On Hanf-equivalence and the number of embeddings of small induced subgraphs

Stephan Kreutzer
Technical University Berlin
stephan.kreutzer@tu-berlin.de

Nicole Schweikardt
Goethe-University Frankfurt am Main
schweika@cs.uni-frankfurt.de

Abstract

Two graphs are Hanf-equivalent with respect to radius r if there is a bijection between their vertex sets which preserves the isomorphism types of the vertices' neighbourhoods of radius r. For $r = 1$ this means that the graphs have the same degree sequence.

In this paper we relate Hanf-equivalence to the graph-theoretical concept of subgraph equivalence. To make this concept applicable to graphs that are not necessarily connected, we first generalise the notion of the radius of a connected graph to general graphs in a suitable way, which we call the *generalised radius*. We say that two graphs G and H are subgraph-equivalent up to generalised radius r if for all graphs S of generalised radius r, the number of induced subgraphs isomorphic to S is the same in G and H. We prove that Hanf-equivalence with respect to radius r is equivalent to subgraph-equivalence up to generalised radius r, thereby relating the purely logical and the graph-theoretical concepts in a very strong way.

The notion of subgraph-equivalence up to order s is defined accordingly, where all graphs S of order at most s are taken into account. As a corollary we obtain that Hanf-equivalence with respect to radius r implies subgraph-equivalence up to order s, provided that $r \geqslant 3s/4$. In particular, this implies that two graphs which are Hanf-equivalent with respect to radius $3s/4$ satisfy exactly the same unions of conjunctive queries of quantifier rank at most s.

Categories and Subject Descriptors F.4.1 [*Mathematical Logic and Formal Languages*]: Mathematical Logic—Computational Logic, Model Theory

General Terms Theory, Algorithms

Keywords finite model theory, graph theory, Hanf-locality, degree sequences

1. Introduction

A degree sequence of a graph G with vertex set $\{v_1, \ldots, v_n\}$ is the sequence $(d_1, \ldots, d_n)$ listing the degrees $d_i := \deg^G(v_i)$ of the vertices v_i in G. As the order of the vertices is not important in this context, degree sequences are arranged in non-decreasing order so that isomorphic graphs have identical degree sequences.

CSL-LICS 2014, July 14–18, 2014, Vienna, Austria.

ACM 978-1-4503-2886-9…$15.00.
http://dx.doi.org/10.1145/2603088.2603148

Degree sequences have been studied intensively in graph theory with many results on classifying the sequences of non-negative integers which can occur as degree sequences of graphs or multigraphs and algorithms constructing graphs realising a given degree sequence. See for instance [3, 4, 7, 10, 12, 19, 21, 22].

Degree sequences are also related to colour refinement algorithms for the graph isomorphism problem such as the Weisfeiler-Lehman algorithm (see e.g. [1, 23, 24]) which, in its first step, essentially computes and compares the degree sequence of the two input graphs. More precisely, the algorithm, given two graphs as input, first labels each vertex in both graphs by its degree and compares the number of vertices in both graphs of every possible degree. In its next steps the algorithm then refines this labelling by computing for each vertex the number of neighbours of any given label and so on until the process becomes stable. With respect to degree sequences, therefore, the first step is to compare the degree sequence of both graphs and then to compare the 1-neighbourhoods of vertices with respect to the labels occurring in them and so forth.

Degree sequences can be generalised in many ways. For instance, instead of simply counting for each vertex the number of its neighbours one can determine the isomorphism type of its 1-neighbourhood and count the number of vertices whose 1-neighbourhood have a given type. This idea can be extended to r-neighbourhoods for any fixed radius r, leading to the concept of a *Hanf sequence* of a graph (or, more generally, a relational structure) which lists the isomorphism types of r-neighbourhoods occurring in the graph and for each such type the number of vertices whose r-neighbourhoods are of this type. This concept has been studied intensively in logic, especially in *finite model theory*. Two graphs H and G are called *Hanf-equivalent with respect to radius* r, denoted by $H \approx_r^{\mathit{Hanf}} G$, if for every possible isomorphism type of r-neighbourhoods the number of vertices in the two graphs whose r-neighbourhoods are of this type is the same. A classical result is Hanf's theorem [11], stating that for every quantifier rank r there is a radius R such that any two graphs which are Hanf-equivalent with respect to radius R satisfy the same first-order formulas of quantifier rank r. This result has found numerous applications for instance in showing that certain properties are not first-order definable (see e.g. [13, 18]) or in evaluation algorithms for first-order logic, such as in [20]. It is known that R can be chosen as 2^{r-1}, and this is essentially optimal (see [15, 16]).

A different way of generalising degree sequences of graphs is used in the *s-dimensional* version of the Weisfeiler-Lehman algorithm [8] (the following description of the algorithm is basically taken from [9]). In its first step, the algorithm labels, in each graph, each s-tuple $\bar{v}$ of vertices of the graph by the isomorphism type of the subgraph induced by the vertices in $\bar{v}$ (viewed as a labeled graph where each vertex is labeled by the positions in the tuple where it occurs). In its next steps the algorithm then refines this labelling by taking into account the colours of all neighbours of $\bar{v}$ in the Hamming metric.

This has led us to consider the concept of *subgraph-equivalence* of graphs. We say that two graphs G and H are *subgraph-equivalent up to order* s if for all graphs S of order at most s (i.e., S is a graph on at most s vertices), the number of induced subgraphs of G isomorphic to S is the same as the number of such subgraphs of H. Of course, instead of the order s of the graphs S, also other graph invariants could be considered. The one that is of particular interest for the present paper is a new notion of a *generalised radius* of a graph S, which is defined as the sum of the radii of the connected components of S plus half the number of connected components of S. The motivation for this definition is that if we have a graph S with k components and we want to add edges to S to obtain a connected graph S', then it is easily seen that the worst diameter of S' we may end up with is at most two times the generalised radius. In fact, we will prove below that the generalised radius of S is an upper bound for the radius of S'.

We write $H \approx^{\textit{subgraph}}_{\textit{genrad } r} G$ to indicate that G and H are subgraph equivalent with respect to generalised radius r. The main result of this paper is to bring the concepts of Hanf-equivalence and subgraph-equivalence together. More precisely, we prove the following theorem.

Theorem. *For all graphs G, H and all $r \geqslant 0$,*

$$G \approx^{\textit{Hanf}}_{r} H \text{ if, and only if, } G \approx^{\textit{subgraph}}_{\textit{genrad } r} H.$$

Note that Hanf's theorem (in the form stated above), implies that any two graphs which are Hanf-equivalent with respect to radius $R = 2^{r-1}$ contain the same subgraphs of order at most r, since the existence of such a subgraph can be described by a (purely existential) first-order sentence of quantifier rank r. By suitably modifying the standard proof of Hanf's theorem (cf., e.g., [6, 18]), one can also obtain that Hanf-equivalence w.r.t. radius $R = 3^{O(r)}$ implies subgraph-equivalence up to order r. Using this proof technique, however, it seems impossible to reduce the required Hanf-equivalence radius R from $3^{O(r)}$ to a number that is polynomially related to r.

A connection between subgraph equivalence with respect to generalised radius and Hanf-equivalence also follows from [17]. In this paper, a logic $\mathcal{LSO}^{*}_{\infty\omega}(\mathbf{C})$ is defined which captures exactly the Hanf-local properties. This logic is an extension of infinitary first-order logic by counting as well as restricted second-order properties. For any fixed graph S, the property that a graph contains a certain number of subgraphs isomorphic to S can be formalised in this logic by a formula whose rank only depends on S and not on the number of copies of S that we want to certify. Here, the rank of a formula is the suitable concept of quantifier rank for this logic. As every property definable in this logic is Hanf-local, the main result of [17] implies that for any graph S there is a number R such that any two structures which are Hanf-equivalent up to radius R contain the same number of isomorphic copies of S. While this already relates subgraph equivalence to Hanf-equivalence, the radius R again depends exponentially on the generalised radius of S. Hence, in this way we cannot get the exact correspondence between Hanf-locality and subgraph equivalence with respect to generalised radius established in our main result.

In this paper we follow a different proof technique which allows us to obtain a precise characterisation, stating that Hanf-equivalence with respect to radius r coincides with subgraph-equivalence up to generalised radius r. While this result is not overly complicated to see for connected subgraphs S, it becomes significantly more complex in case that the subgraphs S are not connected, as Hanf-equivalence only speaks about neighbourhoods, and hence connected subgraphs.

Our theorem also provides an easy way of checking whether a graph S occurs in a graph G the same number of times as in a graph H. For, checking whether G and H are Hanf-equivalent with respect to radius r can be much easier than comparing the number of times S occurs as a subgraph. For instance, if G is the directed cycle C_{2n} on $2n$ vertices and H is the disjoint union of two directed cycles on n vertices each, then it is trivially seen that G and H are r-Hanf-equivalent if, and only if, $2r + 1 < n$. However, even determining whether the graph T_5 consisting of 5 disjoint directed edges occurs in both graphs the same number of times as subgraph requires a little thought. And if instead of T_5 we choose a more complicated graph, the subgraph analysis can get quite complicated. Using our main result immediately implies that T_5, which has generalised radius 7, is contained in G and H for the same number of times as subgraph if $n > 2r + 1$ for any $r \geqslant 7$.

As a corollary of our main result, we also obtain that Hanf-equivalence with respect to radius r implies subgraph-equivalence up to order s, provided that r is of size at least $\lfloor 3s/4 \rfloor$. In particular, this implies that two graphs which are Hanf-equivalent with respect to radius $r \geqslant \lfloor 3s/4 \rfloor$ satisfy the same unions of conjunctive queries of quantifier rank at most s.

Let us note that all our results easily generalise from finite graphs to finite relational structures.

The rest of the paper is organised as follows. The basic notation is fixed in Section 2. In Section 3, we formally introduce the concept of Hanf-equivalence and illustrate it by an example. In Section 4, we define the generalised radius of a graph and the concept of subgraph-equivalence up to generalised radius r. We then prove the main theorem of our paper, showing that Hanf-equivalence and subgraph-equivalence with respect to radius r coincide. The corollary concerning subgraph-equivalence up to the order of the induced subgraphs is presented in Section 5. We conclude in Section 6 where we also state the implication of our results to unions of conjunctive queries. A detailed example is given in the appendix.

2. Preliminaries

This section fixes basic notation used throughout the paper. We assume that the reader is familiar with basic notions from graph theory (cf., e.g., [2, 5]).

We write $\mathbb{N}$ to denote the set of non-negative integers, and we let $\mathbb{N}_{\geqslant 1} := \mathbb{N} \setminus \{0\}$. For $n \in \mathbb{N}_{\geqslant 1}$ we let $[n] := \{1, \ldots, n\} = \{i \in \mathbb{N} : 1 \leqslant i \leqslant n\}$. Letters i, j, k, m, n, r, s, R will always denote non-negative integers, and we will write $r \geqslant 0$ or $r \geqslant 1$ instead of $r \in \mathbb{N}$ or $r \in \mathbb{N}_{\geqslant 1}$. For a set V we write $\binom{V}{2} := \{X \subseteq V : |X| = 2\}$ to denote the set of all 2-element subsets of V.

All graphs considered in this paper are simple and have a *finite* and non-empty vertex set. An *undirected* graph $G = (V, E)$ has edge set $E \subseteq \binom{V}{2}$, a *directed* graph $G = (V, E)$ has edge set $E \subseteq V \times V$. Given a graph G, we write $V(G)$ and $E(G)$ to denote its vertex set and edge set, respectively. The *order* of a graph G is the cardinality of its vertex set, i.e., $\text{order}(G) = |V(G)|$.

