$$
Summing up the equations (1) over all $i = 1, \ldots, 5$, we get

$$\sum_{i=1}^{5} s_i(E(G)) = 1/2 - \sum_{i=1}^{5} \omega(v_i)\left(\omega(v_i) + \omega(v_{i+1})\right)$$

$$= \frac{1}{2} - \frac{1}{2}\sum_{i=1}^{5} \left(\omega(v_i) + \omega(v_{i+1})\right)^2$$

$$\leq \frac{1}{2} - \frac{1}{2} \cdot 5 \cdot \frac{4}{25} = \frac{1}{10},$$

using Jensen's inequality for the function x^2. Thus one of the functions s_i is a sparse half. Note that the proof above did not use the minimum degree condition, while we use it in the other two cases.

d=3: Let $F_3 = \{v_1, v_2, \ldots, v_8\}$. As the minimum degree of (F_3, ω) is at least $5/14$, we have

$$\omega(v_j) + \omega(v_{j+1}) + \omega(v_{j+2}) \geq \delta(F_3, \omega) \geq 5/14, \tag{2}$$

for all $j = 1, 2, \ldots, 8$. Summing these inequalities for $j = i, i+3$ and $j = i+6$, we obtain $\omega(v_j) \geq 1/14$ for $j = 1, 2, \ldots, 8$.

As in the case of $d = 2$, if there exist three consecutive vertices of total weight at least $1/2$, then we are done, since they induce an independent set. Suppose not, then every five consecutive vertices have total weight more than $1/2$. We define the following halves s_i, for each $i = 1, 2, \ldots, 8$ on the vertices of the graph and prove that there is at least one sparse half among them.

$$s_i(v) = \begin{cases} \omega(v), & \text{if } v = v_{i+1}, v_{i+2}, v_{i+3} \\ \frac{1}{2}\left(\frac{1}{2} - (\omega(v_{i+1}) + \omega(v_{i+2}) + \omega(v_{i+3}))\right), & \text{if } v = v_i \text{ or } v = v_{i+4}, \\ 0, & \text{otherwise.} \end{cases}$$

Claim 3.2. *For each $i = 1, 2, \ldots, 8$, s_i is a half.*

7

Proof. It suffices to show that

$$\frac{1}{2}\left(\frac{1}{2} - (\omega(v_{i+1}) + \omega(v_{i+2}) + \omega(v_{i+3}))\right) \le \omega(v_i),$$

$$\frac{1}{2}\left(\frac{1}{2} - (\omega(v_{i+1}) + \omega(v_{i+2}) + \omega(v_{i+3}))\right) \le \omega(v_{i+4}).$$

By symmetry it suffices to prove the first inequality. By (2) we get that

$$\frac{1}{2}\left(\frac{1}{2} - (\omega(v_{i+1}) + \omega(v_{i+2}) + \omega(v_{i+3}))\right) \le \frac{1}{2}\cdot\left(\frac{1}{2} - \frac{5}{14}\right) = \frac{1}{14} \le \omega(v_i).$$

This finishes the proof of the claim. $\qquad\square$

We relegate the proof of the following lemma, which ensures that one of the halves s_i is sparse, to the Appendix.

Lemma 3.3. *Let $1/14 \le x_i \le 1$ for $i = 1, 2, \ldots, 8$. If $\sum_{i=1}^{8} x_i = 1$ and $x_i + x_{i+1} + x_{i+2} \ge 5/14$ for every i, then there exists i such that*

$$\frac{1}{2}\left(\frac{1}{2} - (x_i + x_{i+1} + x_{i+2})\right)(x_i + x_{i+2}) + \frac{1}{4}\left(\frac{1}{2} - (x_i + x_{i+1} + x_{i+2})\right)^2 \le \frac{1}{50}. \quad (3)$$

d=4: Let $F_4 = \{v_1, v_2, \ldots, v_{11}\}$. The minimum degree condition gives us the following inequality

$$\omega(v_i) + \omega(v_{i+1}) + \omega(v_{i+2}) + \omega(v_{i+3}) \ge 5/14, \quad\quad (4)$$

for all $i = 1, 2, \ldots, 11$. It follows, as in the case $d = 3$, $\omega(v_i) \ge 1/14$ for all i. We may assume that no four consecutive vertices have total weight at least $1/2$, otherwise, we are done - these four vertices induce an independent set. Therefore, we can assume that

$$\omega(v_i) + \omega(v_{i+1}) + \cdots + \omega(v_{i+6}) > 1/2,$$

for all $i = 1, 2, \ldots, 11$. This allows us to define halves s_i in the following way:

$$\mathsf{s}_i(v) = \begin{cases} \omega(v), & \text{if } v = v_{i+1}, v_{i+2}, v_{i+3}, v_{i+4}, \\ \frac{1}{2}\left(\frac{1}{2} - (\omega(v_{i+1}) + \omega(v_{i+2}) + \omega(v_{i+3}) + \omega(v_{i+4}))\right), & \text{if } v = v_i \text{ or } v = v_{i+5}, \\ 0, & \text{otherwise.} \end{cases}$$

As in the previous case, it is easy to verify that each s_i is a half, and the following lemma proved in the Appendix implies that at least one of these halves is sparse.

Lemma 3.4. *Suppose $x_1, x_2, \ldots, x_{11}$ are reals such that $1/14 \le x_i \le 1$ for each $i = 1, 2, \ldots, 11$ and $\sum_{i=1}^{11} x_i = 1$. If $x_i + x_{i+1} + x_{i+2} + x_{i+3} \ge 5/14$ for every i, then there exists i such that*

$$\frac{1}{2}\left(\frac{1}{2} - (x_{i+1} + x_{i+2} + x_{i+3} + x_{i+4})\right)(x_{i+1} + x_{i+4})$$

$$+ \frac{1}{4}\left(\frac{1}{2} - (x_{i+1} + x_{i+2} + x_{i+3} + x_{i+4})\right)^2 \le 1/50.$$

8

d=5: Let $F_5 = \{v_1, v_2, \ldots, v_{14}\}$. As in the previous cases, the minimum degree condition gives us $\omega(v_i) \geq \frac{1}{14}$ for every $i = 1, 2, \ldots, 14$. But on the other hand, $\sum_{i=1}^{14} \omega(v_i) = 1$, hence $\omega(v_i) = 1/14$, for each $i = 1, 2, \ldots, 14$. Define

$$\mathsf{s}(v) = \begin{cases} \omega(v), & \text{if } v = v_1, v_2, v_3, v_4, v_5, v_6, v_7 \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to check that s is a half. In fact, it is a sparse half. Indeed, $\mathsf{s}(E(G)) = \frac{3}{14^2} \leq \frac{1}{50}$. $\quad\square$

Proof of Theorem 1.1. Let G be a triangle-free graph on n vertices with minimum degree $\geq 5n/14$. By Lemma 2.1 it suffices to prove that the uniformly weighted graph (G, ω) has a sparse half. By Theorem 2.6 and Lemma 2.7, the graph G admits a surjective homomorphism φ to F_d for $1 \leq d \leq 5$. Clearly, (F_d, ω_φ) has minimum degree $\geq 5/14$ and thus has a sparse half by Theorem 3.1. Theorem 1.1 now follows from Lemma 2.2. $\quad\square$

4. Uniform sparse halves, balanced weights and disturbed subgraphs

The purpose of this section is to develop a set of tools, that, under fairly general circumstances, allow us to show that graphs "close" to a fixed graph have sparse halves. A number of technical definitions will be necessary. We start by a variant of the definition of *edit distance* (see e.g. [11]).

