A Turán-type problem on distances in graphs

Mykhaylo Tyomkyn\textsuperscript{a}, Andrew Uzzell\textsuperscript{b}\textsuperscript{*}

\textsuperscript{a}University of Cambridge, Department of Pure Mathematics and Mathematical Statistics
Centre for Mathematical Sciences, Wilberforce Road
Cambridge, CB3 0WB, England

\textsuperscript{b}Department of Mathematics, Uppsala University
PO Box 480, 751 06, Uppsala, Sweden

August 9, 2012

Abstract

We suggest a new type of problem about distances in graphs and make several conjectures. As a first step towards proving them, we show that for sufficiently large values of \(n\) and \(k\), a graph on \(n\) vertices that has no three vertices pairwise at distance \(k\) has at most \((n - k + 1)^2/4\) pairs of vertices at distance \(k\).

1 Introduction

In \cite{8}, Bollobás and Tyomkyn determined the maximum number of paths of length \(k\) in a tree \(T\) on \(n\) vertices. Here we suggest an extension of this problem to general graphs.

The ‘obvious’ extension, counting paths of a given length in a graph \(G\), has been studied since 1971, see, e.g., \cite{1, 3, 4, 5, 6, 7, 9, 11, 12} and the references therein. On the other hand, counting paths of length \(k\) in trees can be interpreted as counting pairs of vertices at distance \(k\). Therefore, a natural question to ask is the following.

Question. For a graph \(G\) on \(n\) vertices, what is the maximum possible number of pairs of vertices at distance \(k\)?

To the best of our knowledge, this question has not been considered previously. Our aim in this paper is to formulate several conjectures and to prove one of them in the first non-trivial special case.

For a graph \(G\), define the distance-\(k\) graph \(G_k\) to be the graph with vertex set \(V(G)\) and \(\{x, y\} \in E(G_k)\) if and only if \(x\) and \(y\) are at distance \(k\) in \(G\), that is, the shortest path between \(x\) and \(y\) has length \(k\). We call such vertices \(x\) and \(y\) \(k\)-neighbours and the pair \(\{x, y\}\) a \(k\)-distance. We call \(d_{G_k}(x)\) the \(k\)-degree of \(x\).

\textsuperscript{*}Research supported in part by NSF grant DMS-0505550
Figure 1: A 5-broom for $k = 8$ and a 5-broom for $k = 7$.

Observe that if $H$ is an induced subgraph of $G$, then $H_2$ is a subgraph of $G_2$. This need not be the case when $k \geq 3$. It is clear that $G_k \cong H_k$ does not imply that $G \cong H$. If $G_k \cong H_k$, then we say that $G$ is $k$-isomorphic to $H$.

We wish to maximise the number of edges in $G_k$ over all graphs $G$ on $n$ vertices. One attempt to construct a graph with many $k$-neighbours would be to consider what we call $t$-brooms. For even $k \geq 4$ and for $t \geq 2$, define a $t$-broom to be a graph consisting of a central vertex $v$ with $t$ ‘brooms’ attached, each consisting of a path on $(k - 2)/2$ vertices with leaves attached to the ends opposite $v$. In this way, the leaves of different brooms will be at distance $k$. For odd $k \geq 3$, to define a $t$-broom, take a copy of $K_t$ and attach a broom to each vertex, adjusting the length of the path. (See Figure 1.) As in Turán’s theorem, the number of $k$-distances in a $t$-broom will be maximised when the numbers of leaves in the brooms are as equal as possible.

In [8] Bollobás and Tyomkyn proved that if $G$ is a tree, then $e(G_k)$ is maximal when $G$ is a $t$-broom for some $t$.

**Theorem 1.** Let $n \geq k$. If $G$ is a tree on $n$ vertices, then $e(G_k)$ is maximal when $G$ is a $t$-broom. If $k$ is odd, then $t = 2$. If $k$ is even, then $t$ is within 1 of

$$\frac{1}{4} + \sqrt{\frac{1}{16} + \frac{n - 1}{k - 2}}.$$

These results prompt us to make the following conjecture.

**Conjecture 2.** Let $k \geq 3$. There exists $h = h(k)$ such that if $n \geq h(k)$, then $e(G_k)$ is maximised over all $G$ with $|G| = n$ when $G$ is $k$-isomorphic to a $t$-broom for some $t$.

For small values of $n$ there exist better constructions. For example, if $k = 3$ and $n = 7$, the 7-cycle has more 3-distances than any $t$-broom.

We firmly believe Conjecture 2 to be true, but are unable to prove it. In this paper, we approach Conjecture 2 by placing a restriction on $\omega(G_k)$, the clique number of $G_k$, which is the maximal number of vertices at pairwise distance $k$. We formulate the following natural analogue of Conjecture 2 under this condition.

**Conjecture 3.** Let $k \geq 3$ and $t \geq 2$. There is a function $h_2: \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that if $n \geq h_2(k, t)$, then $e(G_k)$ is maximised over all $G$ with $|G| = n$ and $\omega(G_k) \leq t$ when $G$ is $k$-isomorphic to a $t$-broom for some $t$. 

2
In this paper, we shall discuss the case $t = 2$ of Conjecture 3 that is, the case when no three vertices of $G$ are pairwise at distance $k$. Note that in this case the parity of $k$ matters little, as the conjectured optimal example is just a path of length $k - 2$ with equally many leaves attached to each of its endvertices. The number of $k$-distances in such a graph is simply $\lfloor \frac{(n - k + 1)^2}{4} \rfloor$.

