# Hybrid graphs as a framework for the small–world effect

Katharina A. Lehmann*, Hendrik D. Post*, and Michael Kaufmann†

*Wilhelm–Schickard–Institut für Informatik, Universität Tübingen*

(Dated: April 24, 2006)

## Abstract

In this article we formalize the small–world effect which describes the surprising fact that a hybrid graph composed of a local graph component and a very sparse random graph has a diameter of $O(\log n)$ whereby the diameter of both components alone is much higher. We show that a large family of these hybrid graphs shows this effect and that this generalized family also includes classic small–world models proposed by various authors although not all of them are captured by the small–world definition given by Watts and Strogatz. Furthermore, we give a detailed upper bound of the hybrid's graph diameter for different choices of the expected number of random edges by applying a new kind of proof pattern that is applicable to a large number of hybrid graphs. The focus in this paper is on presenting a flexible family of hybrid graphs showing the small–world effect that can be tuned closely to real–world systems.

PACS numbers: 89.75.Fb

* contributed equally

†Electronic address: lehmannk/post/mk@informatik.uni-tuebingen.de

## I. INTRODUCTION

The 'small–worldness' of social networks has long been a part of folklore until Milgram took the first steps to examine it more closely in 1967 [1]. The small–world property describes the fact that people are tightly knit into small social clusters while on the other hand just a short chain of acquaintances is needed to connect almost any two humans in the world. Milgram estimated the number of persons in such a chain to be around six [1] which is why this observation is also known under the title 'six degrees of separation'. The first formal approach to explain this astonishing result was made by Watts and Strogatz in a seminal paper [2] in which they gave a first definition of small–world networks and presented a model for their generation. They defined a small–world to be every network that is on the one hand very regular, in the sense of locally densely connected, and on the other hand random enough to show a low diameter.

Following their publication, several real–world networks such as the WWW or file–sharing communities were analyzed and shown to be small–worlds (e.g. [3–6]). A second research area deals with network–based processes on small–world networks, like the behavior of neural networks on small–worlds [7] or disease spreading in small–worlds [8]. Other directions of research tried to find more rigorous analytical results on the properties of either the classic small–world model or on variants of the small–world model that were easier to analyze or captured new aspects of small–worlds [9–15]. All of these models are either build by starting from a regular network where a very small proportion of the edges is randomly rewired [9, 10], or they are composed of a regular network and a very sparse random graph [11–15]. The latter type of model we will call a hybrid graph of a non–random and a random graph component.

In this article we want to concentrate on the *small–world effect*, which we define as the substantial decrease of the diameter by building the hybrid graph out of two graph components where each of the components alone has a much higher diameter than the hybrid graph. The formal definition is given in Def. 1. The connection between the terms *small–world* and *small–world effect* is as follows: Watts and Strogatz defined a graph to be a small–world if it shows a high clustering coefficient and a low diameter of $O(\log n)$. The clustering coefficient equals the number of neighbors of a vertex that are connected in relation to the maximal number of connected neighbors. This measure is in most cases

a good indicator of localness and regularity: It is for example known that distributing $n$ vertices in a $2d$ space uniformly at random and connecting each vertex with its $k$ next neighbors will result in a clustering coefficient of around 0.58 [20]. But not every graph that is regular and in which vertices are connected that are near to each other has a high clustering coefficient, e.g., all orthogonal lattices have a clustering coefficient of 0. In the strict sense, all small–world models based on regular lattices are not included in the set of small–worlds, e.g. [11, 12]. Also some real-world networks are not classified as a small-world in that strict sense, e.g., data on heterosexual relationship networks that have by definition a clustering coefficient of 0 because they are bipartite [21]. Of course, all these models and data are nonetheless named and recognized as small–worlds and small–world models, respectively. The *small–world effect* is now defined for graph models and not so much for single graphs, and includes by definition all classic small–world models because it does not require a high clustering coefficient. We want to note that there might be graph models that would be classified as a small–world in the strict sense of Watts and Strogatz but do not show the small–world effect. As an example, let a graph be composed of a balanced binary tree and of a second graph, where any two leaves of the tree are connnected with each other if they are connected to the same vertex. Since there are $n/2$ leaves in the tree and every leaf has a clustering coefficient of 1 due to the additional connection, the whole graph has a clustering coefficient greater than 0.5 but the graph does not show the small–world effect because the balanced tree alone has a small diameter of $O(\log n)$. In our intuition, graph models with this property would not be acknowledged as small–world models and to our knowledge there is no such small–world model.