In the following, we will often speak of *graphs* without specifying whether they are directed or undirected. In particular, this paper's main result holds for undirected graphs as well as for directed graphs.

The subgraph of a graph of G *induced by a set* $W \subseteq V(G)$, denoted $G[W]$, is the graph with vertex set W and whose edge set consists of all edges $e \in E(G)$ such that both endpoints of e belong to W. Let G be a graph and $u, v \in V(G)$. A *path* of length k in G from u to v is a sequence $P := (u_0{=}u, u_1, \ldots, u_k{=}v)$ of vertices such that there is an edge from u_i to u_{i+1} for all $i < k$. We allow $k = 0$ in which case $u = v$. The path is *simple* if $u_i \neq u_j$ for all $i \neq j$. An undirected graph G is *connected* if for any pair u, v of vertices, G contains a path from u to v. A *connected component* of an undirected graph G is a maximal connected induced subgraph of G, i.e., a subgraph of G induced by a set $W \subseteq V(G)$ such that $G[W]$ is connected, and there is no W' with $W \subsetneq W' \subseteq V(G)$ such that $G[W']$ is connected.

For a directed graph G, we write $\mathrm{U}(G)$ to denote the *undirected version of* G, i.e., the undirected graph with vertex set $V(G)$ and edge set $\{\{u,v\} : (u,v) \in E,\ u \neq v\}$. A directed graph G is called *connected* if its undirected version $\mathrm{U}(G)$ is connected. The *connected components* of G are defined to be the connected components of $\mathrm{U}(G)$.

For two graphs G and H we write $\pi : G \cong H$ to indicate that π is an *isomorphism* from G to H, i.e., a bijective mapping from $V(G)$ to $V(H)$ such that for all nodes $u, v \in V(G)$ there is an edge from u to v in $E(G)$ iff there is an edge from $\pi(u)$ to $\pi(v)$ in $E(H)$. For $\bar{u} = (u_1, \ldots, u_k) \in V(G)^k$ and $\bar{v} = (v_1, \ldots, v_k) \in V(H)^k$ we write $\pi : (G, \bar{u}) \cong (H, \bar{v})$ to indicate that π is an isomorphism from G to H such that $\pi(u_i) = v_i$ for all $i \in \{1, \ldots, k\}$.

A *homomorphism* from a graph G to a graph H is a mapping $h : V(G) \to V(H)$ such that if $\{u,v\} \in E(G)$ then $\{h(u), h(v)\} \in E(H)$, for all $u, v \in V(G)$. For directed graphs, the definition is analogous, preserving directed edges.

The distance $\mathrm{dist}^G(u,v)$ between two vertices u and v in an undirected graph G is the minimal length of a path between u and v in G (in particular, it is 0 if $u = v$, and it is ∞ if there is no path between u and v). For a directed graph G we let $\mathrm{dist}^G(u,v) := \mathrm{dist}^{\mathrm{U}(G)}(u,v)$.

The greatest distance between any two vertices of a graph G is the *diameter* of G, denoted by $\mathrm{diam}(G)$. I.e.

$$\mathrm{diam}(G) = \max_{x,y \in V(G)} \mathrm{dist}^G(x,y).$$

The *radius* of a graph G, denoted $\mathrm{rad}(G)$, is defined as

$$\mathrm{rad}(G) \ = \ \min_{x \in V(G)} \ \max_{y \in V(G)} \ \mathrm{dist}^G(x,y).$$

A vertex $u \in V(G)$ which achieves this minimum, i.e. for which $\max_{y \in V(G)} \mathrm{dist}^G(u,y) = \mathrm{rad}(G)$, is called *central*. Clearly, G is connected iff $\mathrm{rad}(G) < \infty$ iff $\mathrm{diam}(G) < \infty$. It is straightforward to see that $\mathrm{rad}(G) \leqslant \mathrm{diam}(G) \leqslant 2\,\mathrm{rad}(G)$.

3. Hanf-Equivalence: a Generalisation of the Degree Sequence of Graphs

For a (directed or undirected) graph G, a vertex u of G and a number $r \geqslant 0$, the *r-neighbourhood* of u is the set

$$N_r^G(u) \ := \ \{v \in V(G) \,:\, \mathrm{dist}^G(u,v) \leqslant r\}.$$

We write $\mathcal{N}_r^G(u)$ to denote the subgraph of G induced by $N_r^G(u)$. Clearly, if G is of radius $\mathrm{rad}(G) \leqslant r$ and u is a central vertex of G, then $N_r^G(u) = V(G)$ and $\mathcal{N}_r^G(u) = G$.

Definition 3.1 (neighbourhood types) An *r-neighbourhood type ϱ (with one centre)* is specified by a tuple (F, w), where F is a graph and w is a vertex of F such that $N_r^F(w) = V(F)$. A vertex u of a graph G has *r-neighbourhood type* ϱ iff $\big(\mathcal{N}_r^G(u), u\big) \cong \big(F, w\big)$. We write $\#_\varrho(G)$ to denote the number of vertices $u \in V(G)$ of r-neighbourhood type ϱ.

Example 3.2 *For all $n \geqslant 1$ let C_n be a directed cycle on n nodes. I.e., C_n has vertex set $\{0, \ldots, n{-}1\}$ and edge set $\{(i, i{+}1) : i \in \{0, \ldots, n{-}1\}\}$, where addition is modulo n.*

Let $r \geqslant 0$ and let u be an arbitrary node of C_n. Then, the r-neighbourhood $N_r^{C_n}(u)$ consists of node u, all nodes reachable from u by a path of length at most r, and all nodes from which u is reachable by a path of length at most r.

Thus, if $2r{+}1 < n$, then the r-neighbourhood type ϱ of u in C_n is specified by the tuple (P_{2r+1}, w_{mid}) where P_{2r+1} is a directed path on $2r{+}1$ nodes and w_{mid} is the node in the middle of this path. On the other hand, if $2r{+}1 \geqslant n$, then the r-neighbourhood type ϱ of u in C_n is specified by the tuple (C_n, w) where w is an arbitrary node of C_n. In both cases, all nodes of C_n have the same r-neighbourhood type ϱ, and $\#_\varrho(C_n) = n$.

Definition 3.3 (Hanf-equivalence) Let $r \geqslant 0$. Two graphs G and H are *Hanf-equivalent w.r.t. radius* r, in symbols:

$$G \approx_r^{\textit{Hanf}} H$$

if $\ \#_\varrho(G) = \#_\varrho(H)\ $ for all r-neighbourhood types ϱ (with one centre).

Clearly, $G \approx_r^{\textit{Hanf}} H$ if, and only if, there is a bijection $\beta : V(G) \to V(H)$ such that for all vertices u of G we have

$$\big(\mathcal{N}_r^G(u), u\big) \ \cong \ \big(\mathcal{N}_r^H(\beta(u)), \beta(u)\big).$$

Using this, it is straightforward to see (cf., e.g., the textbook [18]) that $G \approx_r^{\textit{Hanf}} H$ implies that $G \approx_{r'}^{\textit{Hanf}} H$ for all $r' \leqslant r$, and thus $\#_{\varrho'}(G) = \#_{\varrho'}(H)$ for all r'-neighbourhood types ϱ' (with one centre).

Example 3.4 *Recall from Example 3.2 that C_n is the directed cycle on n nodes. Furthermore, let $D_{n,n}$ be the graph consisting of two disjoint copies of C_n.*

Now let $r \geqslant 0$ with $2r{+}1 < n$. As noted in Example 3.2, all nodes of $D_{n,n}$ have the same r-neighbourhood type ϱ, and this r-neighbourhood type is specified by (P_{2r+1}, w_{mid}). In particular, $\#_\varrho(D_{n,n}) = 2n$. Furthermore, since $2r{+}1 < n \leqslant 2n$, Example 3.2 also tells us that all nodes of C_{2n} have r-neighbourhood type ϱ and thus $\#_\varrho(C_{2n}) = 2n$. Therefore, $C_{2n} \approx_r^{\textit{Hanf}} D_{n,n}$.

On the other hand, for $r \geqslant 0$ with $2r{+}1 \geqslant n$, we know that all nodes of $D_{n,n}$ have r-neighbourhood type $\varrho := (C_n, w)$ for an arbitrary node w of C_n. Furthermore, no node of C_{2n} has this r-neighbourhood type. Thus, $\#_\varrho(D_{n,n}) = 2n \neq 0 = \#_\varrho(C_{2n})$. Therefore, $C_{2n} \not\approx_r^{\textit{Hanf}} D_{n,n}$. In summary, we thus obtain for all $n \geqslant 1$ and $r \geqslant 0$ that

$$C_{2n} \approx_r^{\textit{Hanf}} D_{n,n} \quad \iff \quad 2r{+}2 \leqslant n. \tag{1}$$

4. Subgraph-equivalence with respect to the generalised radius of the considered subgraphs

Definition 4.1 (embeddings) An *embedding* of a graph S in a graph G is an injective mapping $\eta : V(S) \to V(G)$ such that η is an isomorphism from S to the subgraph $G\big[\{\eta(v) \,:\, v \in V(S)\}\big]$ of G induced by the image of $V(S)$ under η. We write

- $\textit{emb}_S(G)$ to denote the set of all embeddings of S in G and
- $\#_S(G)$ to denote the number of all embeddings of S in G, i.e., $\#_S(G) = |\textit{emb}_S(G)|$.

If S is a graph with vertex set $[s] = \{1, \ldots, s\}$ (for some $s \geqslant 1$), it will be convenient to identify an embedding η of S in a graph G with the tuple $\big(\eta(1), \ldots, \eta(s)\big)$. Thus, we will identify $\textit{emb}_S(G)$ with the set of all tuples $\bar{x} = (x_1, \ldots, x_s)$ of s pairwise distinct vertices of G such that for all $i, j \in [s]$ there is an edge from x_i to x_j in G iff there is an edge from i to j in S.

For $k \geqslant 1$ and $\bar{r} = (r_1, \ldots, r_k) \in \mathbb{N}^k$ we let

$$f(\bar{r}) \ := \ \Big(\sum_{i=1}^{k} r_i\Big) \ + \ \big\lfloor \tfrac{k}{2} \big\rfloor\,.$$

We define the *generalised radius* genrad(S) of a graph S as follows: Let $k \geqslant 1$, let $S_1, \dots, S_k$ be the connected components of S, and let $r_1, \dots, r_k$ be the radii of these connected components. Then,

$$\mathrm{genrad}(S) \;:=\; f\big((r_1,\dots,r_k)\big) \;=\; \big(\sum_{i=1}^{k} r_i\big) \;+\; \lfloor \tfrac{k}{2} \rfloor .$$

To give some intuition for the definition of the generalised radius, suppose we want to connect the k components $S_1, \dots, S_k$ of S by $k{-}1$ edges to form a connected graph. Then it is easily seen that the diameter of the resulting graph S' is at most twice its generalised radius. In fact, we show in Lemma 4.5 that the radius of S' is at most genrad(S).

Definition 4.2 (Subgraph-equivalence w.r.t. gen. radius) Let $r \geqslant 0$. Two graphs G and H are *subgraph-equivalent w.r.t. generalised radius* r, in symbols:

$$G \approx^{\mathit{subgraph}}_{\mathit{genrad}\ r} H,$$

if $\#_S(G) = \#_S(H)$ for all graphs S with genrad$(S) \leqslant r$.