We prove Conjecture 3 for $t = 2$ and sufficiently large $k$. More precisely, we prove the following assertion.

**Theorem 4.** There is a constant $k_0$ and a function $n_0: \mathbb{N} \to \mathbb{N}$ such that for all $k \geq k_0$, all $n \geq n_0(k)$ and all graphs $G$ of order $n$ with no three vertices pairwise at distance $k$,

$$e(G_k) \leq \frac{(n - k + 1)^2}{4}.$$ 

Moreover, if equality holds, then $G$ is $k$-isomorphic to the double broom.

We do not make any effort to determine $k_0$ and $n_0$ exactly; on the contrary, we are rather generous about them. However, we conjecture that $k = 2$ and $(n, k) = (7, 3)$ are the only exceptions to the optimality of the double-broom.

**Conjecture 5.** In the setting of Theorem 4 we can take $k_0 = 3$, $n_0(3) = 8$ and $n_0(k) = k + 1$ otherwise.

For $k = 2$, we can do a little better than the bound in Theorem 4 as the following construction shows.

**Example 6.** Let $X$ and $Y$ be cliques on $(n + 1)/2$ vertices each, with a vertex $z$ in common. Take vertices $x \in X$ and $y \in Y$. Remove the edges $\{x, z\}$ and $\{y, z\}$ and add the edge $\{x, y\}$; call the resulting graph $G$. Then $G_2$, the distance-2 graph of $G$, is a complete bipartite graph with one edge subdivided, and thus has a total of $(n - 1)^2/4 + 1$ edges.

We believe that for $n \geq 5$, a triangle-free $G_2$ can have no more than $(n - 1)^2/4 + 1$ edges. This is clearly true for $n = 5$, and a computer search verifies that it also holds for $6 \leq n \leq 11$. However, we cannot prove that it is true in general. If this is indeed so, then it shows that for $k = 2$ the quantity $(n - k + 1)^2/4$ is within 1 of the maximum. The same would hold for $k = 3$ and $n = 7$, when the aforementioned 7-cycle wins by 1 over the double broom.

In the general case, with no restriction on $\omega(G_2)$, the maximum value of $e(G_2)$ is straightforward.

**Proposition 7.** Let $G$ be a graph on $n$ vertices. Then

$$e(G_2) \leq \binom{n - 1}{2}.$$ 

Moreover, if equality holds, then $G$ is a star.

**Proof.** The result is clear for $n = 3$. We suppose that for some $n > 3$, the result holds for all graphs on at most $n - 1$ vertices. Let $G$ be a graph on $n$ vertices. We
may assume that $G$ is connected. Otherwise, if $G$ has $\ell \geq 2$ components of sizes $c_1, \ldots, c_\ell$, say, then, by hypothesis,

$$e(G_2) \leq \sum_{i=1}^\ell \binom{c_i - 1}{2} < \binom{n - 1}{2}.$$  

Observe that, setting $d = \text{diam}(G)$,

$$\binom{n}{2} = e(G) + e(G_2) + \cdots + e(G_d).$$

Thus,

$$e(G_2) \leq \binom{n}{2} - e(G). \quad (1.1)$$

Since $G$ is connected, $e(G) \geq n - 1$, and hence $e(G_2) \leq \binom{n - 1}{2}$. Note that if equality holds, then $G$ must be a tree. Moreover, in this case, by (1.1), $\text{diam}(G) = 2$, so $G$ must be a star. \hfill \Box

2 Preliminaries

In this section, we shall prove a straightforward bound on $e(G_k)$. We shall then discuss possible ways of extending this result to the upper bound in Theorem 4. We shall also discuss a useful property of spanning trees.

Since $G_k$ is triangle-free by our assumption, Mantel’s theorem implies that $e(G_k) \leq n^2/4$. In fact, we can do somewhat better by adapting a standard proof of Mantel’s Theorem (see, e.g., [2]) to $k$-distances.

Lemma 8. If $G_k$ is triangle-free, then

$$e(G_k) \leq \frac{n(n-k+1)}{4}.$$  

Proof. For a vertex $x$, let $\nu(x)$ be the number of $k$-neighbours of $x$, or equivalently, the degree of $x$ in $G_k$. We may bound $e(G_k)$ as follows: for each pair of vertices $x, y$ at distance $k$, count the $k$-neighbours of $x$ and of $y$. Note that since $x$ and $y$ have no common $k$-neighbours (otherwise there would be a triangle in $G_k$), we have $\nu(x) + \nu(y) \leq n$. In fact, we can claim that $\nu(x) + \nu(y) \leq n - k + 1$, since none of the $k-1$ internal vertices on the shortest path between $x$ and $y$ are $k$-neighbours of $x$ or of $y$. Summing over all such pairs $\{x, y\}$, we obtain

$$(n - k + 1)e(G_k) \geq \sum_{\{x, y\} \in E(G_k)} (\nu(x) + \nu(y)). \quad (2.1)$$

Observe that for each $x \in V(G)$, the quantity $\nu(x)$ appears $\nu(x)$ times on the right-hand side of (2.1). By the Cauchy-Schwarz inequality, we have

$$\sum_{\{x, y\} \in E(G_k)} (\nu(x) + \nu(y)) = \sum_{x \in V(G)} \nu(x)^2 \geq \frac{1}{n} \left( \sum_{x \in V(G)} \nu(x) \right)^2 = \frac{4}{n} (e(G_k))^2. \quad (2.2)$$
It follows from (2.1) and (2.2) that
\[ e(G_k) \leq \frac{n(n - k + 1)}{4}, \] (2.3)
as claimed.