Definition 4.1. *Given graphs G and H of orders n and k respectively, we say that G can be ε-approximated by H, if there exists a partition $\mathcal{V} = \{V_1, V_2, \ldots, V_k\}$ of $V(G)$ such that*

$$\left| |V_i| - \frac{n}{k} \right| \leq \varepsilon n, \; \text{for all } i = 1, 2, \ldots k, \; \text{and } \mathsf{e}(G \triangle H_\mathcal{V}) \leq \varepsilon n^2,$$

where $H_\mathcal{V}$ is a graph on vertex set $V(G)$ with adjacency defined as follows. Each $H_\mathcal{V}[V_i]$ is an independent set, (V_i, V_j) induces a complete bipartite graph in $H_\mathcal{V}$ if $v_i v_j \in E(H)$ and an empty graph, otherwise.

We will also need a stronger notion of distance defined as follows.

Definition 4.2. *Given a graph G and $F \subseteq E(G)$, we say that $D \subseteq V(G)$ is an ε-covering set for F, if $|D| \leq \varepsilon \mathsf{v}(G)$ and every edge in F has at least one end in D.*

Definition 4.3. *We say that the graph G is an ε-disturbed subgraph of G' for some $0 < \varepsilon < 1$, if the following conditions hold:*

1. *$V(G) = V(G')$,*

2. *there exists an ε-covering set for $E(G) \setminus E(G')$ in G,*

3. *$|N_G(v) \setminus N_{G'}(v)| \leq \varepsilon \mathsf{v}(G')$ for every vertex $v \in V(G')$.*

Note that if G is an ε-disturbed subgraph of some graph G', it does not imply that G is a subgraph of G', as one might expect. However, from our definition, it follows that G is not "far" from some subgraph of G' and that is why we use ε-disturbed term.

For a graph H, let $\mathcal{N}(H)$ be the set of neighborhoods of vertices of H and $\mathcal{I}(H)$ be the set of maximum independent sets of H. Let $\mathcal{I}^*(H) = \mathcal{I}(H) \setminus \mathcal{N}(H)$. We construct a graph H^* as follows. Let $V(H^*) = V(H) \cup \mathcal{I}^*(H)$, H be an induced subgraph of H^*, and let every $I \in \mathcal{I}^*(H)$ be adjacent to every $v \in I$ and no other vertex of H^*. We say that a weighted graph (H^*, ω) is ε-balanced if $|\omega(v) - 1/\mathsf{v}(H)| \leq \varepsilon$ for every $v \in V(H)$ and $\omega(v) \leq \varepsilon$ for every $v \in V(H^*) \setminus V(H)$. We are now ready for our first technical result.

Theorem 4.4. *Let H be a maximal triangle-free graph. For every $\varepsilon > 0$ there exists $\delta > 0$ such that if G is a triangle-free graph which can be δ-approximated by H then there exists a graph G' and a homomorphism φ from G' to H^* with the following properties.*

(i) G is an ε-disturbed subgraph of G',

(ii) (H^*, ω_φ) is ε-balanced, where ω is the uniform weight function defined on G,

*(iii) φ is a **strong** homomorphism, that is, $uv \notin E(G')$ implies $\varphi(u)\varphi(v) \notin E(H^*)$ for every pair of vertices $u, v \in V(G')$.*

Proof. We assume that $V(H) = \{1, 2, \ldots, k\}$. We show that $\delta > 0$ satisfies the theorem if $(k+2)\sqrt{\delta} \leq \min(\varepsilon, 1/k)$. Let $\mathcal{V} = (V_1, V_2, \ldots, V_k)$ be the partition of $V(G)$, and the graph $H_\mathcal{V}$ be as in Definition 4.1. Let $n := \mathsf{v}(G)$, $F := E(G) \triangle E(\mathcal{H}_\mathcal{V})$ and J be the set of all vertices of G incident to at least $\sqrt{\delta}n$ edges in F. We have $\frac{1}{2}\sqrt{\delta}n|J| \leq |F| \leq \delta n^2$. It follows that $|J| \leq 2\sqrt{\delta}n$.

We define a map $\varphi : V(G) \to V(H^*)$, as follows. If $v \in V_i \setminus J$ for some $i \in V(H)$ then $\varphi(v) := i$. Now consider $v \in J$ and let

$$I_0(v) := \{i \in V(G) \mid |N(v) \cap V_i| > \sqrt{\delta}n\}.$$

Then $I_0(v)$ is independent, as otherwise there exist $i, j \in [k]$, such that $ij \in E(H)$, $|N(v) \cap V_i| > \sqrt{\delta}n$ and $|N(v) \cap V_j| > \sqrt{\delta}n$. As G is triangle-free it follows that $\mathsf{e}(H_\mathcal{V} \triangle G) > \varepsilon n^2$, contradicting the choice of $\mathcal{V}$. Let $I(v)$ be a maximal independent set containing $I_0(v)$, chosen arbitrarily. Define $\varphi(v) := i$, if $I(v) = N_H(i)$ for some $i \in V(H)$, and $\varphi(v) := I(v)$, otherwise.

Let G' be the graph with $V(G') = V(G)$ and the vertices $uv \in E(G')$ if and only if $\varphi(u)\varphi(v) \in E(H^*)$. Then φ is a strong homomorphism from G' to H^*. For $i \in V(H)$ we have $V_i \setminus J \subseteq \varphi^{-1}(i) \subseteq V_i \cup J$ and, therefore, $||\varphi^{-1}(i)|/n - 1/k| \leq \delta + 2\sqrt{\delta}$. For $I \in V(H^*) \setminus V(H)$ we have $|\varphi^{-1}(I)| \leq |J| \leq 2\sqrt{\delta}n$. Thus (ii) holds, as $\varepsilon \geq \delta + 2\sqrt{\delta}$.

It remains to verify (i). We will show that J is an ε-covering set for $F' := E(G) \setminus E(G')$ for δ sufficiently small. First, we show that every edge of F' has an end in J. Indeed suppose that $uv \in E(G)$ for some $u \in V_i, v \in V_j$ with $i, j \in V(H)$ not necessarily distinct and $ij \notin E(H)$. Then there exists $h \in V(H)$ adjacent to both i and j, as H is maximally triangle-free. It follows that both v and u have at least $(1/k - \delta - \sqrt{\delta})n$ neighbors in V_h and share a common neighbor if $2(1/k - \delta - \sqrt{\delta}) > 1/k$. Thus our first claim holds, as $\delta + \sqrt{\delta} < 1/k$.