Example 4.3 *Let us revisit the graphs $D_{n,n}$ and C_{2n}. As shown in Example 3.4, $D_{n,n}$ and C_{2n} are* Hanf-equivalent *w.r.t. radius r if, and only if, $2r{+}2 \leqslant n$.*

But for which graphs S do we have $\#_S(D_{n,n}) = \#_S(C_{2n})$? For example, for the graph T_5 consisting of 5 disjoint directed edges, is $\#_{T_5}(D_{n,n}) = \#_{T_5}(C_{2n})$? And for which numbers r are the graphs $D_{n,n}$ and C_{2n} subgraph-equivalent *w.r.t. generalised radius r?*

Our following Theorem 4.4 gives an answer to this question: $D_{n,n} \approx^{\mathit{subgraph}}_{\mathit{genrad}\ r} C_{2n} \iff D_{n,n} \approx^{\mathit{Hanf}}_{r} C_{2n} \iff 2r{+}2 \leqslant n$. In particular, for $n \geqslant 2r{+}2$ and $r \geqslant$ genrad$(T_5) = 7$ we have $\#_{T_5}(D_{n,n}) = \#_{T_5}(C_{2n})$.

The main result of this paper states that Hanf-equivalence w.r.t. radius r and subgraph-equivalence w.r.t. generalised radius r coincide, i.e.:

Theorem 4.4 *For all graphs G, H and all $r \geqslant 0$,*

$$G \approx^{\mathit{Hanf}}_{r} H \text{ if, and only if, } G \approx^{\mathit{subgraph}}_{\mathit{genrad}\ r} H.$$

The remainder of this section is devoted to the proof of Theorem 4.4. For the proof, the following generalisation of r-neighbourhood types (with one centre) to $\bar{r}$-neighbourhood types with k centres is essential.

Let $k \geqslant 1$, let $\bar{r} = (r_1, \dots, r_k) \in \mathbb{N}^k$. For a graph G and a k-tuple $\bar{u} = (u_1, \dots, u_k) \in V(G)^k$, the *$\bar{r}$-neighbourhood* of $\bar{u}$ is the set

$$N^G_{\bar{r}}(\bar{u}) \;:=\; \bigcup_{i=1}^{k} N^G_{r_i}(u_i).$$

We write $\mathcal{N}^G_{\bar{r}}(\bar{u})$ to denote the subgraph of G induced by $N^G_{\bar{r}}(\bar{u})$.

An *$\bar{r}$-neighbourhood type τ (with k centres)* is specified by a tuple $(F, \bar{w})$ where F is a graph and $\bar{w} \in V(F)^k$ such that $N^F_{\bar{r}}(\bar{w}) = V(F)$. A k-tuple $\bar{u}$ of vertices of a graph G has *$\bar{r}$-neighbourhood type* τ iff $\big(\mathcal{N}^G_{\bar{r}}(\bar{u}), \bar{u}\big) \cong \big(F, \bar{w}\big)$. We write

- *type*$^G_{\bar{r}}(\bar{u})$ to denote the $\bar{r}$-neighbourhood type of the k-tuple $\bar{u} \in V(G)^k$,
- *set*$_\tau(G)$ to denote the set of all k-tuples $\bar{u} \in V(G)^k$ of $\bar{r}$-neighbourhood type τ, and
- $\#_\tau(G)$ to denote the number of k-tuples $\bar{u} \in V(G)^k$ of $\bar{r}$-neighbourhood type τ, i.e., $\#_\tau(G) = |\mathit{set}_\tau(G)|$.

In the following, whenever we write $\tau = (F, \bar{w})$, we actually mean that τ is specified by $(F, \bar{w})$.

Lemma 4.5 *Let $k \geqslant 1$, let $\bar{r} = (r_1, \dots, r_k) \in \mathbb{N}^k$, and let τ be an $\bar{r}$-neighbourhood type with k centres, specified by a tuple $(F, \bar{w})$. If F is connected, then* rad$(F) \leqslant f(\bar{r})$.

Proof. Let τ be an $\bar{r}$-neighbourhood type with k centres specified by $(F, \bar{w})$ with $\bar{w} = (w_1, \dots, w_k)$ such that F is connected. In particular, $V(F) = N^F_{r_1}(w_1) \cup \dots \cup N^F_{r_k}(w_k)$. Our goal is to show that rad$(F) \leqslant f(\bar{r})$.

Let G be the *weighted* undirected graph with vertex set $V(G) := \{w_1, \dots, w_k\}$ such that for all $w_i, w_j \in V(G)$ with $w_i \neq w_j$ there is an edge between w_i and w_j if, and only if, the subgraph H of F induced by $N^F_{r_i}(w_i) \cup N^F_{r_j}(w_j)$ is connected. The *weight* of this edge is dist$^H(w_i, w_j)$. Clearly, dist$^H(w_i, w_j) \leqslant r_i + 1 + r_j$.

Since F is connected, we know that also G is connected. Let T_0 be a spanning tree of G. In particular, T_0 has vertex set $V(T_0) = \{w_1, \dots, w_k\}$. The length of a path in T_0 is defined to be the sum of the weights of the edges traversed by the path, and the distance dist$^{T_0}(w_i, w_j)$ between two nodes w_i and w_j of T_0 is defined as the length of the (unique) simple path $(w_i, w_{\ell_1}, \dots, w_{\ell_s}, w_j)$ from w_i to w_j in T_0. Clearly,

$$\begin{aligned} &\mathrm{dist}^{T_0}(w_i, w_j) \\ \leqslant\;& r_i{+}1{+}r_{\ell_1} \;+\; \sum_{\nu=1}^{s-1}(r_{\ell_\nu}{+}1{+}r_{\ell_{\nu+1}}) \;+\; r_{\ell_s}{+}1{+}r_j \\ =\;& r_i + r_j + (s{+}1) + \sum_{\nu=1}^{s} 2r_{\ell_\nu}. \end{aligned}$$

Furthermore, as the vertices $w_i, w_{\ell_1}, \dots, w_{\ell_s}, w_j$ are pairwise distinct, $s{+}2 \leqslant k$ and

$$\mathrm{dist}^{T_0}(w_i, w_j) \;\leqslant\; r_i + r_j + (k{-}1) + \sum_{\nu \in \{1,\dots,k\} \setminus \{i,j\}} 2r_\nu. \qquad (2)$$

Now let T_1 be the tree obtained from T_0 as follows: For each $i \in \{1, \dots, k\}$ and each vertex $v \in N^F_{r_i}(w_i)$ with $v \neq w_i$ let $t(i, v)$ be a new node of T_1 that is attached to node w_i by an edge of weight dist$^F(w_i, v)$. Clearly, T_1 is a tree. Using Equation (2), it is straightforward to see that the distance between any two nodes u, u' of T_1 is

$$\mathrm{dist}^{T_1}(u, u') \;\leqslant\; (k{-}1) \;+\; \sum_{i=1}^{k} 2r_i. \qquad (3)$$

Now let T_2 be the *unweighted* tree obtained from T_1 by replacing every edge of weight $c \in \mathbb{N}$ by a path of length c. Using Equation (3), it is straightforward to see that

$$\mathrm{diam}(T_2) \;\leqslant\; (k{-}1) \;+\; \sum_{i=1}^{k} 2r_i. \qquad (4)$$

It is well-known (and easy to see) that rad$(T) \leqslant \frac{\mathrm{diam}(T)+1}{2}$ for any (unweighted, undirected) tree T. . Together with Equation (4) we therefore obtain that

$$\mathrm{rad}(T_2) \;\leqslant\; \lfloor \tfrac{k}{2} \rfloor + \sum_{i=1}^{k} r_i \;=\; f(\bar{r}). \qquad (5)$$

All that remains to do is to show that rad$(F) \leqslant$ rad(T_2). Towards this aim, we construct a surjective homomorphism $h : V(T_2) \to V(F)$ as follows. For any pair w_i, w_j such that $e_{i,j} := \{w_i, w_j\} \in E(T_1)$ choose a path $P_{i,j}$ in U(F) between w_i and w_j of length $d_{i,j}$, where $d_{i,j}$ is the weight of the edge $e_{i,j}$. This path exists by construction of T_1. Analogously, for any $t(i, v)$ and w_i choose a path $P_{t(i,v)}$ between v and w_i in U(F) of length d where d is the weight of the edge between $t(i, v)$ and w_i in T_2. Now define $h : V(T_2) \to V(F)$ such that

- $h(w_i) = w_i$, for all $i \in \{1, \dots, k\}$,
- $h\big(t(i, v)\big) = v$, for all $i \in \{1, \dots, k\}$ and all $v \in V(F)$ and
- h maps the vertices of the simple path between w_i and w_j in T_2 to the vertices of $P_{i,j}$ and the vertices of the simple path between $t(i, v)$ and w_i in T_2 to $P_{t(i,v)}$, so that for all nodes $u, u' \in V(T_2)$, if there is an edge between u and u' in T_2 then there is an edge between $h(u)$ and $h(u')$ in $\mathrm{U}(F)$.

By construction, h is a surjective homomorphism from T_2 to $\mathrm{U}(F)$. In particular, this implies that

$$\mathrm{dist}^{T_2}(u, u') \;\geqslant\; \mathrm{dist}^{F}\big(h(u), h(u')\big) \tag{6}$$

for all nodes u, u' of T_2.

Now let $u_c \in V(T_2)$ be a central vertex for T_2. Let $v_c := h(u_c)$ be the corresponding vertex in F. Now let v be an arbitrary vertex of F. Since h is surjective, there exists an $u \in V(T_2)$ such that $h(u) = v$. Using Equation (6), we obtain that $\mathrm{dist}^F(v, v_c) \leqslant \mathrm{dist}^{T_2}(u, u_c) \leqslant \mathrm{rad}(T_2)$. Thus, $\mathrm{rad}(F) \leqslant \mathrm{rad}(T_2)$. This completes the proof of Lemma 4.5. ☐

We now turn towards the proof of Theorem 4.4. To this end, we first show that if two graphs are Hanf-equivalent with respect to radius R then every neighbourhood type $(F, \bar{w})$ of generalised radius at most R occurs in both graphs for the same number of times. This will be the main technical tool to show that Hanf-equivalence implies subgraph-equivalence for suitable radii. We first prove this intermediate step for connected types $(F, \bar{w})$, see Lemma 4.6, which forms the base step for the general case in Lemma 4.7.

Lemma 4.6 *Let $k \geqslant 1$, let $\bar{r} = (r_1, \dots, r_k) \in \mathbb{N}^k$, and let $\tau = (F, \bar{w})$ be an $\bar{r}$-neighbourhood type with k centres such that F is connected. If G and H are graphs with $G \approx_R^{\mathit{Hanf}} H$ for some $R \geqslant f(\bar{r})$, then $\#_\tau(G) = \#_\tau(H)$.*

Proof. Let G and H be graphs such that $G \approx_R^{\mathit{Hanf}} H$ for some $R \geqslant f(\bar{r})$. Clearly, as $\approx_R^{\mathit{Hanf}}$ is symmetric, it suffices to show that $\#_\tau(G) \geqslant \#_\tau(H)$, i.e. to exhibit a surjective mapping from $\mathit{set}_\tau(G)$ to $\mathit{set}_\tau(H)$.