We state that for all the models that have been acknowledged as small–world models so far, the small–world effect is a generalizing concept that unifies the existing small-world models.

In the following we will restrict ourselves to those hybrid graphs where one component is a random graph $G(n, p)$, in which every edge exists with the same probability $p$. We show in this article, that the small–world effect can be found in a large family of hybrid graphs with different graph families building the non–random graph component. We introduce a new characteristic for graph families, namely the *regular decomposability* and prove that it is sufficient for any hybrid graph to show the small–world effect if the non–random graph family has this property. We also give strong upper bounds on the diameter of the hybrid

graph in dependence of $p$ and the structure of the non–random graph component.

The paper is organized as follows: In Sec. II we give some basic definitions needed in the concourse of the article. Sec. III is structured into three subsections: Subsec. III A introduces the main model, Subsec. III B gives the upper bound of the diameter for this model, and in Subsec. III C we generalize the analysis to any combination of regularly decomposable graph families with $G(n, p)$. In Sec. IV we discuss the relationship between locally clustered and regularly decomposable graphs on the example of $k$–next neighborhood graphs. Sec. V concludes with a summary and discussion of the results.

## II. DEFINITIONS

A graph family $G(n)$ denotes any set of graphs generated by the same algorithm and parameterized by the number $n$ of vertices in it. For non–random graph families and a fixed set of parameters only one specific graph is generated. For graph families generated partly by probabilistic processes, $G(n)$ is defined as the set of all possible realizations. Statements about $G(n)$ are then interpreted as statements about expected characteristics of this set. We will use the notation $G(n)$ interchangeably for the whole set or a specific realization of this set.

A regular $d$–dimensional, equilateral grid (hypercubical lattice) $G_d(n)$ is defined as a set of vertices placed on integer positions in $d$ dimensions. $a \in \mathbb{N}$ denotes the number of vertices placed in each of the $d$ dimensions. The number of vertices in this grid is then given by $n = a^d$, where every possible position — identified by a $d$–dimensional vector $(1 \leq b_1 \leq a, 1 \leq b_2 \leq a, \ldots, 1 \leq b_d \leq a)$ — is occupied with one vertex. The degree $deg(v)$ of a vertex is defined as the number of incident edges and equals the number of direct neighbors of $v$. Every vertex $v$ is connected by an edge to those vertices that differ in their position by exactly one in exactly one dimension from the position of $v$, i.e., every vertex has at most degree $2d$. For these grids, the graph theoretic distance $d(v, w)$ of any two vertices $v, w$, i.e., the minimal number of traversed edges to walk from $v$ to $w$, coincides with the Manhattan distance $d_M(v, w)$ of these vertices which is defined by:

$$d_M(v, w) = \sum_{1 \leq i \leq d} |b_i(v) - b_i(w)| \tag{1}$$

The diameter $D(G)$ of any graph $G$ is defined as the maximal distance of any two vertices

within the graph. The diameter $D(G_d(n))$ is given by the maximal Manhattan distance of any two vertices in $G_d(n)$ and can be calculated by:

$$D(G_d(n)) = \sum_{1 \leq i \leq d} a - 1 = d(a-1) \tag{2}$$

A graph is *connected* if there is a way from every vertex $v$ to any other vertex $w$.