Consider now $v \in J$ and let $N'(v)$ be the set of neighbors of v in $V(G) \setminus V(J)$ joined to v by edges of F'. Then $|N'(v) \cap V_i| \le \sqrt{\delta}n$ for every $i \in V(H)$ by the choice of $\varphi(v)$. It follows that $|N'(v)| \le k\sqrt{\delta}n$. Therefore v is incident to at most $(k+2)\sqrt{\delta}n$ edges in F, as $|J| \le 2\sqrt{\delta}n$. Thus J is an ε-covering set for F', as $(k+2)\sqrt{\delta}n \le \varepsilon$. $\quad\square$

We say that H is *entwined* if $\mathcal{I}^*(H)$ is intersecting. Note that, using Fact 2.3, one can see that F_d is entwined for every $d \ge 1$, as $\mathcal{I}^*(F_d)$ is empty. It is routine to check that the Petersen graph is entwined. We say that a graph G of order n is *c-maximal triangle-free* if it is triangle-free and adding any new edge to G creates at least cn triangles. For a weighted graph (G, ω), we say that it is *c-maximal triangle-free* if G is triangle-free and adding any new edge to G creates triangles of total weight at least c.

Lemma 4.5. *Let H be an entwined maximal triangle-free graph, and let $0 < \varepsilon < 1/v(H)$. If (H^*, ω) is ε-balanced then it is $(1/v(H) - \varepsilon)$-maximal triangle-free.*

Definition 4.6. *For a weighted graph (G, ω) we call a distribution $\mathbf{s}$ defined on the set of halves of (G, ω) a c-uniform sparse half for some $0 < c \le 1$, if*

1. *for every edge $e \in E(G)$, $\mathbb{E}\left[\mathbf{s}(e)\right] \ge c\,\omega(e)$,*

2. *$\mathbb{E}\left[\mathbf{s}(E(G))\right] \le \frac{1}{50}$.*

Whenever we refer to c-uniform sparse halves in unweighted graphs, they are understood as the c-uniform sparse halves in the corresponding uniformly weighted graphs.

Theorem 4.7. *Let $0 < c < 1$ be real, G' be a c-maximal triangle-free graph and G be a triangle-free $\frac{c^2}{2(1+c)}$-disturbed subgraph of G'. If G' has a c-uniform sparse half then G has a sparse half.*

Proof. Let $\varepsilon = \frac{c^2}{2(1+c)}$, $F := E(G) - E(G')$ and let M be a maximal matching in F. Suppose $|M| = \delta n$ for some $0 \le \delta \le 1/2$. Since G is an ε-disturbed subgraph of G', it has an ε-covering set for F. Let D be the minimum one. By the choice of M every edge of F has at least one end in $V(M)$. It follows that $|D| \le |V(M)| = 2|M|$. By the third condition in the definition of ε-disturbed subgraph, $|F| \le \varepsilon n|D| \le 2\delta\varepsilon n^2$.

Let $F' := E(G') \setminus E(G)$. For an edge $e \in M$, let $T(e)$ be the set of edges $e' \in F'$, such that the only vertex of $V(M)$ that e' is incident to, is an end of e. Let u, v be the ends of e. By the c-maximality of G', we have $|N_{G'}(u) \cap N_{G'}(v)| \ge cn$ for every pair of vertices $u, v \in V(G')$ non-adjacent in G'. Since $e \in E(G)$ and G is triangle-free, for every vertex $w \in N_{G'}(u) \cap N_{G'}(v)$, either $uw \in F'$ or $vw \in F'$. It follows that $|T_e| \ge |N_{G'}(u) \cap N_{G'}(v)| - |V(M)| \ge (c - 2\delta)n$. Thus $|F'| \ge (c - 2\delta)n \cdot \delta n$.

Let $\mathbf{s}$ be a c-uniform sparse half in the graph G'. Then

$$\mathbb{E}\left[\mathbf{s}(E(G)) - \mathbf{s}(E(G'))\right] = \mathbb{E}\left[\mathbf{s}(F) - \mathbf{s}(F')\right]$$
$$= \mathbb{E}\left[\mathbf{s}(F)\right] - \mathbb{E}\left[\mathbf{s}(F')\right],$$

by linearity of the expectation. We have

$$\mathbb{E}\left[\mathbf{s}(F')\right] = \sum_{e \in F'} \mathbb{E}\left[\mathbf{s}(e)\right] \ge \sum_{e \in F'} c\,\omega(e) = c\sum_{e \in F'} \frac{1}{n^2} \ge \frac{c|F'|}{n^2}.$$

On the other hand,

$$\mathbb{E}\left[\mathbf{s}(F)\right] = \sum_{e \in F} \mathbb{E}\left[\mathbf{s}(e)\right] \leq \sum_{e \in F} \omega(e) = \frac{|F|}{n^2}.$$

Finally, note that $\delta \leq \varepsilon$, since every edge in M has at least one of its ends in D. Hence,

$$\mathbb{E}\left[\mathbf{s}(E(G)) - \mathbf{s}(E(G'))\right] \leq \frac{|F|}{n^2} - \frac{c|F'|}{n^2} \leq 2\delta\varepsilon - c(c - 2\delta)\delta$$
$$= \delta(2\delta c + 2\varepsilon - c^2) \leq 0,$$

where the last inequality holds, as $2\delta c + 2\varepsilon - c^2 \leq 2(1 + c)\varepsilon - c^2 = 0$. Therefore, $\mathbb{E}\left[\mathbf{s}(E(G))\right] \leq \mathbb{E}\left[\mathbf{s}(E(G'))\right] \leq \frac{1}{50}$, and the graph G has a sparse half by Lemma 2.1. $\square$

We are now ready to prove the main result of this section.

Theorem 4.8. *Let H be an entwined maximal triangle-free graph. Suppose that there exists $\alpha > 0$ such that, if (H^*, ω) is α-balanced, then (H^*, ω) has an α-uniform sparse half. Then there exists $\delta > 0$ such that every triangle-free graph G which can be δ-approximated by H has a sparse half.*

Proof. Let $\varepsilon := \min\left(\frac{1}{3\mathsf{v}(H)^2}, \frac{\alpha^2}{2(1+\alpha)}\right)$, and let δ be chosen so that there exist φ and G' satisfying the conclusion of Theorem 4.4. The weighted graph (H^*, ω_φ) has an α-uniform sparse half, as $\varepsilon \leq \alpha$. Therefore, the graph G' has an α-uniform sparse half. By Lemma 4.5, the graph (H^*, ω_φ) is $(1/\mathsf{v}(H) - \varepsilon)$-triangle-free. Therefore, the graph G' is ε-triangle-free, because φ is a strong homomorphism. Let $c := \min(\alpha, 1/2\mathsf{v}(H))$. Then $\varepsilon \leq c^2/(2(1 + c))$ and $c \leq 1/\mathsf{v}(H) - \varepsilon$, by the choice of ε. It follows that the conditions of Theorem 4.7 are satisfied for G' and G. Thus G has a sparse half, as desired. $\square$

5. Triangle-free graphs with at least $(1/5 - \gamma)n^2$ edges

In order to establish the conjecture for the triangle-free graphs with average degree $\left(\frac{2}{5} - \gamma\right)n$ we separate the cases when the graphs under consideration are close in the sense of Definition 4.1 to the blowup of C_5 and when they are not. In the second case, we use the main result of Keevash and Sudakov from [8]. Note that this statement follows from their proof method but it is not explicitly stated as such in [8].