As $G \approx_R^{\mathit{Hanf}} H$, there is a bijection $\beta : V(G) \to V(H)$ such that for all vertices u of G there exists an isomorphism

$$\pi_u : \big(\mathcal{N}_R^G(u), u\big) \cong \big(\mathcal{N}_R^H(\beta(u)), \beta(u)\big).$$

In particular, $\pi_u(u) = \beta(u)$. Furthermore, for every tuple $\bar{u} = (u_1, \dots, u_k) \in \mathit{set}_\tau(G)$ let us fix an isomorphism $\alpha_{\bar{u}} : (F, \bar{w}) \cong (\mathcal{N}_{\bar{r}}^G(\bar{u}), \bar{u})$. Analogously, for every $\bar{v} = (v_1, \dots, v_k) \in \mathit{set}_\tau(H)$ we fix an isomorphism $\alpha_{\bar{v}} : (F, \bar{w}) \cong (\mathcal{N}_{\bar{r}}^G(\bar{v}), \bar{v})$.

Let $\bar{w} = (w_1, \dots, w_k)$. Since F is connected, Lemma 4.5 implies that $\mathrm{rad}(F) \leqslant f(\bar{r}) \leqslant R$. Let w_0 be a *central* vertex of F. Clearly, since $\mathrm{rad}(F) \leqslant R$, we have $N_R^F(w_0) = V(F) = N_{\bar{r}}^F(\bar{w})$.

We now define a mapping $\gamma : \mathit{set}_\tau(G) \to V(H)^k$ as follows: for every k-tuple $\bar{u} = (u_1, \dots, u_k) \in \mathit{set}_\tau(G)$ we let $u_0 := \alpha_{\bar{u}}(w_0)$ and $\pi := \pi_{u_0}$ and choose

$$\gamma\big((u_1, \dots, u_k)\big) \;:=\; \big(\,\pi(u_1), \dots, \pi(u_k)\,\big).$$

We claim that γ is a surjective map from $\mathit{set}_\tau(G) \to \mathit{set}_\tau(H)$, i.e.

(1) γ is well-defined and $\gamma(\bar{u}) \in \mathit{set}_\tau(H)$ for all $\bar{u} \in \mathit{set}_\tau(G)$ and

(2) for every $\bar{v} \in \mathit{set}_\tau(H)$ there is a $\bar{u} \in \mathit{set}_\tau(G)$ such that $\bar{v} = \gamma(\bar{u})$.

Towards (1), let $\bar{u} = (u_1, \dots, u_k) \in \mathit{set}_\tau(G)$. By construction, $\alpha_{\bar{u}}$ is an isomorphism from F to $\mathcal{N}_{\bar{r}}^G(\bar{u})$ which maps w_i to u_i, for all $i \in \{1, \dots, k\}$. Furthermore, $N_{\bar{r}}^F(\bar{w}) = V(F) = N_R^F(w_0)$. This implies for $u_0 := \alpha_{\bar{u}}(w_0)$ that $N_{\bar{r}}^G(\bar{u}) \subseteq N_R^G(u_0)$.

By assumption, $\pi := \pi_{u_0}$ is an isomorphism from $\mathcal{N}_R^G(u_0)$ to $\mathcal{N}_R^H(\beta(u_0))$ with $\pi(u_0) = \beta(u_0)$. As $N_{\bar{r}}^G(\bar{u}) \subseteq N_R^G(u_0)$, π is defined on $u_1, \dots u_k$, and therefore γ is well-defined.

Furthermore, the restriction of π to the set $N_{\bar{r}}^G(\bar{u})$ is an isomorphism from $\mathcal{N}_{\bar{r}}^G(\bar{u})$ to $\mathcal{N}_{\bar{r}}^H(\pi(u_1), \dots, \pi(u_k))$ and thus witnesses that $\mathit{type}_{\bar{r}}^H(\gamma(\bar{u})) = \tau$, i.e., $\gamma(\bar{u}) \in \mathit{set}_\tau(H)$.

Towards (2) let $\bar{v} = (v_1, \dots, v_k) \in \mathit{set}_\tau(H)$. By construction, $\alpha_{\bar{v}}$ is an isomorphism from F to $\mathcal{N}_{\bar{r}}^H(\bar{v})$ which maps w_i to v_i, for all $i \in \{1, \dots, k\}$. Furthermore, $N_{\bar{r}}^F(\bar{w}) = V(F) = N_R^F(w_0)$. This implies for $v_0 := \alpha_{\bar{v}}(w_0)$ that $N_{\bar{r}}^H(\bar{v}) \subseteq N_R^H(v_0)$.

Letting $u_0 := \beta^{-1}(v_0)$ and $\pi := \pi_{u_0}$, π is an isomorphism from $\mathcal{N}_R^G(u_0)$ to $\mathcal{N}_R^H(v_0)$ with $\pi(u_0) = v_0$. In particular, since $v_1, \dots, v_k \in N_R^H(v_0)$, we can choose $u_i := \pi^{-1}(v_i)$ for all $i \in \{1, \dots, k\}$. Clearly, for $\bar{u} := (u_1, \dots, u_k)$, the restriction of π to $N_{\bar{r}}^G(\bar{u})$ is an isomorphism from $\mathcal{N}_{\bar{r}}^G(\bar{u})$ to $\mathcal{N}_{\bar{r}}^H(\bar{v})$ which maps u_i to v_i, for all $i \in \{1, \dots, k\}$. Thus, $\mathit{type}_{\bar{r}}^G(\bar{u}) = \tau$ and $\gamma(\bar{u}) = \bar{v}$. In summary, we obtain that γ is a surjective mapping from $\mathit{set}_\tau(G)$ to $\mathit{set}_\tau(H)$. This completes the proof of Lemma 4.6. ☐

In the next lemma we generalise the previous lemma to the case that F is not connected.

Lemma 4.7 *Let $R \geqslant 0$ and let G and H be graphs such that $G \approx_R^{\mathit{Hanf}} H$. For all $k \geqslant 1$ and all $\bar{r} = (r_1, \dots, r_k) \in \mathbb{N}^k$ with $f(\bar{r}) \leqslant R$ we have $\#_\tau(G) = \#_\tau(H)$, for all $\bar{r}$-neighbourhood types τ with k centres.*

Proof. The proof proceeds by induction on the number c of connected components of τ.

Induction base: $c = 1$. The case where $\tau = (F, \bar{w})$ is an $\bar{r}$-neighbourhood type with k centres such that F is connected was already established in Lemma 4.6.

Induction step: $c \to c{+}1$. We can assume the following

> **Induction hypothesis:** $\#_{\tau'}(G) = \#_{\tau'}(H)$, for all $k' \geqslant 1$, all $\bar{r}' = (r'_1, \dots, r'_{k'}) \in \mathbb{N}^{k'}$ with $f(\bar{r}') \leqslant R$ and all $\bar{r}'$-neighbourhood types $\tau' = (F', \bar{w}')$ with k' centres such that F' has at most c connected components.

Let $k \geqslant 1$ and $\bar{r} = (r_1, \dots, r_k) \in \mathbb{N}^k$ with $f(\bar{r}) \leqslant R$. Let $\tau = (F, \bar{w})$ be an $\bar{r}$-neighbourhood type with k centres such that F consists of $c{+}1$ connected components. Our goal is to show that $\#_\tau(G) = \#_\tau(H)$.

Let $\bar{w} = (w_1, \dots, w_k)$. Let F_1 be the connected component of F that contains vertex w_1, and let F_2 be the graph obtained from F by removing F_1. Then

- F is the disjoint union of F_1 and F_2,
- F_1 is connected and
- F_2 consists of c connected components.

For a set $I \subseteq \{1, \dots, k\}$ and a k-tuple $\bar{a} = (a_1, \dots, a_k)$ we write $\bar{a}_I$ to denote the tuple of length $|I|$ obtained from $\bar{a}$ by deleting all components that do not belong to I. Let

$$\begin{aligned} I_1 \;&:=\; \{i \in \{1, \dots, k\} \,:\, \text{vertex } w_i \text{ belongs to } F_1\}, \\ I_2 \;&:=\; \{1, \dots, k\} \setminus I_1 \\ &=\; \{i \in \{1, \dots, k\} \,:\, \text{vertex } w_i \text{ belongs to } F_2\}, \end{aligned}$$

and consider the neighbourhood types $\tau_1 := (F_1, \bar{w}_{I_1})$ and $\tau_2 := (F_2, \bar{w}_{I_2})$. Clearly,

$$\#_\tau(G) = |\{\bar{u} \in V(G)^k : \textit{type}_{\bar{r}}^G(\bar{u}) = \tau\}| = \left| \left\{ \bar{u} \in V(G)^k : \begin{array}{c} \textit{type}_{\bar{r}}^G(\bar{u}) = \tau \text{ and} \\ \textit{type}_{\bar{r}_{I_1}}^G(\bar{u}_{I_1}) = \tau_1 \text{ and} \\ \textit{type}_{\bar{r}_{I_2}}^G(\bar{u}_{I_2}) = \tau_2 \end{array} \right\} \right| \quad (7)$$

and

$$\#_\tau(H) = |\{\bar{v} \in V(H)^k : \textit{type}_{\bar{r}}^H(\bar{v}) = \tau\}| = \left| \left\{ \bar{v} \in V(H)^k : \begin{array}{c} \textit{type}_{\bar{r}}^H(\bar{v}) = \tau \text{ and} \\ \textit{type}_{\bar{r}_{I_1}}^H(\bar{v}_{I_1}) = \tau_1 \text{ and} \\ \textit{type}_{\bar{r}_{I_2}}^H(\bar{v}_{I_2}) = \tau_2 \end{array} \right\} \right|. \quad (8)$$

Note that, for $i \in \{1,2\}$, τ_i is an $\bar{r}_{I_i}$-neighbourhood type with $k_i := |I_i|$ centres and $f(\bar{r}_{I_i}) \leqslant f(\bar{r}) \leqslant R$. Since F_i has at most c connected components, by induction hypothesis, $\#_{\tau_i}(G) = \#_{\tau_i}(H)$ for each $i \in \{1,2\}$. Since $\#_{\tau_i}(G) = |\textit{set}_{\tau_i}(G)|$ and $\#_{\tau_i}(H) = |\textit{set}_{\tau_i}(H)|$, we obtain

$$\begin{aligned} & \left| \left\{ \bar{u} \in V(G)^k : \begin{array}{c} \textit{type}_{\bar{r}_{I_1}}^G(\bar{u}_{I_1}) = \tau_1 \text{ and} \\ \textit{type}_{\bar{r}_{I_2}}^G(\bar{u}_{I_2}) = \tau_2 \end{array} \right\} \right| \\ &= |\textit{set}_{\tau_1}(G) \times \textit{set}_{\tau_2}(G)| \\ &= |\textit{set}_{\tau_1}(H) \times \textit{set}_{\tau_2}(H)| \\ &= \left| \left\{ \bar{v} \in V(H)^k : \begin{array}{c} \textit{type}_{\bar{r}_{I_1}}^H(\bar{v}_{I_1}) = \tau_1 \text{ and} \\ \textit{type}_{\bar{r}_{I_2}}^H(\bar{v}_{I_2}) = \tau_2 \end{array} \right\} \right|. \end{aligned} \quad (9)$$

Now let T be the set of all $\bar{r}$-neighbourhood types $\tau' = (F', \bar{w}')$ with k centres such that:

(1) $\textit{type}_{\bar{r}_{I_1}}^{F'}(\bar{w}'_{I_1}) = \tau_1$,

(2) $\textit{type}_{\bar{r}_{I_2}}^{F'}(\bar{w}'_{I_2}) = \tau_2$, and

(3) F' consists of at most c connected components.