The clustering coefficient $C(v)$ of a vertex $v$ is defined as the ratio of the number of edges $e(v)$ between direct neighbors of $v$ and the maximal possible number of edges between direct neighbors [2]:

$$C(v) = \frac{e(v)}{deg(v)(deg(v)-1)} \tag{3}$$

A $G(n,p)$ random graph is defined as an instance of all possible graphs with $n$ vertices where every of the $\binom{n}{2}$ edges exists with probability $p$ [16].

A graph $G$ or a graph family $G(n)$ is *clustered* if (for every $n$) there is a partitioning such that the fraction of realized edges within the subgraphs given by the partition is much higher than the fraction of realized edges between the subgraphs, i.e., the subgraphs are *locally dense* and *globally sparse*. This definition is quite broad because there are many different measurements that try to quantify how good such a clustering is, e.g., coverage, intra– and inter–cluster conductance (for an overview see [17]), or modularity [18], to name but a few. Here, we just want to concentrate on the *locally dense* vs. *globally sparse* part of that definition that is the intuive basis for most of these measures. A graph or a graph family $G(n)$ is *locally clustered* if (for every $n$) vertices have a position in a $d$–dimensional space and at least one partition exists that divides the space into subspaces such that the subgraphs within these subspaces are locally dense and the graph is globally sparse. If a graph is only given by its adjacency matrix and the vertices have no position in space, the graph can be said to be *potentially locally clustered* if a distribution of the vertices in a $d$–dimensional space can be found such that the graph is locally clustered. Note that all classic small-world models are based on locally clustered graph families like chordal rings in which every vertex is placed on a ring and is connected to its $k$-next neighbors on each side [2, 15] or d-dimensional lattices [11–14].

We will use the following theorem on the diameter of random graphs $G(n,p)$ [19]:

**Theorem 1** *If* $pn/\log n \to \infty$ *and* $\log n/\log(np) \to \infty$ *then* $D(G(n,p))$ *is asymptotically equal to* $\log n/\log(np)$ *with high probability.*

Note that this theorem implicitly includes that the random graph is connected with high probability. To simplify the following proofs we will use a stricter version of the theorem and require additionally that $p \geq (\log n)^{1+\epsilon}/n$, where $\epsilon > 0, \in \mathbb{R}$.

## III. A FRAMEWORK FOR HYBRID GRAPHS WITH A SMALL–WORLD EFFECT

The following definition of the small–world effect describes the decrease of the diameter by combining two graph components that have a much higher diameter as single components:

**Definition 1** *A hybrid graph family is defined as any combination $G_{LR}(n)$ of a clustered graph family $G_L(n)$ and a random graph family $G_R(n)$. $G_{LR}(n)$ shows the small–world effect if the diameter $D(G_{LR}(n))$ is at most scaling polylogarithmically and if the following relations hold for $n \to \infty$:*

$$\frac{D(G_L(n))}{D(G_{LR}(n))} \to \infty \quad \textit{and} \quad \frac{D(G_R(n))}{D(G_{LR}(n))} \to \infty \tag{4}$$

We will now present a rather general proof pattern with which the upper bound on the diameter of certain hybrid graphs can be given. The number of required properties of these hybrid graphs is very small and thus a large family of hybrid graphs falls into this category of graph families with a small–world effect. For didactic purposes we start with a very simple model that will be generalized later to a large family hybrid graphs that fall into the above given definition.

### A. A first Starting Point

We start with a simple hybrid graph composed of a random graph $G(n, p)$ and a regular $d$–dimensional grid in the following way: The basic regular graph is the $d$–dimensional grid of $n$ vertices, where each vertex is connected to its $2d$ next neighbors, combined with a $G(n, p)$ random graph on the same $n$ vertices. We will denote by $G_d(n, p)$ a graph from our model, which is given by the combination of a $G_d(n)$ regular grid and a random graph $G(n, p)$.

The remaining part of this section gives an answer to the following question: How does the diameter of regular networks combined with a small set of random edges scale?