Theorem 5.1. *Let G be a triangle-free graph on n vertices such that one of the following conditions holds*

(a) *either $\frac{1}{n}\sum_{v \in V(G)} d^2(v) \geq \left(\frac{2}{5}n\right)^2$ and $\Delta(G) < \left(\frac{2}{5} + \frac{1}{135}\right)n$, or*

(b) *$\Delta(G) \geq \left(\frac{2}{5} + \frac{1}{135}\right)n$ and $\frac{1}{n}\sum_{v \in V(G)} d(v) \geq \left(\frac{2}{5} - \frac{1}{125}\right)n$.*

Then G has a sparse half.

The next theorem represents the main technical step in the proof of Theorem 1.2.

Theorem 5.2. *For every $\varepsilon > 0$ there exists $\delta > 0$ such that the following holds. If G is a triangle-free graph with $v(G) = n$ and $e(G) \geq (\frac{1}{5} - \delta)n^2$ then either*

(i) *G can be ε-approximated by C_5,*

(ii) *or at least δn vertices of G have degree at least $\left(\frac{2}{5} + \delta\right) n$,*

(iii) *or there exists an $F \subseteq E(G)$ with $|F| \leq \varepsilon n^2$ such that the graph $G - F$ is bipartite.*

The proof of Theorem 5.2 uses the following technical lemmas.

Lemma 5.3. *For $\delta > 0$ let G be a graph with $v(G) = n$ and $e(G) \geq (\frac{1}{5} - \delta)n^2$. Then either*

(1) *at least δn vertices have degree at least $(\frac{2}{5} + \delta)n$,*

(2) *or at most $2\sqrt{\delta}n$ vertices have degree at most $(\frac{2}{5} - 2\sqrt{\delta})n$.*

Proof. Suppose that the outcome (1) does not hold. Let tn be the number of vertices that have degree at most $(\frac{2}{5} - 2\sqrt{\delta})n$. Then

$$2\left(\frac{1}{5} - \delta\right)n^2 \leq 2e(G) = \sum_{v \in V} d(v)$$

$$= \sum_{\substack{v \in V \\ d(v) \leq (\frac{2}{5} - 2\sqrt{\delta})n}} d(v) + \sum_{\substack{v \in V \\ (\frac{2}{5} - 2\sqrt{\delta})n < d(v) < (\frac{2}{5} + \delta)n}} d(v) + \sum_{\substack{v \in V \\ d(v) \geq (\frac{2}{5} + \delta)n}} d(v)$$

$$< tn\left(\frac{2}{5} - 2\sqrt{\delta}\right)n + (1 - t)n\left(\frac{2}{5} + \delta\right)n + \delta n^2$$

$$= \left(\frac{2}{5} - \sqrt{\delta}t + 2\delta - t\delta\right)n^2.$$

Thus $-2\delta < 2\delta - 2t\sqrt{\delta} - \delta t$, and

$$t < \frac{4\delta}{2\sqrt{\delta} + \delta} < 2\sqrt{\delta},$$

as desired. □

Lemma 5.4. *Let H be a graph on n vertices with minimum degree at least $\left(\frac{2}{5} - \delta\right) n$. Let φ be a surjective homomorphism from H to C_5. Then*

$$\left(\frac{1}{5} - 3\delta\right) n \leq |\varphi^{-1}(v)| \leq \left(\frac{1}{5} + 2\delta\right) n, \tag{5}$$

for every $v \in V(C_5)$.

Proof. Let the vertices of C_5 be labelled $v_1, v_2, \ldots, v_5$ such that the neighbors of v_i are v_{i+2} and v_{i+3}. Define $V_i := \varphi^{-1}(v_i)$ for $i = 1, 2, \ldots, 5$. By the minimum degree condition on H, we have

$$|V_i| + |V_{i+1}| \geq \left(\frac{2}{5} - \delta\right) n \tag{6}$$

$$|V_{i+2}| + |V_{i+3}| \geq \left(\frac{2}{5} - \delta\right) n \tag{7}$$

$$|V_{i+4}| + |V_i| \geq \left(\frac{2}{5} - \delta\right) n, \tag{8}$$

therefore

$$|V_i| + \left(\frac{2}{5} - \delta\right) n + \left(\frac{2}{5} - \delta\right) n \leq |V_i| + (|V_{i+1}| + |V_{i+2}|) + (|V_{i+3}| + |V_{i+4}|) \leq n,$$

which gives us the desired upper bound.

For the lower bound summing inequalities (6)-(8) we get

$$n + |V_i| \geq \sum_{j=1}^{5} |V_j| + |V_i| \geq 3 \left(\frac{2}{5} - \delta\right) n,$$

which gives us $|V_i| \geq \left(\frac{1}{5} - 3\delta\right) n$, as desired. $\qquad\square$

Proof of Theorem 5.2. We show that $\delta := (\varepsilon/40)^2$ satisfies the theorem for $\varepsilon \leq 1$. We apply Lemma 5.3. The first outcome of Lemma 5.3 corresponds to the outcome (ii) of the theorem. Therefore we assume that the second outcome holds: There exists $|S| \leq \frac{\varepsilon}{20} n$ such that every vertex in $V(G) \setminus S$ has degree at least $\left(\frac{2}{5} - \frac{\varepsilon}{20}\right) n$ in G. It follows that $G' := G - S$ has the minimum degree at least $\left(\frac{2}{5} - \frac{\varepsilon}{10}\right) n$.

Since $\left(\frac{2}{5} - \frac{\varepsilon}{10}\right) n \geq \frac{3}{8} |V(G')|$, Theorem 2.5 implies that there exists a homomorphism φ from G' to C_5. Let $V(C_5) = \{v_1, v_2, \ldots, v_5\}$ and $V_i = \varphi^{-1}(v_i)$ for each $i = 1, 2, \ldots, 5$.

If the homomorphism φ is not surjective then by Lemma 2.7 the graph G' is bipartite and therefore the graph $G^* = (V(G), E(G'))$ is also bipartite. We have $e(G) - e(G') \leq |S|n \leq \varepsilon n^2$. Thus outcome (iii) holds. Hence we can suppose that the homomorphism φ is surjective. Applying Lemma 5.4 to the graph G' with $\delta = \frac{\varepsilon}{10}$, we get that

$$\left(\frac{1}{5} - \frac{2\varepsilon}{5}\right) n \leq |V_i| \leq \left(\frac{1}{5} + \frac{\varepsilon}{5}\right) n. \tag{9}$$

Let $\mathcal{V} := (V_1 \cup S, V_2, \ldots, V_5)$ be a partition of $V(G)$. From (9) we have $\left||V_i| - \frac{n}{5}\right| \leq \varepsilon n$. Let $H_{\mathcal{V}}$ be as in Definition 4.1. Then

$$e(G \triangle H_{\mathcal{V}}) \leq |S|n + e(H_{\mathcal{V}}) - e(G') \leq \frac{\varepsilon}{20} n^2 + \frac{n^2}{5} - \frac{1}{2} \left(\frac{2}{5} - \frac{\varepsilon}{10}\right) \left(1 - \frac{\varepsilon}{20}\right) n^2$$
$$\leq \varepsilon n^2.$$

Thus outcome (i) holds. $\qquad\square$

If outcome (i) of Theorem 5.2 holds, our goal is to apply Theorem 4.8. To do that we need to ensure that a c-balanced weighting of C_5 has a c-uniform sparse half for some $c > 0$.