It is straightforward to see that

$$\left\{ \bar{u} \in V(G)^k : \begin{array}{c} \textit{type}_{\bar{r}_{I_1}}^G(\bar{u}_{I_1}) = \tau_1 \text{ and} \\ \textit{type}_{\bar{r}_{I_2}}^G(\bar{u}_{I_2}) = \tau_2 \text{ and} \\ \textit{type}_{\bar{r}}^G(\bar{u}) \neq \tau \end{array} \right\} = \dot{\bigcup_{\tau' \in T}} \{\bar{u} \in V(G)^k : \textit{type}_{\bar{r}}^G(\bar{u}) = \tau'\} \quad (10)$$

and

$$\left\{ \bar{v} \in V(H)^k : \begin{array}{c} \textit{type}_{\bar{r}_{I_1}}^H(\bar{v}_{I_1}) = \tau_1 \text{ and} \\ \textit{type}_{\bar{r}_{I_2}}^H(\bar{v}_{I_2}) = \tau_2 \text{ and} \\ \textit{type}_{\bar{r}}^H(\bar{v}) \neq \tau \end{array} \right\} = \dot{\bigcup_{\tau' \in T}} \{\bar{v} \in V(H)^k : \textit{type}_{\bar{r}}^H(\bar{v}) = \tau'\}, \quad (11)$$

where "$\dot{\bigcup}$" indicates that this is a union of pairwise disjoint sets. By the induction hypothesis, $\#_{\tau'}(G) = \#_{\tau'}(H)$ for all $\tau' \in T$. Hence, for every $\tau' \in T$,

$$\begin{aligned} & |\{\bar{u} \in V(G)^k : \textit{type}_{\bar{r}}^G(\bar{u}) = \tau'\}| = \#_{\tau'}(G) \\ = \; & \#_{\tau'}(H) = |\{\bar{v} \in V(H)^k : \textit{type}_{\bar{r}}^H(\bar{v}) = \tau'\}|. \end{aligned} \quad (12)$$

The equations (10), (11), and (12) imply that

$$\left| \left\{ \bar{u} \in V(G)^k : \begin{array}{c} \textit{type}_{\bar{r}_{I_1}}^G(\bar{u}_{I_1}) = \tau_1 \text{ and} \\ \textit{type}_{\bar{r}_{I_2}}^G(\bar{u}_{I_2}) = \tau_2 \text{ and} \\ \textit{type}_{\bar{r}}^G(\bar{u}) \neq \tau \end{array} \right\} \right| = \left| \left\{ \bar{v} \in V(H)^k : \begin{array}{c} \textit{type}_{\bar{r}_{I_1}}^H(\bar{v}_{I_1}) = \tau_1 \text{ and} \\ \textit{type}_{\bar{r}_{I_2}}^H(\bar{v}_{I_2}) = \tau_2 \text{ and} \\ \textit{type}_{\bar{r}}^H(\bar{v}) \neq \tau \end{array} \right\} \right| \quad (13)$$

The equations (13) and (9) imply that

$$\left| \left\{ \bar{u} \in V(G)^k : \begin{array}{c} \textit{type}_{\bar{r}_{I_1}}^G(\bar{u}_{I_1}) = \tau_1 \text{ and} \\ \textit{type}_{\bar{r}_{I_2}}^G(\bar{u}_{I_2}) = \tau_2 \text{ and} \\ \textit{type}_{\bar{r}}^G(\bar{u}) = \tau \end{array} \right\} \right| = \left| \left\{ \bar{v} \in V(H)^k : \begin{array}{c} \textit{type}_{\bar{r}_{I_1}}^H(\bar{v}_{I_1}) = \tau_1 \text{ and} \\ \textit{type}_{\bar{r}_{I_2}}^H(\bar{v}_{I_2}) = \tau_2 \text{ and} \\ \textit{type}_{\bar{r}}^H(\bar{v}) = \tau \end{array} \right\} \right| \quad (14)$$

From the equations (14), (7), and (8) we obtain that $\#_\tau(G) = \#_\tau(H)$. This completes the proof of Lemma 4.7. □

Next, we will transfer the result of Lemma 4.7 to the notion of subgraph-equivalence w.r.t. generalised radius r.

Lemma 4.8 *Let $R \geqslant 0$ and let G and H be graphs such that $G \approx_R^{\textit{Hanf}} H$. Let S be a graph. If* $\text{genrad}(S) \leqslant R$ *then $\#_S(G) = \#_S(H)$.*

Proof. W.l.o.g. we assume that $V(S) = [s]$ for some $s \geqslant 1$. Let $k \geqslant 1$ be the number of connected components of S and let $S_1, \ldots, S_k$ be the connected components of S of radii $r_1, \ldots, r_k$. For each $i \in \{1, \ldots, k\}$ let $c_i \in V(S)$ be a central vertex for S_i — these vertices $c_1, \ldots, c_k$ will be fixed throughout the remainder of this proof. Let $\bar{r} = (r_1, \ldots, r_k)$ and $\bar{c} = (c_1, \ldots, c_k)$. Clearly, $N_{\bar{r}}^S(\bar{c}) = V(S)$.

An $\bar{r}$-neighbourhood type $\tau = (F, \bar{w})$ with k centres is called *compatible* with $(S, \bar{c})$ if there is an embedding $\bar{z} = (z_1, \ldots, z_s) \subseteq V(F)^s$ of S in F such that $z_{c_i} = w_i$ for all $i \in \{1, \ldots, k\}$. Recall that we represent embeddings of S by s-tuples of elements in F, see the remark after Definition 4.1.

We let T be the set of all $\bar{r}$-types with k centres which are compatible with $(S, \bar{c})$. Clearly, $f(\bar{r}) = \text{genrad}(S) \leqslant R$. Hence, we obtain from Lemma 4.7 that $\#_\tau(G) = \#_\tau(H)$, for all $\tau \in T$. I.e., for each $\tau \in T$, there is a bijection $\gamma_\tau : \textit{set}_\tau(G) \to \textit{set}_\tau(H)$. Thus, for each $\tau \in T$ and each $\bar{u} \in \textit{set}_\tau(G)$ and for $\bar{v} := \gamma_\tau(\bar{u})$ we know that $\textit{type}_{\bar{r}}^G(\bar{u}) = \tau$ and $\textit{type}_{\bar{r}}^H(\bar{v}) = \tau$, and hence there is an isomorphism $\pi_{\bar{u}} : \big(\mathcal{N}_{\bar{r}}^G(\bar{u}), \bar{u}\big) \cong \big(\mathcal{N}_{\bar{r}}^H(\bar{v}), \bar{v}\big)$ — this isomorphism $\pi_{\bar{u}}$ will be fixed throughout the remainder of this proof.

Our goal is to show that $\#_S(G) = \#_S(H)$. Note that, by symmetry (since $G \approx_R^{\textit{Hanf}} H$ iff $H \approx_R^{\textit{Hanf}} G$), it suffices to show that $\#_S(G) \geqslant \#_S(H)$.

We define a mapping $\beta : \textit{emb}_S(G) \to V(H)^s$ as follows: For every $\bar{x} = (x_1, \ldots, x_s) \in \textit{emb}_S(G)$ we let

$$\bar{u} := (u_1, \ldots, u_k) := \big(x_{c_1}, \ldots, x_{c_k}\big),$$

and we let $\tau = \textit{type}_{\bar{r}}^G(\bar{u})$ be the $\bar{r}$-neighbourhood type (with k centres) of $\bar{u}$ in G. Since $\bar{x} \in \textit{emb}_S(G)$, we know that τ is compatible with $(S, \bar{c})$, and thus $\tau \in T$. Therefore, we can choose $\pi := \pi_{\bar{u}}$ and let

$$\beta(\bar{x}) := \big(\pi(x_1), \ldots, \pi(x_s)\big).$$

To conclude the proof, it suffices to show that β is a surjective mapping from $\textit{emb}_S(G)$ to $\textit{emb}_S(H)$, i.e. that

(1) β is well-defined, $\beta(\bar{x}) \in emb_S(H)$ for all $\bar{x} \in emb_S(G)$, and

(2) for every $\bar{y} \in emb_S(H)$ there is a $\bar{x} \in emb_S(G)$ such that $\bar{y} = \beta(\bar{x})$.

Concerning (1) let $\bar{x} = (x_1, \ldots, x_s) \in emb_S(G)$ and let $\bar{u} := (x_{c_1}, \ldots, x_{c_k})$ be its centres. Let $\tau := type^G_{\bar{r}}(\bar{u})$ and let $\pi := \pi_{\bar{u}}$. We know that π is an isomorphism witnessing that $(\mathcal{N}^G_{\bar{r}}(\bar{u}), \bar{u}) \cong (\mathcal{N}^H_{\bar{r}}(\bar{v}), \bar{v})$ for $\bar{v} := \gamma_\tau(\bar{u})$.

Since $\bar{x}$ is an embedding of S in G, $V(S) = N^S_{\bar{r}}(\bar{c})$, and $(x_{c_1}, \ldots, x_{c_s}) = \bar{u}$, we know that $\{x_1, \ldots, x_s\} \subseteq N^G_{\bar{r}}(\bar{u})$. Thus, π is defined on $x_1, \ldots, x_s$, and hence β is well-defined.

Furthermore, since π is an isomorphism from $\mathcal{N}^G_{\bar{r}}(\bar{u})$ to $\mathcal{N}^H_{\bar{r}}(\bar{v})$, for all $i, j \in \{1, \ldots, s\}$ there is an edge from x_i to x_j in G if, and only if, there is an edge from $\pi(x_i)$ to $\pi(x_j)$ in H. Thus, since $\bar{x}$ is an embedding of S in G, also $\bar{y} = (\pi(x_1), \ldots, \pi(x_s))$ is an embedding of S in H, i.e., $\bar{y} \in emb_S(H)$.

Concerning (2) let $\bar{y} = (y_1, \ldots, y_s) \in emb_S(H)$. Let $\bar{v} := (y_{c_1}, \ldots, y_{c_k})$ and let $\tau = type^G_{\bar{r}}(\bar{v})$ be the $\bar{r}$-neighbourhood type (with k centres) of $\bar{v}$ in H. Since $\bar{y} \in emb_S(H)$, we know that τ is compatible with $(S, \bar{c})$, and hence $\tau \in T$ and $\{y_1, \ldots, y_s\} \subseteq N^H_{\bar{r}}(\bar{v})$. We choose $\bar{u} = (u_1, \ldots, u_k) := \gamma_\tau^{-1}(\bar{v})$. We let $\pi := \pi_{\bar{u}}$ and choose $\bar{x} = (x_1, \ldots, x_s)$ with $x_i := \pi^{-1}(y_i)$ for every $i \in \{1, \ldots, s\}$. Clearly, $\beta(\bar{x}) = \bar{y}$. This concludes the proof of Lemma 4.8. □

Now we are ready for the proof of Theorem 4.4.

Proof of Theorem 4.4.
For the "only if" direction let $G \approx^{Hanf}_r H$. Our goal is to show that $G \approx^{subgraph}_{genrad\ r} H$.
To this end, let S be a graph on vertex set $V(S) = [s]$ for $s \geqslant 1$, such that genrad$(S) \leqslant r$. From Lemma 4.8 we obtain that $\#_S(G) = \#_S(H)$. This shows that $G \approx^{subgraph}_{genrad\ r} H$.