Since the basic network is a $d$–dimensional grid, the diameter of this component without any added random edges will scale with $a-1$ for a fixed dimension $d$: $D(G_d(n)) = d \cdot (a-1)$. If the added random graph has a probability greater than or equal to $(\log p)^{1+\epsilon}/n$ then the combined graph will have a diameter that is dominated by the diameter of the random graph and thus is asymptotical to at most $\log n / \log(np)$ (Theorem 1). Thus, a hybrid graph with a dense random graph component does not show the small–world effect.

In the following we will show what happens in the regime where $p$ lies below $(\log p)^{1+\epsilon}/n$ and describe the regime in which the diameter of the hybrid graph will scale at most (poly–) logarithmically.

### B. The Diameter of $G_d(n,p)$–Graphs

For the above given model of a graph $G_d(n,p)$ the following lemma holds:

**Lemma 1** *For $p = \frac{1}{cn}$, $c \in \mathbb{R}^+$ the diameter of $G_d(n,p)$ is asymptotically bounded by at most*

$$d \cdot \left( \left\lceil \sqrt[d]{c \cdot (\log n)^{1+\epsilon}} \right\rceil - 1 \right) \cdot \left( \frac{\log n}{(1+\epsilon)\log\log n - \log 2} + 1 \right) \tag{5}$$

The proof proceeds in four steps:

1. To prove the lemma we partition $G_d(n,p)$ into $n_S$ connected $d$–dimensional equilateral subgraphs $S_i, 1 \leq i \leq n_S$ with a side length $l$ such that each subgraph contains at least $s = l^d \geq c \cdot (\log n)^{1+\epsilon}$ vertices (Figure 1).

2. For any $a$, we can only build $\lfloor a/l \rfloor$ full subgraphs per dimension. $n^*$ denotes the number of all vertices contained in a full subgraph. We will show that the $n - n^*$vertices that are not contained in any full subgraph build a vanishing fraction of all vertices for $n \to \infty$. We will thus base our proof on a reduced regular $d$–dimensional grid of size $n^*$ that contains only the full subgraphs.

3. We construct a supergraph $G_S(n_S) = (S, E')$ where each vertex $v_i \in S$ uniquely represents the subgraph $S_i$ for $1 \leq i \leq n_S$. Edge $e = (v_i, v_j)$ is member of $E'$ iff there is at least one random edge from any vertex in $S_i$ to any vertex in $S_j$. We will prove that Theorem 1 can be applied to $G_S(n_S)$.

a) b)

FIG. 1: Valid partitions for a 2–dimensional grid with side length $a$. Full equilateral subgraphs with side length $l$ may be placed arbitrarily as long as their number is maximal. Therefore numerous partitions exist and for each pair of vertices numerous partitions can be found where both are contained in full subgraphs.

4. Then we will expand $G_S(n_S)$ to gain a bound on the diameter of the original but reduced graph $G_d(n^*, p)$. The diameter of $G_d(n^*, p)$ is bounded by the product of the diameter of the subgraphs $D(S_i)$ and the diameter $D(G_S)$. We will show that there are numerous partitions of $G_d(n, p)$ into $n_S$ subgraphs. Especially, for any pair of vertices $v, w$ there is at least one partition of $G_d(n)$ such that both, $v$ and $w$, are contained in full subgraphs. Since every supergraph based on a possible partition obeys Theorem 1, we will therefore have shown that the whole graph $G_d(n, p)$ obeys Lemma 1 and the case is proven.