Theorem 5.5. *Any $(1/50)$-balanced weighted graph (C_5, ω) has a $(1/30)$-uniform sparse half.*

Proof. Let $\delta := 1/50$, $V(C_5) = \{v_1, v_2, \ldots, v_5\}$ and $E(C_5) = \{v_i v_{i+2}\}_{i=1}^5$, as in the proof of Theorem 3.1. We define a distribution on the set of halves of the graph C_5. Recall the halves s_i, $1 \le i \le 5$ that we have defined earlier in the proof of the Theorem 3.1.

Let the probability mass of the distribution s be $\frac{1}{5}$ on every s_i, $i = 1, 2, \ldots, 5$. We show that s is a $\frac{1}{30}$-uniform sparse half. Let us begin by showing that $\mathbb{E}\left[\mathsf{s}(e)\right]) \ge \frac{1}{30} \omega(e)$ for every $e \in E(C_5)$. Let $e = (v_i, v_{i+2})$, then

$$\mathbb{E}\left[\mathsf{s}(e)\right] = \frac{1}{5} \cdot \omega(v_i) \left(\frac{1}{2} - (\omega(v_i) + \omega(v_{i+1}))\right).$$

Hence,

$$\mathbb{E}\left[\mathsf{s}(e)\right] \ge \frac{1}{5} \left(\frac{1}{5} - \delta\right) \left(\frac{1}{2} - 2\left(\frac{1}{5} + \delta\right)\right) = \frac{1}{5} \cdot \left(\frac{1}{5} - \delta\right) \left(\frac{1}{10} - 2\delta\right).$$

On the other hand,

$$\omega(e) = \omega(v_i) \cdot \omega(v_{i+2}) \le \left(\frac{1}{5} + \delta\right)^2.$$

Thus it suffices to show that

$$\frac{1}{5} \left(\frac{1}{5} - \delta\right) \left(\frac{1}{10} - 2\delta\right) \ge \frac{1}{30} \left(\frac{1}{5} + \delta\right)^2,$$

which can be easily verified. It is shown in the proof of Theorem 3.1 that $\mathbb{E}\left[\mathsf{s}(E(G))\right] \le \frac{1}{50}$. Thus, s is a $\frac{1}{30}$-uniform sparse half of G, as claimed. $\square$

We need a final technical lemma.

Lemma 5.6. *For every $\delta > 0$ there exists $\gamma > 0$ such that if G is a graph on n vertices with at least $\left(\frac{1}{5} - \gamma\right) n^2$ edges and at least δn vertices of degree at least $(\frac{2}{5} + \delta)n$ then*

$$\frac{1}{n} \sum_{v \in V(G)} d^2(v) \ge \left(\frac{2}{5}n\right)^2.$$

Proof. Suppose that G contains αn vertices of degree at most $\frac{2}{5}n$ and βn vertices of degree at least $\left(\frac{2}{5} + \delta\right) n$. We may assume that the average degree of G is less than $\frac{2}{5}n$, as otherwise lemma clearly holds. Thus,

$$\frac{2}{5}n > (1 - \alpha - \beta) \frac{2}{5}n + \beta \left(\frac{2}{5} + \delta\right)n,$$

hence $\alpha > \frac{5}{2}\beta\delta \geq \frac{5}{2}\delta^2$. We have

$$n \sum_{v \in V(G)} d^2(v) - \left(\sum_{v \in V(G)} d(v) \right)^2 = \frac{1}{2} \sum_{u \neq v} (d(u) - d(v))^2$$

$$\geq \alpha\delta \left(\frac{2}{5} + \delta - \frac{2}{5} \right)^2 n^4 > \frac{5}{2}\delta^5 n^4.$$

Hence

$$\frac{1}{n} \sum_{v \in V(G)} d^2(v) \geq \frac{1}{n^2} \left(\sum_{v \in V(G)} d(v) \right)^2 + \frac{5}{2}\delta^5 n^2 \geq 4 \left(\frac{1}{5} - \gamma \right)^2 n^2 + \frac{5}{2}\delta^5 n^2$$

$$\geq \frac{4}{25} n^2 + \left(\frac{5}{2}\delta^5 - \frac{8}{5}\gamma \right) n^2 \geq \frac{4}{25} n^2,$$

if we choose $\gamma = \frac{25}{16}\delta^5$. $\qquad\square$

Proof of Theorem 1.2. By Theorem 5.5, $\alpha = 1/50$ satisfies the conditions in the statement of Theorem 4.8 for $H := C_5$. Thus by Theorem 4.8 there exists $0 < \varepsilon \leq 1/50$ such that every triangle-free graph G that can be ε-approximated by C_5 has a sparse half. Let δ be such that Theorem 5.2 holds, and finally let γ be such that Lemma 5.6 holds. We show that Theorem 1.2 holds for this choice of γ.

We distinguish cases based on the outcome of Theorem 5.2 applied to G.

Case (i): If G can be ε-approximated by C_5 then the theorem holds by the choice of ε.

Case (ii): Now suppose that at least δn vertices of G have degree at least $\left(\frac{2}{5} + \delta \right) n$. If $\Delta(G) \geq \left(\frac{2}{5} + \frac{1}{135} \right) n$ then Theorem 5.1 (b) implies that there is a sparse half in G. Therefore we assume that $\Delta(G) < \left(\frac{2}{5} + \frac{1}{135} \right) n$. By Lemma 5.6 and the choice of γ we have $\frac{1}{n} \sum_{v \in V(G)} d^2(v) \geq \left(\frac{2}{5} n \right)^2$. Hence Theorem 5.1 (a) implies that there is a sparse half in G.

Case (iii): Lastly, suppose there exists an $F \subseteq E(G)$ with $|F| \leq \varepsilon n^2$ such that the graph $G' = (V(G), E(G) \setminus F)$ is bipartite with bipartition (U, V). Then either $|U| \geq \frac{n}{2}$ or $|V| \geq \frac{n}{2}$. Without loss of generality suppose $|U| \geq \frac{n}{2}$. The set U is independent in G', while in G it might not be, but $e(G[U]) \leq |F| \leq \varepsilon n^2 \leq \frac{n^2}{50}$. Hence U supports a sparse half in graph G. $\qquad\square$

6. Neighbourhood of the Petersen Graph

The uniform blowup of the Petersen graph P, is an extremal example for Conjecture 1.1, that is every set of $\lfloor n/2 \rfloor$ vertices spans at least $n^2/50$ edges. Here we show that the Conjecture 1.1 holds for any graph that is close to a uniform blowup of Petersen graph in the sense of Definition 4.1. By Theorem 4.8 it is enough to show that sufficiently balanced blowups of P^* (see Figure 4) have uniform sparse halves.