The bound that we have just proved is about halfway between the trivial \( n^2/4 \) and the desired \((n - k + 1)^2/4\). There are two natural ways in which one could try to improve (2.3). One is to try to find vertices that have no \( k \)-neighbours at all. We say that such a vertex is an interior vertex; otherwise, a vertex is called an exterior vertex. The inspiration for this terminology is as follows. If \( k \leq \text{diam}(G) \leq 2k - 1 \), then a vertex at or near the centre of the graph has no \( k \)-neighbours, while a vertex that is far from the centre of the graph will have one or more \( k \)-neighbours. If \( G \) has at least \( r \) interior vertices, then adapting the proof of Lemma 8 improves the bound in (2.3) to
\[ e(G_k) \leq \frac{(n - r)(n - k + 1)}{4}. \] (2.4)
Thus, we are done if we can find at least \( k - 1 \) interior vertices (which holds in the case when \( G \) is the double-broom). So, we may assume that \( r < k - 1 \).

The other way to improve (2.3) would be to find many pairs of vertices \( \{u, v\} \) such that many vertices \( z \) are \( k \)-neighbours of neither \( u \) nor \( v \). We say that a vertex \( v \in V \) is \textit{k-unaffiliated} with a vertex \( u \in V \) if \( d(u, v) \neq k \). A vertex \( v \) is \textit{k-unaffiliated} with a set \( U \subseteq V \) if it is \( k \)-unaffiliated with each \( u \in U \). Otherwise, we say that \( v \) is \( k \)-affiliated with \( U \). Thus, an interior vertex is \( k \)-unaffiliated with \( V \). If \( G \) has \( r \) interior vertices and each \( \{u, v\} \in E(G_k) \) has \( p \) \( k \)-unaffiliated vertices, then (2.3) improves to
\[ e(G_k) \leq \frac{(n - r)(n - p)}{4} \leq \left( \frac{n - r + p}{2} \right)^2. \] (2.5)
In particular, we are done if \( p \geq 2k - r - 2 \), i.e., if every \( \{u, v\} \in E(G_k) \) has at least \( k - r - 1 \) \( k \)-unaffiliated vertices other than those on the shortest path between \( u \) and \( v \).

In order to prove Theorem 4 we shall show that for every pair of \( k \)-neighbours in \( G \), \( p \geq 2k - r - 2 \). To do so, we shall need to study paths in \( G \). In connection with this, we shall need the following easy result about the lengths of paths in a spanning tree. If \( P \) is a path, we write \( |P| \) to denote the number of vertices in \( P \).

**Lemma 9.** Let \( r \geq 2 \) and let \( G \) be a graph on at least \( r + 1 \) vertices. If \( G \) has at most \( r \) interior vertices, then every spanning tree of \( G \) either contains no path of length at least \( r + 1 \) or contains a path of length \( 2k - r \).

**Proof.** Let \( T \) be a spanning tree of \( G \) such that \( T \) contains a path of length at least \( r + 1 \). Let \( P \) be a longest path in \( T \) and let \( u \) and \( v \) be its endpoints. Suppose that \(|P| < 2k - r \). Let \( x_1, \ldots, x_{r+1} \) be the \( r + 1 \) central vertices of \( P \), indexed consecutively in order of increasing distance from \( u \). (If \(|P| \) is even, then there are two choices for the \( r + 1 \) central vertices of \( P \); choose one arbitrarily.) Suppose that some \( x_i \) had a \( k \)-neighbour in \( G \), called \( y \). Let \( Q \) be the path in \( T \) from \( x_i \)
to \( y \). Since, for all \( x, y \in G \), \( d_G(x, y) \leq d_T(x, y) \), we know that \(|Q| \geq k + 1 \). Let 
\( z \) be the furthest vertex from \( x_i \) at which \( P \) and \( Q \) coincide; by symmetry, we 
may assume that \( z \) is closer to \( v \) than \( x_i \) is. We shall construct a path in \( T \) that 
is longer than \( P \), in contradiction to the assumption. Define \( P_+ \) to be the path 
formed by \( Q \) and the portion of \( P \) between \( u \) and \( x_i \). Then

\[
|P_+| = |Q| + d_T(x_1, x_i) + d_T(u, x_1) \geq (k + 1) + (i - 1) + d_T(u, x_1).
\]

Because \( x_1, \ldots, x_{r+1} \) are the \( r + 1 \) central vertices of \( P \), we have

\[
d_T(u, x_1) \geq \frac{|P| - 1 - d_T(x_1, x_{r+1})}{2} \geq \frac{|P| - r - 2}{2},
\]

from which we deduce that

\[
|P_+| \geq k + i + \frac{|P| - r - 2}{2} \geq \frac{|P| + 2k - r}{2} > |P|.
\]

Thus, no \( x_i \) can have a \( k \)-neighbour in \( G \), which means that all of the \( x_i \) must be 
interior vertices. \( \square \)

The following concepts will play key roles in the proof of Theorem \( \ref{thm:4} \). Let \( v \) and 
\( w \) be two vertices at distance \( k \). Recall that a geodesic is a shortest path between 
two vertices in a graph. Define the \( vw \)-path \( P \) to be a shortest path between \( v \) and \( w \). Order the vertices of \( G \) and conduct a breadth-first search starting from 
\( v \) such that the resulting tree \( T_v \) contains \( P \). We define the \( v \)-path \( P_v \) to be the 
longest path in \( T_v \). (If there is more than one longest path in \( T_v \), then we choose one arbitrarily.) Similarly, we define \( T_w \) to be breadth-first tree with respect to \( w \) 
containing \( P \), and the \( w \)-path \( P_w \) to be the longest path in \( T_w \).