For the "if" direction let $G \approx^{subgraph}_{genrad\ r} H$. Our goal is to show that $G \approx^{Hanf}_r H$, i.e. that $\#_{(F,w)}(G) = \#_{(F,w)}(H)$ for all neighbourhood types (F, w) with one centre and radius $\leqslant r$. W.l.o.g. we will always assume that $V(F) = \{1, \ldots, |V(F)|\}$ and $w = 1$.

Let T be a set of all such r-neighbourhood types which for each isomorphism type only contains one copy, i.e. such that for all $\varrho = (F, 1) \in T$ and $\varrho' = (F', 1) \in T$, if $(F, 1) \cong (F', 1)$, then $F = F'$. Our goal is to show that $\#_\varrho(G) = \#_\varrho(H)$ for all $\varrho \in T$.

For each $\varrho = (F, 1) \in T$ we know that $V(F) = N^F_r(1)$, and hence rad$(F) \leqslant r$. Furthermore, as F is connected, genrad$(F) =$ rad$(F) \leqslant r$. Thus, $G \approx^{subgraph}_{genrad\ r} H$ implies that $\#_F(G) = \#_F(H)$. Note that this does not immediately imply that $\#_\varrho(G) = \#_\varrho(H)$ for all neighbourhood types $\varrho = (F, w)$ of radius at most r as neighbourhood types have a distinguished element w. We therefore proceed as follows.

The *size* of $\varrho = (F, 1)$ is defined to be $|V(F)|$. Let $\varrho = (F, 1) \in T$ be of size s. We say that an r-neighbourhood type $\varrho' = (F', 1) \in T$ is *compatible* with ϱ if there exists at least one embedding $(z_1, \ldots, z_s) \in V(F')^s$ of F in F' with $z_1 = 1$ (i.e., the embedding associates the centre node 1 of F with the centre node 1 of F'). Clearly, the following is true:

> If $\varrho' = (F', 1) \in T$ is compatible with $\varrho = (F, 1) \in T$, then either $(F', 1) \cong (F, 1)$ (and thus $\varrho' = \varrho$, according to our choice of T) or the size of ϱ' is strictly larger than the size of ϱ. (15)

For $\varrho \in T$ we write $comp(\varrho)$ to denote the set of all $\varrho' \in T$ which are compatible with ϱ.
For $\varrho = (F, 1) \in T$ and $\varrho' = (F', 1) \in T$ we let $set_\varrho(\varrho')$ be the set of all embeddings $(z_1, \ldots, z_s) \in V(F')^s$ of F in F' with $z_1 = 1$, where s is the size of ϱ.

Claim 1 *For every $\varrho = (F, 1) \in T$ and every graph G the following is true:*

$$\#_F(G) \;=\; \sum_{\varrho' \in comp(\varrho)} \#_{\varrho'}(G) \cdot \#_\varrho(\varrho').$$

Proof of the claim.
Let s be the size of ϱ. By definition, we know that

$$\#_F(G) \;=\; |\{\bar{x} = (x_1, \ldots, x_s) \in V(G)^s \;:\; \bar{x} \in emb_F(G)\}|.$$

We let U be the set of all $u \in V(G)$ for which there exists an embedding $(x_1, \ldots, x_s) \in emb_F(G)$ such that $x_1 = u$. For every $u \in U$ we let X_u be the set of all $(x_1, \ldots, x_s) \in emb_F(G)$ with $x_1 = u$. Clearly,

$$\#_F(G) \;=\; \sum_{u \in U} |X_u|.$$

Let T' be the set of all $\varrho' \in T$ for which there exists an $u \in U$ whose r-neighbourhood type in G is ϱ'. It is straightforward to see that the following is true for each $\varrho' \in T'$:

- ϱ' is compatible with ϱ.
- For each $v \in V(G)$ of r-neighbourhood type ϱ' we have $v \in U$.
- For each $u \in U$ of r-neighbourhood type ϱ' we have $|X_u| = \#_\varrho(\varrho')$.

Thus,

$$\begin{aligned} \#_F(G) \;&=\; \textstyle\sum_{u \in U} |X_u| \;=\; \sum_{\varrho' \in T'} \sum_{\substack{v \in V(G) \\ \text{of type } \varrho'}} \#_\varrho(\varrho') \\ &=\; \textstyle\sum_{\varrho' \in T'} \#_{\varrho'}(G) \cdot \#_\varrho(\varrho'). \end{aligned}$$

We already know that $T' \subseteq comp(\varrho)$. On the other hand, for all $\varrho' = (F', 1) \in comp(\varrho)$ with $\#_{\varrho'}(G) \neq 0$, there must be a $v \in V(G)$ of r-neighbourhood type ϱ'. Since ϱ' is compatible with ϱ, we obtain that there is an embedding $(x_1, \ldots, x_s) \in V(G)^s$ of F in G with $x_1 = v$. Thus, $v \in U$ and $\varrho' \in T'$. Therefore, we obtain that $\#_{\varrho'}(G) = 0$ for all $\varrho' \in comp(\varrho) \setminus T'$. In summary, this leads to

$$\begin{aligned} \#_F(G) \;&=\; \textstyle\sum_{\varrho' \in T'} \#_{\varrho'}(G) \cdot \#_\varrho(\varrho') \\ &=\; \textstyle\sum_{\varrho' \in comp(\varrho)} \#_{\varrho'}(G) \cdot \#_\varrho(\varrho'). \end{aligned}$$

This completes the proof of the claim. ⊣

Recall that we already know that $\#_F(G) = \#_F(H)$ for all $\varrho = (F, 1) \in T$. Our aim is to show that $\#_\varrho(G) = \#_\varrho(H)$. We proceed by backward induction on the size of ϱ.

Induction base:
Let $s_{\max}$ be the maximum size of an r-neighbourhood type ϱ (with one centre) with $\#_\varrho(G) \geqslant 1$ or $\#_\varrho(H) \geqslant 1$. Hence, for all $\varrho \in T$ of size $s > s_{\max}$ we have $\#_\varrho(G) = \#_\varrho(H) = 0$.

Induction step: $s{+}1 \to s$
By the induction hypothesis we know that $\#_{\varrho'}(G) = \#_{\varrho'}(H)$ for all $\varrho' \in T$ of size at least $s{+}1$.
Now let $\varrho = (F, 1) \in T$ be of size s. Our goal is to show that $\#_\varrho(G) = \#_\varrho(H)$. By the claim above we know that

$$\#_F(G) = \textstyle\sum_{\varrho' \in comp(\varrho)} \#_{\varrho'}(G) \cdot \#_\varrho(\varrho') \qquad \text{and}$$
$$\#_F(H) = \textstyle\sum_{\varrho' \in comp(\varrho)} \#_{\varrho'}(H) \cdot \#_\varrho(\varrho').$$

Furthermore, from (15) we know that all $\varrho' \in comp(\varrho)$ with $\varrho' \neq \varrho$ have size at least $s{+}1$. Hence, the induction hypothesis tells us that

$\#_{\varrho'}(G) = \#_{\varrho'}(H)$ for all $\varrho' \in \mathit{comp}(\varrho) \setminus \{\varrho\}$.
In summary, we thus have

$$\begin{aligned} \#_{\varrho}(G) \cdot \#_{\varrho}(\varrho) &= \#_F(G) - \sum_{\varrho' \in \mathit{comp}(\varrho)\setminus\{\varrho\}} \#_{\varrho'}(G) \cdot \#_{\varrho'}(\varrho) \\ &= \#_F(H) - \sum_{\varrho' \in \mathit{comp}(\varrho)\setminus\{\varrho\}} \#_{\varrho'}(H) \cdot \#_{\varrho'}(\varrho) \\ &= \#_{\varrho}(H) \cdot \#_{\varrho}(\varrho). \end{aligned}$$

Since $\#_{\varrho}(\varrho) \geqslant 1$, we obtain that $\#_{\varrho}(G) = \#_{\varrho}(H)$.
This completes the proof of Theorem 4.4. □

5. Subgraph-equivalence with respect to the order of the considered subgraphs

In the previous section we have established a tight relation between Hanf-equivalence and subgraph-equivalence with respect to the generalised radius. Instead of using the generalised radius, subgraph-equivalence can also naturally be defined with respect to the order (i.e. the number of vertices) of the subgraphs. This concept is the object of study in this section and again we show that Hanf-equivalence implies subgraph-equivalence with respect to the order. We first formally define this alternative notion of subgraph-equivalence.

Definition 5.1 Let $s \geqslant 1$. Two graphs G and H are *subgraph-equivalent w.r.t. order s*, in symbols:

$$G \approx^{\textit{subgraph}}_{\textit{order } s} H,$$

if for all graphs S of order at most s we have $\#_S(G) = \#_S(H)$.

The goal of this section is to determine the relation between Hanf-equivalence with respect to radius r and subgraph-equivalence with respect to order s. We will do so by reducing this question to the results of the previous section. As a first step towards our aim we derive an upper bound for the generalised radius of a graph with respect to its order. As we will show below (see Example 5.3), this bound is tight.

Lemma 5.2 *Let S be a graph of order $s := |V(S)|$. Then* $\text{genrad}(S) \leqslant \lfloor 3s/4 \rfloor$.

Proof. Let $k \geqslant 1$ be the number of connected components of S and let $r_1 \leqslant \cdots \leqslant r_k$ be the radii of these connected components. By definition we have $\text{genrad}(S) = \big(\sum_{i=1}^k r_i\big) + \lfloor \frac{k}{2} \rfloor$. Let $j \geqslant 0$ be such that $r_i = 0$ for all $i \leqslant j$ and $r_i \geqslant 1$ for all $i \geqslant j{+}1$.

Observe that a connected graph of radius $r_i = 0$ contains exactly 1 node, and a connected graph of radius $r_i \geqslant 1$ contains at least $2r_i$ nodes. To see the latter, note that the worst case is a graph consisting of a path on $2r_i$ nodes $v_1, \ldots v_{2r_i}$. Then, each of the nodes v_{r_i} and v_{r_i+1} is a central node, and the radius of this graph is exactly r_i.

By this observation we know that $|V(S_i)| = 1$ for all $i \leqslant j$ and $|V(S_i)| \geqslant 2r_i$ for all $i \geqslant j{+}1$. Thus,

$$s = |V(S)| = \sum_{i=1}^k |V(S_i)| \geqslant j + \sum_{i=j+1}^k 2r_i = j + 2\sum_{i=1}^k r_i,$$

and hence $\sum_{i=1}^k r_i \leqslant \lfloor \frac{s-j}{2} \rfloor$. Furthermore, since $r_i \geqslant 1$ for all $i \geqslant j{+}1$, we have

$$s \geqslant j + \sum_{i=j+1}^k 2r_i \geqslant j + \sum_{i=j+1}^k 2 = j + 2(k-j) = 2k - j,$$

and hence $k \leqslant (s+j)/2$, i.e., $\lfloor \frac{k}{2} \rfloor \leqslant \lfloor \frac{s+j}{4} \rfloor$. In summary,

$$\begin{aligned} \text{genrad}(S) &\leqslant \lfloor \tfrac{s-j}{2} \rfloor + \lfloor \tfrac{s+j}{4} \rfloor \leqslant \lfloor \tfrac{s-j}{2} + \tfrac{s+j}{4} \rfloor \\ &= \lfloor \tfrac{3s-j}{4} \rfloor \leqslant \lfloor \tfrac{3s}{4} \rfloor. \end{aligned}$$

This completes the proof of Lemma 5.2. □

The following example shows that the bound achieved in Lemma 5.2 is optimal.