Lemma 6.1. *Let the weighted graph (P^*, ω) be $(1/500)$-balanced. Then it has a $\frac{1}{80}$-uniform sparse half.*
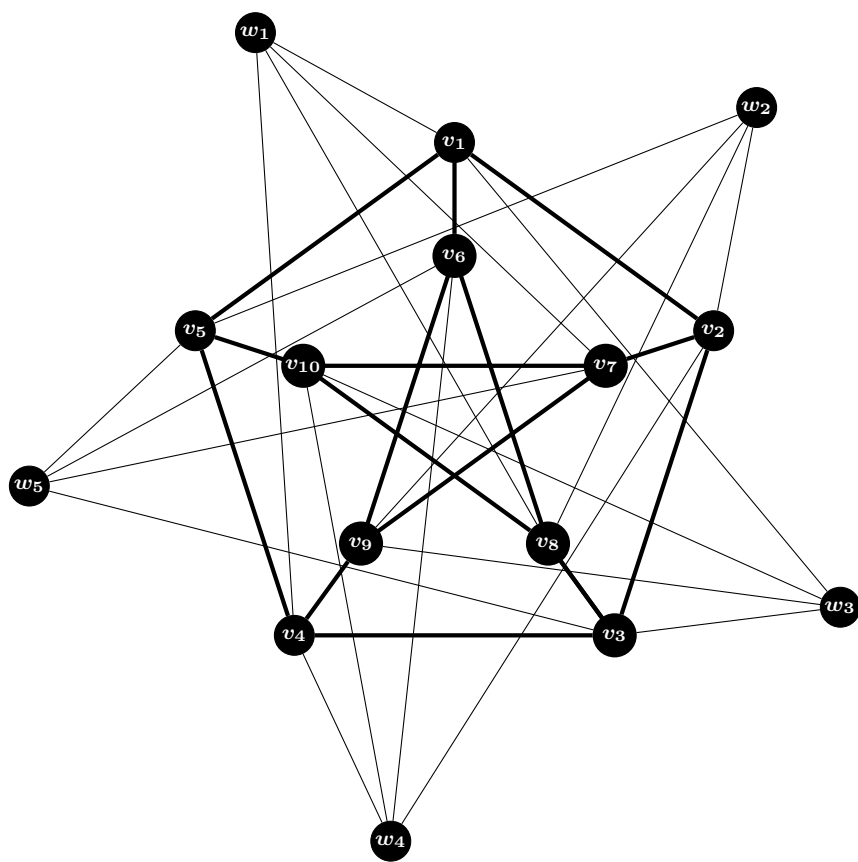
Figure 4: The graph P^*.

Proof. Let $\delta := 1/500$ and the vertices of P^* be labeled as in Figure 4, where $V :=$ $V(P) = \{v_1, v_2, \ldots, v_{10}\}$ and $W := V(P^*) \setminus V(P) = \{w_1, w_2, \ldots, w_5\}$. We define a collection $\{\mathsf{s}_{i,j}\}_{i\in[5],j\in[4]}$ of halves of (P^*, ω). Fix $i \in [5]$ and consider the vertex w_i. Let $M_i := V \setminus N(w_i)$ and not that M_i induces a matching of size three. Choose vertices $\{v_{i_1}, v_{i_2}, v_{i_3}\} \subseteq V(M_i)$ such that they are independent and there exist unique $w_{i_j} \neq w_i$ such that every $v \in V(M_i) \setminus \{v_{i_1}, v_{i_2}, v_{i_3}\}$ is adjacent to w_{i_j}. Note that for every i there exist four such choices of $\{v_{i_1}, v_{i_2}, v_{i_3}\}$, fix one of them and assign

$$\mathsf{s}_{i,j}(w_q) = \begin{cases} \omega(w_q), & \text{if } q = i, \\ \frac{1}{4}\omega(w_q), & \text{if } q = i_j, \\ 0, & \text{otherwise.} \end{cases}$$

and

$$\mathsf{s}_{i,j}(v_k) = \begin{cases} 0, & \text{if } v_k \notin V(M_i) \\ \omega(v_k), & \text{if } k = j_1, j_2, j_3, \\ \frac{1}{3}\left(\frac{1}{2} - (\omega(v_{i_1}) + \omega(v_{i_2}) + \omega(v_{i_3}) + \omega(w_i) + \frac{1}{4}\omega(w_{i_j}))\right), & \text{otherwise.} \end{cases}$$

It is easy to check that every $\mathsf{s}_{i,j}$ is a half. Let s be a distribution concentrated on $\{\mathsf{s}_{i,j}\}_{i\in[5],j\in[4]}$ with each of the halves having the same probability $1/20$. We show that s is a $(1/80)$-uniform sparse half for ω.

First, we show that $\mathbb{E}\left[\mathsf{s}(e)\right] \geq \frac{1}{80} \cdot \omega(e)$ for every edge $e \in E(P^*)$. It can be routinely checked that for our choice of δ one has

$$\frac{1}{3}\left(\frac{1}{2} - (\omega(v_{i_1}) + \omega(v_{i_2}) + \omega(v_{i_3}) + \omega(w_i) + \frac{1}{4}\omega(w_{i_j}))\right) \geq \frac{1}{3}\omega(v_k), \qquad (10)$$

for every $v_k \in V(M_i) \setminus \{v_{i_1}, v_{i_2}, v_{i_3}\}$. If both ends v, v' of $e \in E(P^*)$ lie in V then, using (10), we have

$$\mathbb{E}\left[\mathsf{s}(e)\right] \geq \frac{4}{20} \cdot \frac{1}{3}\omega(v)\omega(v') = \frac{1}{15}\omega(e) \geq \frac{1}{80}\omega(e).$$

If e joins $v \in V$ and $w \in W$ then

$$\mathbb{E}\left[\mathsf{s}(e)\right] \geq \frac{3}{20} \cdot \frac{1}{4}\omega(w_i) \cdot \frac{1}{3}\omega(v_j) = \frac{1}{80}\omega(e)$$

It remains to prove that $\mathbb{E}\left[\mathsf{s}(E(G))\right] \leq \frac{1}{50}$. Note that,

$$\mathsf{s}_{i,j}(E(G)) = \left(\frac{1}{2} - (\omega(v_{i_1}) + \omega(v_{i_2}) + \omega(v_{i_3}) + \omega(w_i) + \frac{1}{4}\omega(w_{i_j}))\right) \times$$
$$\times \left(\frac{1}{4}\omega(w_{i_j}) + \frac{1}{3}\left(\omega(v_{i_1}) + \omega(v_{i_2}) + \omega(v_{i_3})\right)\right).$$

We finish the proof using the following technical lemma, the proof of which is included in the appendix.

Lemma 6.2. *Suppose given are $x_1, x_2, \ldots, x_{10}, y_1, y_2, \ldots y_5$ reals and*

$$L(y_i) = \{x_{i+1}, x_{i+2}, x_{i+4}, x_{i+5}, x_{i+8}, x_{i+9}\},$$

for each $1 \leq i \leq 5$ such that $0 \leq x_i \leq 1$, $0 \leq y_j \leq 1$ and $\sum_{i=1}^{10} x_i + \sum_{j=1}^{5} y_j = 1$. If there exists some $0 < \delta \leq \frac{1}{90}$ such that $x_i \geq \frac{1}{10} - \delta$ for each $i = 1, 2, \ldots, 10$ then

$$\sum_{\substack{i \neq j \\ x_{i,j_1}, x_{i,j_2}, x_{i,j_3} \\ \in L(y_i) \cap L(y_j)}} \left(\frac{1}{2} - x_{i,j_1} - x_{i,j_2} - x_{i,j_3} - y_i - \frac{1}{4}y_j \right) \left(\frac{1}{4}y_j + \frac{1}{3} \left(x_{i,j_1} + x_{i,j_2} + x_{i,j_3} \right) \right) \leq \frac{2}{5}.$$

$$(11)$$

It is easy to see that $\mathbb{E}\left[\mathbf{s}(E(G))\right] \leq \frac{1}{50}$. follows from Lemma 6.2 applied with $x_i := \omega(v_i)$ and $y_j := \omega(w_j)$. Thus $\mathbf{s}$ is a $1/80$-uniform sparse half, as claimed. $\square$

As promised, Lemma 6.1 implies the main theorem of this section.