Consider the breadth-first tree \( T_v \) and the \( v \)-path \( P_v \). Let \( x \) and \( y \) be the 
endpoints of \( P_v \). Moving along \( P_v \) from \( x \) to \( y \), or in fact along any path in \( T_v \), 
the distance from \( v \) will first decrease, then increase — this is a fundamental 
property of breadth-first search trees, for the depth of a vertex \( w \) in such a tree 
equals \( d_G(v, w) \). Thus, \( P_v \) can be divided into two geodesics. Let \( z \) be a vertex of 
\( P_v \) at minimal distance from \( v \). Let \( P_1 \) denote the portion of \( P_v \) between \( x \) and 
z and \( P_2 \) the portion of \( P_v \) between \( z \) and \( y \); one of these may be empty.

By our assumption that \( r < k - 1 \), if \( v \) and \( w \) are \( k \)-neighbours, then \( T_v \) 
must contain a path of length at least \( r + 1 \). Hence, by Lemma \( \ref{lem:9} \) \( P_v \) contains at 
least \( 2k - r \) vertices. Note also that at most two vertices on \( P_v \) can be at distance \( k \) 
from \( v \).

### 3 Proof of Theorem \( \ref{thm:4} \)

From now on, we shall assume that \( G \) satisfies \( e(G_k) \geq (n - k + 1)^2/4 \). As noted 
above, we shall also assume that \( r < k - 1 \).

Let us briefly discuss the proof of Theorem \( \ref{thm:4} \). First, for \( k \) and \( n \) large enough, 
we shall prove a simple condition under which a pair of \( k \)-neighbours must have at 
least \( 2k - r - 2 \) \( k \)-unaffiliated vertices. We shall deduce from this that each pair of 
\( k \)-neighbours in \( G \) has almost enough \( k \)-unaffiliated vertices to achieve the desired
bound on $e(G_k)$. Second, we shall deduce our key lemma, which says that if some pair of $k$-neighbours $\{v, w\}$ does not have enough $k$-unaffiliated vertices, then all geodesics in $G$ must have at least a few vertices apart from each of $P_v$ and $P_w$. Third, we shall show that in this case, every other $k$-neighbour of $v$ is at distance $o(k)$ from $w$, and vice-versa. Moreover, for some $\delta = o(1)$, we shall show that all vertices that are at distance at least $\delta k$ from both $v$ and $w$ have very few $k$-neighbours, which, by Turán's theorem, will contradict our assumption that $e(G_k) \geq (n - k + 1)^2/4$. Finally, we shall deduce that in this case, $G$ has at most as many $k$-distances as the double broom, which will imply that $e(G_k) = (n - k + 1)^2/4$. Moreover, we shall show that in this case $G$ is $k$-isomorphic to the double broom.

**Lemma 10.** For every $\varepsilon > 0$ there exists a constant $K(\varepsilon)$ such that for all $k \geq K(\varepsilon)$, if some geodesic in $G$ contains $\varepsilon k$ vertices that are $k$-affiliated with either $v \in E(G_k)$ or $w \in E(G_k)$, then we can find $2k$ vertices that are $k$-unaffiliated with both $v$ and $w$.

**Proof.** Suppose that $Q$ is a geodesic and that $\varepsilon k$ vertices of $Q$ are $k$-affiliated with either $v$ or $w$. By the pigeonhole principle, we can assume that $m \geq \varepsilon k/2$ of them are at distance $k$ from $w$. Index them consecutively by $x_1, \ldots, x_m$. For $1 \leq i \leq m$, let $Q_i$ denote a shortest path in $G$ from $x_i$ to $w$. Let us 7-colour the vertices of $Q$ so that $x_i$ is coloured with colour $j$ if $i \equiv j \mod 7$ and choose a colour (red, say) that belongs to at least $m/7$ of the $x_i$. For each red $x_i$, move two steps along the path $Q_i$. In this way, we obtain $m' \geq m/7$ vertices at distance $k - 2$ from $w$ and at distance at least $7 - 2 - 2 = 3$ from each other; let us call them $y_1^1, y_2^1, \ldots, y_m^1$. If a vertex $y_j^1$ is not at distance $k$ from $v$, set $z_j^1 = y_j^1$. If $y_j^1$ is at distance $k$ from $v$, take $z_j^1$ to be the vertex obtained by moving from $y_j^1$ one step towards $v$ (see Figure 2). Since the $y_j^1$ were at pairwise distance at least 3, all of the $z_j^1$ will be distinct vertices at distance between $k - 1$ and $k - 3$ from $w$ and not at distance $k$ from $v$, i.e., they are $k$-unaffiliated with both $v$ and $w$.

Similarly, by 13-colouring $Q$, defining ‘red’ to be the largest colour class, and moving 5 steps along the paths $Q_i$, we find $m'' \geq m/13$ vertices $y_1^2, y_2^2, \ldots, y_m^2$, at distance $k - 5$ from $w$ and at distance at least $13 - 5 - 5 = 3$ from each other. This gives rise to $m/13$ distinct $k$-unaffiliated vertices $z_j^2$ at distance between $k - 4$ and $k - 6$ from $w$ and not at distance $k$ from $v$, i.e., the $z_j^2$ are disjoint from the previously constructed $z_j^1$.