Example 5.3 *For $k \geqslant 1$ let S be the graph consisting of k connected components, each of which consists of a single edge. Then, S is of order $s = |V(S)| = 2k$ and*

$$\text{genrad}(S) = \Big(\sum_{i=1}^k r_i\Big) + \lfloor \tfrac{k}{2} \rfloor = k + \lfloor \tfrac{k}{2} \rfloor.$$

If $k = 2k'$ for $k' \in \mathbb{N}_{\geqslant 1}$, then $\text{genrad}(S) = 3k' = \lfloor 3s/4 \rfloor$.

Using Theorem 4.4 and Lemma 5.2, we immediately obtain the following consequence.

Corollary 5.4 *Let $s \geqslant 1$ and let $r \geqslant \lfloor 3s/4 \rfloor$. If G and H are graphs with $G \approx^{\textit{Hanf}}_r H$, then $G \approx^{\textit{subgraph}}_{\textit{order } s} H$.*

Proof. Let G and H be graphs with $G \approx^{\textit{Hanf}}_r H$. By Theorem 4.4, $G \approx^{\textit{subgraph}}_{\textit{genrad } r} H$. I.e., for all graphs S with $\text{genrad}(S) \leqslant r$ we have $\#_S(G) = \#_S(H)$.

Furthermore, by Lemma 5.2, $\text{genrad}(S) \leqslant \lfloor 3s/4 \rfloor \leqslant r$ for all graphs S of order at most s. Thus, $\#_S(G) = \#_S(H)$ for all graphs S of order at most s, i.e., $G \approx^{\textit{subgraph}}_{\textit{order } s} H$. □

The next example shows that Corollary 5.4 cannot be strengthened to any $r \leqslant \lfloor 3s/8 \rfloor$.

Example 5.5 *Recall from Example 3.2 and Example 3.4 that C_n is a directed cycle of n nodes, and $D_{n,n}$ is the graph consisting of two disjoint unions of C_n.*

Recall that, for a graph G, a set $E \subseteq E(G)$ is an induced matching *if the subgraph of G induced by the (endpoints of the) edges in E has no connected component containing more than one edge. For example, for any $n \geqslant 2$,*

$$\begin{aligned} E_n &:= \{(1,2),(4,5),(7,8),\ldots\} \\ &= \{(3i{-}2, 3i{-}1) : 1 \leqslant i \leqslant \lfloor n/3 \rfloor\} \end{aligned}$$

is a maximal induced matching for C_n. Hence, the maximum size of an induced matching in C_n is $\lfloor n/3 \rfloor$.

For each $k \geqslant 1$ let T_k be the graph consisting of k connected components, each of which consists of a single directed edge. Clearly, for any graph G we have $\#_{T_k}(G) \geqslant 1$ iff the maximum size of an induced matching in G is at least k.

Now let $k = 2k'{+}1$ for $k' \geqslant 1$, let $n = 3k'{+}2$, and let G be the graph $D_{n,n}$. Note that any induced matching E in G has size $\leqslant 2k'$, since E can contain at most k' edges of each of the two connected components of G. Thus, since $2k' < k$ we have $\#_{T_k}(G) = 0$.

Furthermore, let H be the graph C_{2n}, i.e., H is the directed cycle on $2n = 2 \cdot (3k'+2) = 3k+1$ nodes. Since $\lfloor 2n/3 \rfloor = k$, we know that the maximum size of an induced matching of H is k, and hence $\#_{T_k}(H) \geqslant 1$. Thus, $\#_{T_k}(G) \neq \#_{T_k}(H)$, and hence for $s = |V(T_k)| = 2k$ we have $G \not\approx^{\textit{subgraph}}_{\textit{order } s} H$.

Recall that $H = C_{2n}$ and $G = D_{n,n}$ for $n = 3k'{+}2$. As noted in Example 3.4, $G \approx^{\textit{Hanf}}_r H$ for $r = \lfloor 3k'/2 \rfloor$. If $k' = 2k''$ for $k'' \geqslant 1$, we have $r = 3k''$ and $s = 2k = 4k'{+}2 = 8k''{+}2$. Thus, $\lfloor 3s/8 \rfloor = \lfloor 3k'' + (3/4) \rfloor = 3k'' = r$.

In summary, for $r = \lfloor 3s/8 \rfloor$, G and H witness that $G \approx^{\textit{Hanf}}_r H$, but $G \not\approx^{\textit{subgraph}}_{\textit{order } s} H$.

6. Conclusion

In this paper we proved a strong relation between the logical notion of Hanf-equivalence and the graph-theoretical concept of subgraph-equivalence, with respect to the generalised radius as well as the order of subgraphs.

This result has implications to the equivalence of graphs under purely existential formulas. For example, we immediately obtain the following corollary, where, by *union of conjunctive query* we mean a finite disjunction of first-order sentences of the form $\exists x_1 \cdots \exists x_s\, \psi(x_1, \ldots, x_s)$, where ψ is a conjunction of atoms and negated atoms.

Corollary 6.1 *Let G, H be graphs, $s \geqslant 1$, and $r \geqslant \lfloor 3s/4 \rfloor$. If $G \approx_r^{Hanf} H$ then G and H satisfy the same unions of conjunctive queries of quantifier rank at most s.*

Proof. By Corollary 5.4, as $G \approx_r^{Hanf} H$, we have $G \approx_{order\ s}^{subgraph} H$. Let φ be a union of conjunctive queries of quantifier rank at most s. W.l.o.g. we may assume that φ is of the form $\bigvee_{i=1}^{t} \varphi_i$ with $\varphi_i = \exists x_1 \cdots \exists x_s\, \psi_i$, where ψ_i is a maximal conjunction of literals, i.e., ψ_i is a conjunction of atomic or negated atomic formulas so that for all pairs x_ℓ, x_m of variables either $E(x_\ell, x_m)$ or $\neg E(x_\ell, x_m)$ and either $x_\ell{=}x_m$ or $\neg x_\ell{=}x_m$ occurs in ψ_i.

It follows that each satisfiable φ_i defines, up to isomorphism, a graph on at most s vertices. As the number of such graphs is the same in G and H, it follows that the sentence φ_i is true in G if, and only if, it is true in H. Thus, $G \models \varphi$ if, and only if, $H \models \varphi$. □

In fact, the proof of the corollary shows that not only φ_i is true in G if, and only if, it is true in H, but the number of witnesses is the same in G and H. It is likely, therefore, that the previous result can be generalised from unions of conjunctive queries to suitable counting versions of existential first-order logic, for instance to the very general counting logics studied in [13, 16]. Furthermore, aggregate functions, as they are commonly used in relational database querying, can be formalised in such counting logics [14]. Hence, it could be possible to use our result for proving expressivity bounds for fragments of languages with aggregation. We defer these generalisations to the full version of this paper.

Acknowledgments

We would like to thank the anonymous referees for their valuable comments. Furthermore, the second author would like to thank Mario Szegedy for an inspiring discussion on Hanf- and subgraph-equivalence and for posing the question that has led to the research presented in this paper.

References

[1] L. Babai. Automorphism groups, isomorphism, reconstruction (chapter 27). In R. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics – Volume 2*, pages 1447–1540. Elsevier, 1995.

[2] J. Bang-Jensen and G. Z. Gutin. *Digraphs – Theory, Algorithms and Applications*. Springer, 2nd edition, 2010.

[3] C. Berge. *Graphs and Hypergraphs*. North-Holland, 1973.

[4] S. A. Choudam. A simple proof of the Erdos-Gallai theorem on graphic sequences. *Bull. Austral. Math. Soc.*, 36:67–70, 1986.

[5] R. Diestel. *Graph Theory*. Springer, 4th edition, 2010.

[6] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*. Springer, 1999.

[7] P. Erdős and T. Gallai. Graphs with prescribed degrees of vertices. *Mat. Lapok*, 11:264–274, 1960.

[8] S. Evdokimov, M. Karpinski, and I. Ponomarenko. On a new high dimensional Weisfeiler-Lehman algorithm. *J. Algebraic Combinatorics*, 10:29–45, 1999.

[9] M. Grohe and O. Verbitsky. Testing graph isomorphism in parallel by playing a game. In *Proc. 33rd International Colloquium on Automata, Languages and Programming (ICALP'06)*, volume 4051 of *Lecture Notes in Computer Science*, pages 3–14. Springer, 2006.

[10] S. Hakimi. On the realizability of a set of integers as the degrees of the vertices of a graph. *SIAM J. Appl. Math.*, 10:496–506, 1962.

[11] W. Hanf. Model-theoretic methods in the study of elementary logic. In J. Addison, L. Henkin, and A. Tarski, editors, *The Theory of Models*, pages 132–145. North-Holland, 1965.

[12] M. Havel. A remark on the existence of finite graphs. *Casopis Pest. Mat.*, 80:477–480, 1955.

[13] L. Hella, L. Libkin, and J. Nurmonen. Notions of locality and their logical characterizations over finite models. *Journal of Symbolic Logic*, 64(4):1751–1773, 1999.

[14] L. Hella, L. Libkin, J. Nurmonen, and L. Wong. Logics with aggregate operators. *J. ACM*, 48(4):880–907, 2001.

[15] N. Immerman. *Descriptive Complexity*. Springer, 1998.

[16] L. Libkin. Logics with counting and local properties. *ACM Trans. Comput. Log.*, 1(1):33–59, 2000.

[17] L. Libkin. Logics capturing local properties. *ACM Trans. Comput. Log.*, 2(1):135–153, 2001.

[18] L. Libkin. *Elements of Finite Model Theory*. Springer, 2004.

[19] H. J. Ryser. Combinatorial properties of matrices of zeros and ones. *Canad. J. Math.*, 9:371–377, 1957.

[20] D. Seese. Linear time computable problems and first-order descriptions. *Math. Structures in Computer Science*, 5:505–526, 1996.

[21] A. Tripathi, S. Venugopalan, and D. B. West. A short proof of the Erdos-Gallai characterization of graphic lists. *Discrete Math.*, 310: 843–844, 2010.

[22] D. L. Wang and D. J. Kleitman. On the existence of n-connected graphs with prescribed degrees ($n{>}2$). *Networks*, 3:225–239, 1973.

[23] B. Weisfeiler, editor. *On construction and identification of graphs*, volume 558 of *Lecture Notes in Mathematics*. Springer, 1976.

[24] B. Weisfeiler and A. A. Lehman. A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Technicheskaya Informatsia, Seriya 2*, 9:12–16, 1968. (In Russian).

APPENDIX

A. An Example

In this appendix, we pick up the paper's running example of considering the graph C_{2n} and $D_{n,n}$, where C_{2n} is a directed cycle on $2n$ nodes and $D_{n,n}$ is the disjoint union of two directed cycles on n nodes (cf., Example 3.2 and Example 3.4).

For all $\ell \geqslant 1$ let P_ℓ be the graph consisting of a single directed path on ℓ nodes (i.e., the path has length $\ell{-}1$). It is not difficult to see that P_ℓ has radius $\text{rad}(P_\ell) = \lfloor \ell/2 \rfloor$.