Theorem 6.3. *There exists $\delta > 0$ such that any triangle-free graph G on n vertices which can be δ-approximated by the Petersen graph has a sparse half.*

Proof. The theorem follows from Theorem 4.8, as the Petersen graph satisfies the requirements of that theorem with $\alpha = 1/500$ by Lemma 6.1. $\square$

References

[1] B. Andrásfai, P. Erdős, and V. T. Sos, *On the connection between chromatic number, maximal clique and minimal degree of a graph*, Discrete Math. (1974), no. 8, 205–218.

[2] C.C. Chen, G.P. Jin, and K.M. Koh, *Triangle-free graphs with large degree*, Combinatorics, Probability and Computing (1997), no. 6, 381–396.

[3] P. Erdős, *Problems and results in graph theory and combinatorial analysis*, Proceedings of the Fifth British Combinatorial Conference (1975), no. 15(1976), 169–192.

[4] P. Erdős, *Some old and new problems in various branches of combinatorics*, Graphs and Combinatorics, Discrete Math (1995,1997), 165/166 , 227–231.

[5] P. Erdős, R. Faudree, C. Rousseau, and R. Schelp, *A local density condition for triangles*, Discrete Math (1994), no. 127, 153–161.

[6] R. Häggkvist, *Odd cycles of specified length in non-bipartite graphs*, Annals of Discrete Mathematics **13** (1982), 89–100.

[7] G.P. Jin, *Triangle-free chromatic grpahs*, Discrete Mathematics (1995), no. 145, 151–170.

[8] P. Keevash and B. Sudakov, *Sparse halves in triangle-free graphs*, Journal of Combinatorial Theory (2006), no. 96, 614–620.

[9] M. Krivelevich, *On the edge distribution in triangle-free graphs*, J. Comb. Theory, Ser. B **63** (1995), no. 2, 245–260.

[10] L. Lovász, *Subgraph densities in signed graphons and the local Simonovits-Sidorenko conjecture*, The Electronic Journal of Combinatorics [electronic only] **18** (2011), no. 1, Research Paper P127, 21 p., electronic only (eng).

[11] L. Lovász and B. Szegedy, *Testing properties of graphs and functions*, Israel Journal of Mathematics (2010), no. 178, 113–156.

[12] A. A. Razborov, *On the Caccetta-Häggkvist Conjecture with Forbidden Subgraphs*, J. Graph Theory **74** (2013), 236–248.

[13] A.A. Razborov, *On Turán's (3,4)-problem with forbidden subgraphs*, Mathematical Notes **95** (2014), no. 1-2, 245–252.

[14] D.R. Woodall, *The binding number of a graph and its Anderson number*, J. Combin. Theory (1973), 225–255.

7. Appendix

Proof of Lemma 3.3

Suppose the Lemma is false. Then for each $i = 1, 2, \ldots, 8$

$$\frac{1}{2}\left(\frac{1}{2} - x_i - x_{i+1} - x_{i+2}\right)(x_i + x_{i+2}) + \frac{1}{4}\left(\frac{1}{2} - x_i - x_{i+1} - x_{i+2}\right)^2 > \frac{1}{50}. \quad (12)$$

Summing up these inequalities over all $i = 1, 2, \ldots, 8$ we get that

$$\frac{8}{50} < \sum_{i=1}^{8} \frac{1}{2}\left(\frac{1}{2} - x_i - x_{i+1} - x_{i+2}\right)(x_i + x_{i+2}) + \sum_{i=1}^{8} \frac{1}{4}\left(\frac{1}{2} - x_i - x_{i+1} - x_{i+2}\right)^2$$

$$= \frac{1}{2}\sum_{i=1}^{8} x_i - \sum_{i=1}^{8}(x_i + x_{i+1} + x_{i+2})x_i + \frac{1}{4}\cdot\frac{1}{4}\cdot 8 - \frac{1}{4}\sum_{i=1}^{8}(x_i + x_{i+1} + x_{i+2})$$

$$+ \frac{1}{4}\sum_{i=1}^{8}(x_i + x_{i+1} + x_{i+2})^2$$

$$= \frac{1}{2} - \sum_{i=1}^{8}(x_i + x_{i+1} + x_{i+2})x_i + \frac{1}{2} - \frac{3}{4} + \frac{1}{4}\sum_{i=1}^{8}(x_i + x_{i+1} + x_{i+2})^2$$

$$= \frac{1}{4} - \frac{1}{4}\sum_{i=1}^{8} x_i^2 - \frac{1}{2}\sum_{i=1}^{8} x_i x_{i+2}. \quad (13)$$

Let us find the maximum value under the conditions of the lemma. To find the maximum value of the expression in (13), we need to find the minimum value of

$$S := \frac{1}{4} \sum_{i=1}^{8} {x_i}^2 + \frac{1}{2} \sum_{i=1}^{8} x_i x_{i+2}.$$

Claim 7.1. *For every* $1 \le i \le 8$

$$x_{i+1} + x_{i+2} + x_{i+3} < 0.394$$

Proof. By inequality (12)

$$\frac{1}{2}\left(\frac{1}{2} - (x_{i+1} + x_{i+2} + x_{i+3})\right)(x_{i+1} + x_{i+3}) + \frac{1}{4}\left(\frac{1}{2} - (x_{i+1} + x_{i+2} + x_{i+3})\right)^2 > \frac{1}{50}.$$

Let $\alpha = x_{i+1} + x_{i+2} + x_{i+3}$. Then $x_{i+1} + x_{i+3} = \alpha - x_{i+2} \le \alpha - \frac{1}{14}$. Therefore

$$\frac{1}{2}\left(\frac{1}{2} - \alpha\right)\left(\alpha - \frac{1}{14}\right) + \frac{1}{4}\left(\frac{1}{2} - \alpha\right)^2 = -\frac{\alpha^2}{4} + \frac{\alpha}{28} + \frac{5}{112} > \frac{1}{50},$$

which reduces to the inequality

$$\frac{\alpha^2}{4} - \frac{\alpha}{28} - \frac{69}{2800} < 0.$$

This quadratic inequality gives us the desired $\alpha < 0.393412$ bound. $\square$

It is easy to check that the following claim is true.

Claim 7.2.

$$S \ge \frac{1}{2} \sum_{i=1}^{4} {z_i}^2 + \sum_{i=1}^{4} z_i z_{i+2},$$

where $z_i = (x_i + x_{i+4})/2$ *for all* $1 \le i \le 4$.

Claim 7.3. *For every* $1 \le i \le 4$, $z_i > 0.106$.