Repeating this procedure for all $6\ell + 1$-colourings up to $\ell = \lfloor m/6 \rfloor$, we obtain a total of at least

$$\left(\frac{1}{7} + \frac{1}{13} + \cdots + \frac{1}{6\lfloor m/6 \rfloor + 1}\right) m \geq \frac{1}{12} \cdot \frac{\varepsilon k}{2} \cdot \frac{\varepsilon k}{2}$$

$k$-unaffiliated vertices, which is greater than $2k$ for $k$ large enough. \[\square\]

**Lemma 11** has the following important corollary.

**Corollary 11.** Let $\{v, w\} \in G_k$. Let $P_v$ and $P_w$ be geodesics as defined above. Then either $|P_v| - \varepsilon k \geq (2k - r) - \varepsilon k$ vertices on $P_v$ are $k$-unaffiliated with $v$ and $w$ or we can find $2k$ vertices that are $k$-unaffiliated with $v$ and $w$ elsewhere. The same is true of $P_w$. \[\square\]
Figure 2: The set of vertices $z_i$ that are $k$-unaffiliated with both $v$ and $w$. 
Corollary 11 and equation (2.5) have the following immediate consequence, which is an approximate version of the bound in Theorem 4.

**Corollary 12.** For every $\varepsilon > 0$ there is a constant $K(\varepsilon)$ such that for all $k \geq K(\varepsilon)$, if $G_k$ is triangle-free then

$$e(G_k) \leq \frac{(n - (1 - \varepsilon)k)^2}{4}.$$ 

\[ \square \]

Lemma 10 and Corollary 11 also give us very useful information about the structure of the graph. The following lemma is the main tool in the remainder of the proof of Theorems 4.

**Lemma 13.** Let $\varepsilon > 0$ and let $k \geq K(\varepsilon)$, where $K(\varepsilon)$ is as in Lemma 11. Suppose that $v$ and $w$ are $k$-neighbours in a graph $G$ that have fewer than $2k - r - 2$ $k$-unaffiliated vertices. Then any geodesic in $G$ contains fewer than $2\varepsilon k$ vertices disjoint from $P_v$, and similarly for $P_w$.

**Proof.** Suppose that $Q$ is a geodesic in $G$ with at least $2\varepsilon k$ vertices disjoint from $P_v$. First, if $\varepsilon k$ of these vertices are $k$-unaffiliated with $v$ and $w$, then, by Corollary 11, $v$ and $w$ have at least $2k \geq 2k - r - 2$ $k$-unaffiliated vertices. Second, if not, then $Q \setminus P_v$ must contain at least $\varepsilon k$ vertices that are $k$-affiliated with $v$ and $w$. In this case, by Lemma 10, we again obtain at least $2k \geq 2k - r - 2$ $k$-unaffiliated vertices.

In either case, we reach a contradiction. \[ \square \]

Let $P$ denote the $vw$-path and let $P'$ denote the $vw'$-path, as defined in Section 2. Recall that the $v$-path $P_v$ splits into two geodesics, which we call $P_1$ and $P_2$, along each of which the distance from $v$ is strictly monotone. Similarly, the $w$-path $P_w$ splits into two geodesics, which we call $P_3$ and $P_4$. Note that since $P_v$ was defined on a tree $T_v$ that contains $P$, we have that $P_v \cap P$ is an interval of $P$, lying entirely in $P_1$ or in $P_2$, and analogously for $P_w \cap P$. For the remainder of the proof, without loss of generality, let $P_v \cap P \subseteq P_1$ and let $P_w \cap P \subseteq P_3$.

As was already mentioned in Section 2, by Lemma 8 we are done if any pair of $k$-neighbours has $2k - r - 2$ $k$-unaffiliated vertices. So let us assume for the sake of contradiction that some vertices $v$ and $w$ at distance $k$ have fewer than that many $k$-unaffiliated vertices. Let $w'$ be another $k$-neighbour of $v$. How large can the distance between $w$ and $w'$ be? The following lemma shows that this distances is either close to $2k$ or close to 0.

**Lemma 14.** Let $\varepsilon > 0$ and let $k \geq K(\varepsilon)$, where $K(\varepsilon)$ is as in Lemma 11. Let $v$ and $w$ be $k$-neighbours with fewer than $2k - r - 2$ $k$-unaffiliated vertices. Let $w'$ be another $k$-neighbour of $v$. Then either $d(w, w') = (2 - o(1))k$ or $d(w, w') = o(k)$.

**Proof.** Recall that $P_v$ and $P_w$ denote the $v$-path and the $w$-path, respectively. By Corollary 11 $P_v$ contains at least $|P_v| - \varepsilon k \geq 2k - r - \varepsilon k$ $k$-unaffiliated vertices. Since, by assumption, $v$ and $w$ have at most $2k - r - 2$ $k$-unaffiliated vertices, by Lemma 13 we must have

$$|P_v| \leq 2k - r + \varepsilon k.$$
Figure 3: When $u' \in P_1$, all $k$-neighbours of $v$ are close together. The dashed segments represent the path $P_1$.