We will start by partitioning a $G_d(n, p)$ graph. Let $S_i, 1 \leq i \leq n_S$ denote an equilateral subgraph that has a side length of $l = \left\lceil \sqrt[d]{c \cdot (\log n)^{1+\epsilon}} \right\rceil$ in each dimension. The number $s$ of vertices contained in one (full) subgraph is bounded by:

$$c \cdot (\log n)^{1+\epsilon} \leq s = \left\lceil \sqrt[d]{c \cdot (\log n)^{1+\epsilon}} \right\rceil^d < 2^d \cdot c \cdot (\log n)^{1+\epsilon} \tag{6}$$

We will now partition $G_d(n, p)$ into the subgraphs as shown in Figure 1. Obviously, incomplete subgraphs exist if $a/l$ is not integer. The leftover vertices can be placed arbitrarily between full subgraphs as indicated in Figure 1 b). For simplicity we will consider instead of $G_d(n)$ a smaller hypercube $G_d(n^*)$ containing all full subgraphs. Note that now $a^*$ with $\sqrt[d]{n}^* = a^* \leq a$ is the maximal integer smaller than $a$ that is a multiple of $l$. Let $q = a^*/l$ denote the number of subgraphs in one dimension.

The relative fraction of vertices not contained in full subgraphs is approaching 0 for $n \to \infty$:

$$\frac{n - n^*}{n} \leq \frac{(l \cdot (q+1))^d - (l \cdot q)^d}{(l \cdot q)^d} \tag{7}$$

$$= \left(\frac{q+1}{q}\right)^d - 1 \tag{8}$$

Since $q \to \infty$ for $n \to \infty$, the relative fraction of ignored vertices is asymptotically 0. Note that $n_S = \frac{n^*}{s} \geq \frac{n}{2^d \cdot c \cdot (\log n)^{1+\epsilon}} \to \infty$. Thus for $n \to \infty$ we may safely use

$$n \geq n^* > n/2 \tag{9}$$

In $G_d(n^*, p)$ there are $s^2$ possible random edges between any vertex from subgraph $S_i$ and any vertex from subgraph $S_j$. Each of these edges exists independently with probability $p$. It follows that for $G_S$ the probability $p_S$ is exactly $\frac{s^2}{cn}$.

We will now prove that Theorem 1 can be applied to $G_S(n)$. A basic observation is that for $n \to \infty$, also $n_S \to \infty$. Additionally, we must show that $\frac{p_S n_S}{\log n_S} \to \infty$ and $\frac{\log n_S}{\log(n_S p_S)} \to \infty$ for $n_S \to \infty$.

Regarding, that for all $n_S > 1$, $n^* > n/2$ (eq. 9) the following two equations hold:

$$\begin{aligned}
\frac{p_S \cdot n_S}{\log n_S} &= \frac{s^2}{cn} \cdot \frac{n^*}{s} \cdot \frac{1}{\log \frac{n^*}{s}} && (10)\\
&\geq \frac{s}{2c(\log n^* - \log s)} && (11)\\
&\geq \frac{(\log n)^{1+\epsilon}}{2\log n - 2\log s} && (12)
\end{aligned}$$

such that $\frac{p_S \cdot n_S}{\log n_S} \to \infty$ for $n \to \infty$ and

$$\begin{aligned}
\frac{\log n_S}{\log(p_S \cdot n_S)} &= \frac{\log \frac{n^*}{s}}{\log \frac{s \cdot n^*}{cn}} && (13)\\
&\geq \frac{\log n/2 - \log(2^d \cdot c(\log n)^{1+\epsilon})}{\log(2^d (\log n)^{1+\epsilon})} && (14)
\end{aligned}$$

such that also $\frac{\log n_S}{\log(p_S \cdot n_S)} \to \infty$. By Theorem 1 we know that thus $G_S$ has a diameter asymptotical to $\frac{\log n_S}{\log(p_S \cdot n_S)}$. Regarding that $n^*/n > 1/2$ this is bounded by

$$\begin{aligned}
D(G_S) &= \frac{\log n_S}{\log(p_S \cdot n_S)} && (15)\\
&\leq \frac{\log n}{\log \frac{s}{2c}} && (16)\\
&\leq \frac{\log n}{(1+\epsilon)\log\log n - \log 2} && (17)\\
&\leq \frac{\log n}{\log\log n} && (18)
\end{aligned}$$

Where the last inequality is valid for all $n$ with $\epsilon \log\log n > \log 2$.