Proof. For every $1 \le i \le 4$ we have

$$1 = (x_i + x_{i+4}) + (x_{i+1} + x_{i+2} + x_{i+3}) + (x_{i+5} + x_{i+6} + x_{i+7}) \overset{(7.1)}{<} 2z_i + 2 \cdot 0.394,$$

therefore $z_i > 0.106$. $\square$

Let $\beta = z_1 + z_3$. Then

$$S \overset{(7.2)}{\ge} \frac{1}{2}(z_1 + z_3)^2 + \frac{1}{2}(z_2 + z_4)^2 + z_1 z_3 + z_2 z_4$$

$$= \frac{1}{2}\beta^2 + \frac{1}{2}\left(\frac{1}{2} - \beta\right)^2 + z_1 z_3 + z_2 z_4$$

$$\overset{(7.3)}{\ge} \frac{1}{2}\beta^2 + \frac{1}{2}\left(\frac{1}{2} - \beta\right)^2 + 0.106 \cdot (\beta - 0.106) + 0.106\left(\frac{1}{2} - \beta - 0.106\right)$$

$$= \beta^2 - \frac{1}{2}\beta + 0.155528$$

$$> 0.093$$

The last two inequalities hold because for fixed β the expression $z_1 z_3 + z_2 z_4$ achieves its minimum value when $z_1 = z_2 = 0.106$, $z_3 = \beta - 0.106$ and $z_4 = \frac{1}{2} - \beta - 0.106$. The expression $\beta^2 - \frac{1}{2}\beta + 0.155528$ achieves its minimum for $\beta = \frac{1}{4}$. It follows from the inequality above that

$$\frac{1}{4} - \frac{1}{4}\sum_{i=1}^{8}{x_i}^2 - \frac{1}{2}\sum_{i=1}^{8}x_i x_{i+2} = \frac{1}{4} - S < \frac{8}{50},$$

a contradiction that finishes the proof of the lemma.

Proof of Lemma 3.4

Suppose the lemma is false. Then for all $1 \leq i \leq 11$ we have

$$\frac{1}{2}\left(\frac{1}{2} - (x_{i+1} + x_{i+2} + x_{i+3} + x_{i+4})\right)(x_{i+1} + x_{i+4})$$
$$+ \frac{1}{4}\left(\frac{1}{2} - (x_{i+1} + x_{i+2} + x_{i+3} + x_{i+4})\right)^2 > 1/50.$$

Summing these inequalities for $1 \leq i \leq 11$ we obtain

$$\frac{11}{50} < \sum_{i=1}^{11}\frac{1}{2}\left(\frac{1}{2} - (x_{i+1} + x_{i+2} + x_{i+3} + x_{i+4})\right)(x_{i+1} + x_{i+4})$$
$$+ \sum_{i=1}^{11}\frac{1}{4}\left(\frac{1}{2} - (x_{i+1} + x_{i+2} + x_{i+3} + x_{i+4})\right)^2$$
$$= \frac{1}{4}\sum_{i=1}^{11}(x_{i+1} + x_{i+4}) - \sum_{i=1}^{11}(x_{i+1} + x_{i+2} + x_{i+3} + x_{i+4})x_{i+1}$$
$$+ \frac{11}{16} - \frac{1}{4}\sum_{i=1}^{11}(x_{i+1} + x_{i+2} + x_{i+3} + x_{i+4}) + \frac{1}{4}\sum_{i=1}^{11}(x_{i+1} + x_{i+2} + x_{i+3} + x_{i+4})^2$$
$$= \frac{3}{16} - \sum_{i=1}^{11}(x_{i+1} + x_{i+2} + x_{i+3} + x_{i+4})x_{i+1} + \frac{1}{4}\sum_{i=1}^{11}(x_{i+1} + x_{i+2} + x_{i+3} + x_{i+4})^2$$
$$= \frac{3}{16} + \frac{1}{2}\sum_{i=1}^{11}x_i(x_{i+1} - x_{i+3})$$
$$\leq \frac{3}{16} + \frac{1}{2}\left(\sum_{i=1}^{11}x_i x_{i+1} - \frac{11}{196}\right). \tag{14}$$

We claim that

$$f(x_1, \ldots, x_{11}) := \sum_{i=1}^{11}x_i x_{i+1} \leq \frac{77}{784}.$$

Indeed, consider any pair (x_i, x_j), such that $j \neq i \pm 1$, let $\alpha = x_i + x_j$ is fixed. Note that f is linear as a function of (x_i, x_j). Therefore $f(x_i, x_j)$ achieves its maximum value

22

on the region $R := \{0 \le x_i \le 1/14 \text{ for } 1 \le 1 \le 14, \sum_{i=1}^{11} x_i = 1\}$, when $x_i = \frac{1}{14}$ and $x_j = \alpha - \frac{1}{14}$, or when $x_j = \frac{1}{14}$ and $x_i = \alpha - \frac{1}{14}$. Thus f attains its maximum on R when all variables are equal $\frac{1}{14}$ except possibly two of them whose indices are consecutive, without loss of generality, say x_{10} and x_{11}. It is easy to see that the maximum is achieved for $x_{10} = x_{11} = \frac{5}{28}$ and is equal to $\frac{77}{784}$, as claimed.

Thus (14) implies

$$\frac{11}{50} < \frac{3}{16} + \frac{1}{2}\left(\sum_{i=1}^{11} x_i x_{i+1} - \frac{11}{196}\right) \le \frac{3}{16} + \frac{1}{2}\cdot\left(\frac{77}{784} - \frac{11}{196}\right) = \frac{327}{1568} < \frac{11}{50},$$

a contradiction that finishes the proof.

Proof of Lemma 6.2

Let $Y := \sum_{i=1}^{5} y_i$. Summing inequalities (11) over all i, j such that $i \ne j$ and $x_{i,j_1}, x_{i,j_2}, x_{i,j_3} \in L(y_i) \cap L(y_j)$. We get

$$\sum_{i \ne j}\left(\frac{1}{2} - (x_{i,j_1} + x_{i,j_2} + x_{i,j_3} + y_i + \frac{1}{4}y_j)\right)\left(\frac{1}{4}y_j + \frac{1}{3}(x_{i,j_1} + x_{i,j_2} + x_{i,j_3})\right)$$

$$= \frac{1}{8}\sum_{i \ne j} y_j - \frac{1}{4}\sum_{i \ne j} y_j(x_{i,j_1} + x_{i,j_2} + x_{i,j_3}) + \frac{1}{6}\sum_{i \ne j}(x_{i,j_1} + x_{i,j_2} + x_{i,j_3})$$

$$- \frac{1}{3}\sum_{i \ne j}(x_{i,j_1} + x_{i,j_2} + x_{i,j_3})^2 - \frac{1}{3}\sum_{i \ne j}\left(y_i + \frac{1}{4}y_j\right)(x_{i,j_1} + x_{i,j_2} + x_{i,j_3}) - \frac{1}{4}\sum_{i \ne j} y_j\left(y_i + \frac{1}{4}y_j\right)$$

$$\le \frac{1}{2}Y - 3Y\left(\frac{1}{10} - \delta\right) + (1 - Y) - \frac{36(1-Y)^2}{60} - 5Y\left(\frac{1}{10} - \delta\right)$$

$$= \frac{2}{5} - \frac{Y}{10} - \frac{3}{5}Y^2 + 8\delta Y$$

$$\le \frac{2}{5},$$

since $\delta \le \frac{1}{90}$.