Observe that $G$ contains at most $\varepsilon k$ vertices that are $k$-unaffiliated with $v$ and $w$ and are not on $P_v$, or else we are done by Corollary 11. The same assertions as above hold for $P_w$ in place of $P_v$. By Lemma 13, for the vertex sets of the paths,

$$|P \cap P_v| \geq (1 - 2\varepsilon)k \text{ and } |P \cap P_w| \geq (1 - 2\varepsilon)k.$$  \hfill (3.1)

Let $u$ be the point furthest from $v$ at which $P$ and $P_v$ coincide. By hypothesis, $u \in P_1$. Similarly, let $u'$ be the furthest point from $v$ at which $P'$ and $P_v$ coincide. By Lemma 13, $d(u, w) \leq 2\varepsilon k$ and $d(u', w') \leq 2\varepsilon k$. It follows that

$$(1 - 2\varepsilon)k \leq d(v, u), d(v, u') \leq k.$$  \hfill (3.2)

Now we consider two cases: when $u' \in P_1$ and when $u' \in P_2$.

Suppose that $u'$ lies on $P_1$. Suppose first that $u$ is closer to $v$ than $u'$ is (see Figure 3). Then $d(u, w') = d(u, w) \leq 2\varepsilon k$, thus,

$$d(w, w') \leq d(w, u) + d(u, w') \leq 4\varepsilon k.$$  

If $u'$ is closer to $v$ than $u$ is, then by a similar argument, $d(u', w') = d(u', w') \leq 2\varepsilon k$, and so

$$d(w, w') \leq d(w, u') + d(u', w') \leq 4\varepsilon k.$$  

If, however, $u' \in P_2$, then $d(w, w')$ depends on the length of $P_w \setminus P$. Because $u' \in P_2$, we have $|P_2| \geq (1 - 2\varepsilon)k$. Since $|P_1| + |P_2| = |P_v| + 1 \leq 2k + \varepsilon k$ (and similarly for $|P_3|$, $|P_4|$ and $|P_w|$), we have

$$(1 - 2\varepsilon)k \leq |P_i| \leq (1 + 3\varepsilon)k \text{ for } i = 1, 2.$$  \hfill (3.3)

Now we consider the geodesics $P_3$ and $P_4$ that comprise $P_w$. We have assumed that $P_3$ contains $P_w \cap P$. Then, because $P_4 \cap P = \emptyset$, we have $P_1 \cap P_4 \subseteq P_1 \setminus P$, hence,

$$|P_1 \cap P_4| \leq |P_1 \setminus P| \leq (1 + 3\varepsilon)k - (1 - 2\varepsilon)k = 5\varepsilon k.$$  

Thus, by Lemma 13,

$$|P_4 \setminus P_2| = |P_4 \cap P_1| + |P_4 \setminus P_v| \leq 5\varepsilon k + 2\varepsilon k = 7\varepsilon k.$$
Figure 4: When \( u' \in P_2 \) and \( |P_4| \) is large, all \( k \)-neighbours of \( v \) are again close together. The dashed segments represent the path \( P_4 \). The vertices \( a \) and \( b \) denote the endpoints of \( P_2 \).

Now we shall show that if \( |P_4| \) is at all large, then \( d(w, w') = o(k) \), while if \( |P_4| \) is very small, then \( d(w, w') = 2k - o(k) \).

Case 1: Suppose first that \( |P_4| > 7\varepsilon k \). Then, because \( |P_4 \setminus P_2| \leq 7\varepsilon k \), we have \( P_2 \cap P_4 \neq \emptyset \). Let \( t \) denote the vertex at which \( P_4 \) meets \( P_2 \). Then by Lemma 13 we have \( d(t, w) < 2\varepsilon k \). Let \( q \) be the vertex of \( P_2 \cap P_4 \) that is closest to \( w \) (see Figure 4). Then \( d(q, w) \leq |P_4 \setminus P_2| + d(t, w) < 9\varepsilon k \).

Thus, \( k + 9\varepsilon k \geq d(q, v) \geq k \). Then
\[
d(w, w') \leq d(w, q) + d(q, w') \leq 9\varepsilon k + 11\varepsilon k = 20\varepsilon k,
\]
which completes the proof of Case 1.

Case 2: Suppose instead that \( |P_4| \leq 7\varepsilon k \). Then, by (3.1) and Lemma 13 we have
\[
|P_2 \cap P_3| = |P_2| - |P_2 \setminus P_w| - |P_2 \cap P_4| \geq (k - 2\varepsilon k) - 2\varepsilon k - 7\varepsilon k = k - 13\varepsilon k. \tag{3.4}
\]

Since \( P_3 \) contains \( P_w \cap P \) and \( P_2 \) is edge-disjoint from \( P \), it follows from (3.1) and (3.4) that
\[
|P_2 \cap P_3| + |P \cap P_3| \geq 2k - 13\varepsilon k. \tag{3.5}
\]

Let \( q' \) be the vertex of \( P_2 \cap P_3 \) that is furthest away from \( w \) (see Figure 5). Then, by (3.3) and (3.5),
\[
2k + 3\varepsilon k \geq d(q', v) + d(v, w) \geq d(q', w) \geq |P_2 \cap P_3| + |P \cap P_3| \geq 2k - 13\varepsilon k. \tag{3.6}
\]
Figure 5: When $u' \in P_2$ and $|P_4|$ is small, all $k$-neighbours of $v$ are at distance $(2 - o(1))k$ from one another. The dashed segments represent the portion of the path $P_3$ that is disjoint from $P$. The vertices $a$ and $b$ denote the endpoints of $P_2$.