For all $k \geqslant 1$ let $S_{k,\ell}$ be the graph consisting of a disjoint union of k copies of P_ℓ. More precisely, $S_{k,\ell}$ is the graph with vertex set $\{1, \ldots, k\ell\}$, and for all $i \in \{1, \ldots, k\}$, the edges of the i-th path of $S_{k,\ell}$ go from vertex $(i{-}1)\ell + j$ to vertex $(i{-}1)\ell + j{+}1$, for all $j \in \{1, \ldots, \ell{-}1\}$. In particular, $S_{k,\ell}$ has size $k\ell$ and generalised radius

$$\text{genrad}(S_{k,\ell}) \;=\; k \cdot \lfloor \ell/2 \rfloor \;+\; \lfloor k/2 \rfloor\,. \tag{16}$$

The next two lemmas give an answer to the following question: How often can the graph $S_{k,\ell}$ be embedded in C_{2n} and $D_{n,n}$, respectively?

Lemma A.1 *For all $k, \ell, n \geqslant 1$, the following is true.*

$$\#_{S_{k,\ell}}(C_n) \;=\; \begin{cases} 0 & \text{if } n \;<\; k \cdot (\ell{+}1) \\[2ex] n \cdot (k{-}1)! \cdot \binom{n-k\ell-1}{k-1} & \text{otherwise.} \end{cases}$$

In particular, if $n \geqslant k \cdot (\ell{+}1)$, then $\#_{S_{k,\ell}}(C_n) = n \cdot \frac{(n-k\ell-1)!}{(n-k\ell-k)!}$.

Proof. It is straightforward to see that $S_{k,\ell}$ can be embedded in C_n if, and only if, $n \geqslant k \cdot (\ell+1)$. Thus, for the remainder of the proof it suffices to consider the case where $n \geqslant k \cdot (\ell+1)$.

Now let $\bar{x} = (x_1, \ldots, x_{k\ell})$ be an arbitrary embedding of $S_{k,\ell}$ in C_n. Then, for each $i \in [k]$, the starting point of the embedding of the i-th path of $S_{k,\ell}$ is the vertex $y_i := x_{(i-1)\ell+1}$.

We define the tuple $d(\bar{x}) = \big(v, \pi, (g_1, \ldots, g_k)\big)$, which will serve as a unique description of $\bar{x}$, as follows: $v := x_1 = y_1$ is the starting point of the embedding of the first path of $S_{k,\ell}$, and π is the permutation of $\{2, \ldots, k\}$ and $(g_1, \ldots, g_k)$ is the tuple of natural numbers, such that the following is true (where addition is modulo n) for $v_1 := v$ and $v_{i+1} := v_i + \ell + g_1$:

- $n = \sum_{i=1}^{k} (\ell + g_i)$.
- For all $i \in \{2, \ldots, k\}$, vertex v_i is the starting point of the embedding of the $\pi(i)$-th path of $S_{k,\ell}$, i.e., $v_i = y_{\pi(i)}$.
- $v_1 = v_k + \ell + g_k$.

Since $\bar{x}$ is an embedding of $S_{k,\ell}$ in C_n, we know that $g_i \geqslant 1$ for all $i \in \{1, \ldots, k\}$. Furthermore, since $n = \sum_{i=1}^{k} (\ell + g_i)$, we have $\sum_{i=1}^{k} g_i = n - k\ell$. Therefore, for every $\bar{x} \in \mathit{emb}_{S_{k,\ell}}(C_n)$, the description $d(\bar{x})$ belongs to the set D defined as follows:

> D consists of all tuples $\big(v, \pi, (g_1, \ldots, g_k)\big)$ such that $v \in V(C_n)$, π is a permutation of $\{2, \ldots, k\}$, and $(g_1, \ldots, g_k) \in \mathbb{N}_{\geqslant 1}^k$ with $\sum_{i=1}^{k} g_i = n - k\ell$.

It is straightforward to see that $d(\bar{x}) \neq d(\bar{y})$ whenever $\bar{x}, \bar{y} \in \mathit{emb}_{S_{k,\ell}}(C_n)$ with $\bar{x} \neq \bar{y}$. Furthermore, for every tuple $t \in D$, there is an embedding $\bar{x} \in \mathit{emb}_{S_{k,\ell}}(C_n)$ such that $d(\bar{x}) = t$. Thus, $\#_{S_{k,\ell}}(C_n) = |D|$. Clearly, $|D| = n \cdot (k-1)! \cdot \gamma(k, n-k\ell)$ where, for all $k, n \geqslant 1$, $\gamma(k,n) := \big|\{(g_1, \ldots, g_k) \in \mathbb{N}_{\geqslant 1}^k : \sum_{i=1}^{k} g_i = n\}\big|$ is the number of *compositions* of n into an ordered sum of k positive integers. It is well-known (and easy to see) that $\gamma(k,n) = \binom{n-1}{k-1}$. In summary, we obtain that $\#_{S_{k,\ell}}(C_n) = n \cdot (k-1)! \cdot \binom{n-k\ell-1}{k-1}$. □

Lemma A.2 *For all $k, \ell, n \geqslant 1$, the following is true:*

$$\#_{S_{k,\ell}}(D_{n,n}) \;=\; \sum_{j=0}^{k} \binom{k}{j} \cdot \#_{S_{j,\ell}}(C_n) \cdot \#_{S_{k-j,\ell}}(C_n),$$

where we fix $\#_{S_{0,\ell}}(G) := 1$ for all graphs G.

Proof. For each $I \subseteq \{1, \ldots, k\}$ let X_I be the set of all embeddings of $S_{k,\ell}$ in $D_{n,n}$ where, for each $i \in I$, the i-th path of $S_{k,\ell}$ is embedded in the first cycle of $D_{n,n}$, and for each $i' \in \{1, \ldots, k\} \setminus I$, the i'-th path of $S_{k,\ell}$ is embedded in the second cycle of $D_{n,n}$. Clearly, the set $\mathit{emb}_{S_{k,\ell}}(D_{n,n})$ is the disjoint union of the sets X_I for all $I \subseteq \{1, \ldots, k\}$. I.e., $\#_{S_{k,\ell}}(D_{n,n}) = \sum_{I \subseteq \{1,\ldots,k\}} |X_I|$. It is straightforward to see that for every $j \in \{1, \ldots, k-1\}$ and for each $I \subseteq \{1, \ldots, k\}$ of size j, we have

$$|X_I| \;=\; \#_{S_{j,\ell}}(C_n) \cdot \#_{S_{k-j,\ell}}(C_n).$$

Furthermore, $|X_\emptyset| = |X_{\{1,\ldots,k\}}| = \#_{S_{k,\ell}}(C_n) = \#_{S_{0,\ell}}(C_n) \cdot \#_{S_{k,\ell}}(C_n)$ for $\#_{S_{0,\ell}}(C_n) := 1$. In summary, we obtain that

$$\begin{aligned}\#_{S_{k,\ell}}(D_{n,n}) \;&=\; \sum_{j=0}^{k} \Big(\sum_{\substack{I \subseteq \{1,\ldots,k\} \\ \text{with } |I| = j}} \#_{S_{j,\ell}}(C_n) \cdot \#_{S_{k-j,\ell}}(C_n) \Big) \\ &=\; \sum_{j=0}^{k} \binom{k}{j} \cdot \#_{S_{j,\ell}}(C_n) \cdot \#_{S_{k-j,\ell}}(C_n).\end{aligned}$$

This completes the proof of Lemma A.2. □

k	$o(U_k)$	$g(U_k)$	$\#(C_{12})$	$\#(D_{12,12})$	$\#(C_{24})$
1	3	1	12	24	24
2	6	3	60	408	408
3	9	4	24	$4,368$	$4,368$
4	12	6	0	$23,904$	$23,760$
5	15	7	0	$28,800$	$40,320$
6	18	9	0	$11,520$	$2,880$
7	21	10	0	0	0

Table 1. Results concerning Example A.3: Let $U_k := S_{k,3}$ for all $k \geqslant 1$; "o", "g", and "$\#$" are abbreviations for order, genrad, and $\#_{U_k}$. Using equation (16), Lemma A.1, and Lemma A.2, one obtains the values indicated in the table.

k	$o(T_k)$	$g(T_k)$	$\#(C_6)$	$\#(D_{6,6})$	$\#(C_{12})$
1	2	1	6	12	12
2	4	3	6	84	84
3	6	4	0	216	240
4	8	6	0	216	72
5	10	7	0	0	0

Table 2. Results concerning Example A.4: Let $T_k := S_{k,2}$ for all $k \geqslant 1$; "o", "g", and "$\#$" are abbreviations for order, genrad, and $\#_{T_k}$. Using equation (16), Lemma A.1, and Lemma A.2, one obtains the values indicated in the table.

Example A.3 *Consider the graphs C_{2n} and $D_{n,n}$ for $n = 12$, i.e., the graphs C_{24} and $D_{12,12}$. By equation (1) we know that*

$$C_{24} \approx_5^{\mathit{Hanf}} D_{12,12} \qquad \text{and} \qquad C_{24} \not\approx_6^{\mathit{Hanf}} D_{12,12}.$$

Theorem 4.4 thus tells us that

(i) $\#_S(C_{24}) = \#_S(D_{12,12})$
for all graphs S with $\operatorname{genrad}(S) \leqslant 5$, *and*

(ii) $\#_{S'}(C_{24}) \neq \#_{S'}(D_{12,12})$
for some graph S' with $\operatorname{genrad}(S') = 6$.

In fact, Table 1 shows that S' can be chosen to be the graph $U_4 := S_{4,3}$.

Letting $U_k := S_{k,3}$ (for all $k \geqslant 1$), Table 1 shows that $\#_{U_k}(C_{24}) = \#_{U_k}(D_{12,12})$ for $k \in \{1, 2, 3\}$ and for $k \geqslant 7$, whereas $\#_{U_k}(C_{24}) \neq \#_{U_k}(D_{12,12})$ for all $k \in \{4, 5, 6\}$.

Example A.4 *Consider the graphs C_{2n} and $D_{n,n}$ for $n = 6$, i.e., the graphs C_{12} and $D_{6,6}$. By equation (1) we know that*

$$C_{12} \approx_2^{\mathit{Hanf}} D_{6,6} \qquad \text{and} \qquad C_{12} \not\approx_3^{\mathit{Hanf}} D_{6,6}. \tag{17}$$

Theorem 4.4. thus tells us that

(i) $\#_S(C_{12}) = \#_S(D_{6,6})$
for all graphs S with $\operatorname{genrad}(S) \leqslant 2$, *and*

(ii) $\#_{S'}(C_{12}) \neq \#_{S'}(D_{6,6})$
for some graph S' with $\operatorname{genrad}(S') = 3$.

It is straightforward to see that S' can be chosen to be the graph P_6, i.e., the directed path on 6 nodes, for which $\#_{P_6}(C_{12}) = 12$ and $\#_{P_6}(D_{6,6}) = 0$.

On the other hand, Table 2 shows that $T_2 := S_{k,2}$ is a graph of generalised radius 3 where still $\#_{T_2}(C_{12}) = \#_{T_2}(D_{6,6})$.

Letting $T_k := S_{k,2}$ (for all $k \geqslant 1$), Table 2 shows that $\#_{T_k}(C_{12}) = \#_{T_k}(D_{6,6})$ for k=1, k=2, and $k \geqslant 5$, whereas $\#_{T_k}(C_{12}) \neq \#_{T_k}(D_{6,6})$ for k=3 and k=4.