Also, it follows from (3.3) and (3.4) that

$$k + 3\varepsilon k \geq |P_2| \geq d(q', v) \geq |P_2 \cap P_3| \geq k - 11\varepsilon k.$$  

It follows from this and (3.2) that $d(q', u') \leq 13\varepsilon k$ and therefore that $d(q', w') \leq 15\varepsilon k$. We obtain from this and (3.6) that

$$d(w, w') \geq d(q', w) - d(q', w') \geq 2k - 28\varepsilon k.$$  

This proves the lemma.

We shall now show that the assumption that $d(w, w') = (2 - o(1))k$ for some $w'$ leads to a contradiction. Let $\delta = O(\varepsilon) = o(1)$. For $v \in V(G)$, define the cluster of $v$ to be the set $C_v$ of vertices at distance at most $\delta k$ from $v$.

Lemma 15. Fix $\varepsilon > 0$. For $k$ and $n$ large enough, let $G$ be a graph on $n$ vertices and let $v$ and $w$ be $k$-neighbours with fewer than $2k - r - 2$ unaffiliated vertices. Then every $k$-neighbour of $v$ is at distance $O(\varepsilon k) = o(k)$ from $w$, and vice versa.

Proof. Suppose that $w' \neq w$ is a $k$-neighbour of $v$ such that $d(w, w') = (2 - o(1))k$. In this case, our graph $G$ is ‘flat’, i.e., it contains a geodesic $P_3$ of length $(2 - o(1))k$ and every vertex is $o(k)$ away from $P_3$. The vertex $v$ lies close to the centre of $P_3$, whereas $w$ and $w'$ lie near the opposite ends: every $k$-neighbour of $v$ lies within $o(k)$ of either $w$ or $w'$. Since there are no vertices at distance $(2 - o(1))k$ from $v$, switching $v$ and $w$ in the statement of Lemma 14 yields that all $k$-neighbours of $w$ are close to $v$. Since, by assumption, $v$ and $w$ have fewer than $2k$ unaffiliated vertices, all but at most $2k$ vertices are contained in one of the clusters $C_v$, $C_w$, and $C_{w'}$. Every vertex on $P_3$ not lying within $2\delta k$ of either $v$, $w$ or $w'$ cannot have a $k$-neighbour in any of these clusters. Therefore any such vertex has at most $2k$ $k$-neighbours. So, we have $c = (2 - o(1))k$ vertices of $k$-degree at most $2k$. Applying Turán’s theorem to the remaining vertices, we obtain

$$e(G_k) \leq \left(\frac{n - c}{2}\right)^2 + 2ck < \left(\frac{n - k + 1}{2}\right)^2,$$

provided that $n$ is sufficiently large compared to $k$, contradicting our hypothesis that $e(G_k) \geq (n - k + 1)^2/4$. \qed
Figure 6: The vertices of $S_1$, shown in grey, are just outside of $C_v$, and so have no $k$-neighbours in either $C_v$ or $C_w$.

Since all but at most $2k$ of the vertices in $G$ are at distance $k$ from either $v$ or $w$, we can conclude that all but at most $2k$ vertices lie either in $C_v$, that is, within distance $o(k)$ of $v$, or in $C_w$. This is a fairly strong structural property of $G$ and from here it is a short step to completing the proof of Theorem 4.

Proof of Theorem 4. Observe that by the definition of $\delta$, every vertex on the $vw$-path $P$ that is at distance more than $\delta k$ from both $v$ and $w$ cannot have a $k$-neighbour in either cluster. We shall use these vertices, together with a small set that we shall now construct, to produce a contradiction as in the proof of Lemma 15.

We shall now show that $|P_2|$ must be very small and that $|P_1|$ cannot be much larger than $k$. It will then follow that $|P_v|$ cannot be much larger than $k$, either.

If $|P_2| > 10\delta k$, then $P_2$ contains $4\delta k$ vertices at distance between $2\delta k$ and $6\delta k$ from $v$. By Lemma 13 at least $4\delta k - 2\varepsilon k > 3\delta k$ of these vertices are contained either in $P_3$ or in $P_4$. We shall consider two cases: when at least $\delta k$ of these vertices are in $P_3$ and when at least $2\delta k$ of them are in $P_4$. We shall show below that either case produces a contradiction as in the proof of Lemma 15. Suppose first that $\delta k$ of them are in $P_3$. Let $S_1 = \{ x \in P_3 : 2\delta k \leq d(v, x) \leq 6\delta k \}$, let $y$ be the closest vertex of $P_3$ to $v$ and let $s \in P_2$ be such that $d(v, s) = 2\delta k$ (see Figure 6). By Lemma 13 and the triangle inequality, we have $d(y, s) \geq 2\delta k - 2\varepsilon k$. So,

$$d(w, s) \geq k - 2\varepsilon k + 2\delta k - 2\varepsilon k \geq k + \delta k,$$

which means that the vertices of $S_1$ are all at distance at least $k + \delta k$ from $w$ and therefore have no $k$-neighbours in the clusters.

Suppose instead that $2\delta k$ of them are in $P_4$. We shall show that this implies that $|P_4| \sim k$. Let $x$ be the closest vertex of $P_4$ to $w$ and let $z$ be the point at which $P_4$ meets $P_2$. Let $a$ be the closest point of $P_2$ to $v$ (see Figure 7). We may assume that

$$3\delta k < 4\delta k - 2\varepsilon k \leq d(a, z) \leq d(v, z) \leq 4\delta k.$$

Then, because $a, x \in P$,

$$|P_4| \geq d(x, z) \geq d(a, x) - d(a, z) \geq k - 4\varepsilon k - 4\delta k. \tag{3.7}$$

Also, since $|P_w| \leq (2 + \varepsilon)k$ and $|P_3| \geq (1 - 2\varepsilon)k$, it follows that

$$|P_4| \leq (1 + 3\varepsilon)k. \tag{3.8}$$

13
Figure 7: The vertices of $S_2$, shown in grey, have no $k$-neighbours in either cluster.

Then $|P_4| \sim k$. Let $S_2$ denote the $2\delta k$ central vertices of $P_4$. Observe that if $q \in S_2$, then both $d(q, w)$ and $d(q, v)$ are bounded away from both 0 and $k$. Indeed, after a bit of calculation, it follows from (3.7) and (3.8) that

$$k - 4\varepsilon k - 4\delta k \leq d(q, w) \leq \frac{(1 + 3\varepsilon)k}{2} + \delta k$$

and that

$$3\delta k + \frac{k - 4\varepsilon k - 4\delta k}{2} - \delta k \leq d(q, v) \leq 4\delta k + \frac{(1 + 3\varepsilon)k}{2} + \delta k.$$

Together with the aforementioned vertices on $P$, in each case we obtain $c \geq k + \delta k$ vertices that are not in either cluster and whose $k$-degree is at most $2k$. By applying Turán’s theorem to these vertices as in the proof of Lemma 15, we obtain a contradiction for large enough $n$. Thus, $|P_3| \leq 10\delta k$.

Similarly, we claim that $|P_1| \leq k + 10\delta k$. If not, then $P_1$ contains at least $3\delta k$ vertices that are at distance between $k + 2\delta k$ and $k + 5\delta k$ from $v$, and at distance between $2\delta k - 2\varepsilon k$ and $5\delta k + 2\varepsilon k$ from $w$, and we obtain a contradiction as above.

We therefore have

$$|P_v| \leq k + 20\delta k.$$

Let $m$ be the midpoint of $P$, or one of the midpoints if $k$ is odd. Suppose that there exist $k$-neighbours $x$ and $y$ in $G$ such that every $k$-geodesic between $x$ and $y$ misses $m$. Let $Q$ be such a geodesic. Then it must miss either the path $P_{vm}$ between $v$ and $m$ or the path $P_{wm}$ between $w$ and $m$. In either case, $Q$ will contain
a geodesic of length at least $k/4$ disjoint from $P$, a contradiction. Hence, every pair of $k$-neighbours is connected by a geodesic containing $m$.

It follows that $G$ is $k$-isomorphic to $T_m$, the breadth-first tree with respect to $m$. By Theorem 1, $T_m$ has at most as many $k$-distances as the double broom does. Thus, $e(G) = (n - k + 1)^2/4$ and $G$ is $k$-isomorphic to the double broom. The proof of Theorem 4 is complete. 

4 Discussion

We believe that Conjecture 3 may be susceptible to a similar approach to the one above. In Section 2 we obtained our first non-trivial bound, Lemma 8, by adapting a proof of Mantel’s theorem, which is the simplest case of Turán’s theorem. Unfortunately we were not able find a straightforward generalisation of this approach to $K_{t+1}$-free distance-$k$ graphs when $t \geq 3$. A possible solution might be to adapt a proof of Turán’s theorem that works for all $t$. However, it seems difficult to generalise Lemma 9 for values of $t$ greater than 2.

We also note that in [10], Csikvári asked a similar question about maximising or minimising the number of (closed) walks of length $k$ in a connected graph $G$ on $n$ vertices and $m$ edges. In the same paper Csikvári settled the case of closed walks and $m = n - 1$, that is, when $G$ is a tree. The answer for general walks on trees was given in [8], but the general case remains open.

5 Acknowledgments

We would like to thank Béla Bollobás for introducing us to the problem and for helpful discussions. We would like to thank the anonymous referees for their helpful comments. We would also like to thank Paul Balister for helpful comments and for running the computer search mentioned after Example 6.

References

[1] R. Ahlswede and G.O.H. Katona. Graphs with maximal number of adjacent pairs of edges. Acta Math. Acad. Sci. Hungar., 32:97–120, 1978.

[2] M. Aigner. Turán’s graph theorem. Amer. Math. Monthly, 102(9):808–816, 1995.

[3] N. Alon. On the number of subgraphs of prescribed type of graphs with a given number of edges. Israel J. Math., 38:116–130, 1981.

[4] N. Alon. On the number of certain subgraphs contained in graphs with a given number of edges. Israel J. Math., 53:97–120, 1986.

[5] B. Bollobás and P. Erdős. Graphs of extremal weights. Ars Combin., 50:225–233, 1998.

[6] B. Bollobás and A. Sarkar. Paths in graphs. Studia Sci. Math. Hungar., 38:115–137, 2001.
[7] B. Bollobás and A. Sarkar. Paths of length four. *Discrete Math.*, 265:357–363, 2003.

[8] B. Bollobás and M. Tyomkyn. Walks and paths in trees. *J. Graph Theory*, 70(1):54–66, 2012.

[9] O.D. Byer. Maximum number of 3-paths in a graph. *Ars Combin.*, 61:73–79, 2001.

[10] P. Csikvári. On a poset of trees. *Combinatorica*, 30(2):125–137, 2010.

[11] Z. Füredi. Graphs with maximum number of star-forests. *Studia Sci. Math. Hungar.*, 27:403–407, 1992.

[12] M. Katz. Rearrangements of (0–1) matrices. *Israel J. Math.*, 9:53–72, 1971.