We will now expand $G_S(n)$ in order to get an upper bound for the diameter of $G_d(n,p)$.

Let $v$ and $w$ be two vertices in the original graph $G_d(n,p)$. First, we will reduce $G_d(n,p)$ to $G_d(n^*,p)$ in such a way that $v$ and $w$ are contained in $G_d(n^*,p)$. Then we know that there is a path from subgraph $S_i$ containing $v$ to subgraph $S_j$ containing $w$ with a length of no more than $D(G_S)$. This path is denoted by $(e_1, e_2, ..., e_k)$, the sequence of edges to traverse to walk from $S_i$ to $S_j$.

To use this path in the original graph $G_d(n,p)$, we will first have to walk from vertex $v$ to that vertex $v'$ from $S_i$ that is attached to $e_1$. This will at most take $D(S_i) = d \cdot (\left\lceil \sqrt[d]{c \cdot (\log n)^{1+\epsilon}} \right\rceil - 1)$ steps. For every entered subgraph $S_x$ on the way to subgraph $S_j$, an additional distance of $D(S_x)$ has at most to be added to get from the random edge entering the subgraph to the edge leaving this subgraph. Thus, the distance of $v, w$ in the original graph $G_d(n,p)$ is asympotically given by at most

$$D(S_i) \cdot (D(G_S) + 1) \leq$$
$$d \cdot \left( \left\lceil \sqrt[d]{c \cdot (\log n)^{1+\epsilon}} \right\rceil - 1 \right) \cdot \left( \frac{\log n}{(1+\epsilon) \log \log n - \log 2} + 1 \right)$$

With this, Lemma 1 is proven.

In the following we want to discuss what happens if the degree of the underlying grid graph is enlarged.

As stated in Lemma 1, the diameter of a $G_d(n,p)$ graph is asymptotically at most $D(S_i) \cdot (D(G_S) + 1)$. Let $G_d(n,k,p)$ denote an extended regular grid, in which every vertex is connected to its $k$ next neighbors, combined with an additional $G(n,p)$ graph. The diameter $D(S_i)$ depends on the degree of the vertices in the underlying grid graph. Thus, if we want to reduce the diameter of the $G_d(n,k,p)$ graph we just have to add some more edges to the grid. For example, $D(S_i)$ is reduced to 1 if for $p = \frac{1}{c \cdot n}$ we add edges from every vertes to its $c \cdot (\log n)^{1+\epsilon}$ next neighbors. The combined graph $G_d(n, (\log n)^{1+\epsilon}, p)$ has now a diameter of at most $D(G_S)$.

### C. Generalizing the Family of Hybrid Graphs with Small–World Effect

In this section we will generalize Lemma 1 in two ways:

1. The probability $p$ of the added random graph $G(n,p)$ can be as small as $\frac{1}{f(n) \cdot n}$ as long as $(\log n)^{1+(\epsilon/2)} \leq f(n) \leq n^{1-\delta}$ for some constants $\delta, \epsilon > 0$ and $n \to \infty$.

2. The basic regular $d$–dimensional grid can be replaced by certain graph families. This was already indicated at the end of section III B.

These two extensions lead finally to our generalized theorem on the diameter of hybrid graphs with a small–world effect.

*1. Generalizing the random graph component*

At first, we explain in which range $p$ can be chosen, such that the proof–technique can still be applied. In section III B, we kept $p = 1/cn$. For smaller $p = \frac{1}{f(n)\cdot n}$, the size $s$ of the subgraphs has to be chosen larger such that Theorem 1 can be applied. Let again $n_S$ denote the number of vertices and $p_s$ denote the probability of an edge in $G_S$.

For simplicity we assume that $n_S = n/s \in \mathbb{N}$ and $p \cdot s \cdot n = (\log n)^{1+\epsilon} \in \mathbb{N}$. The general case follows the argumentation above.

The number of nodes in each subgraph will be chosen such that
$s = \frac{(\log n)^{1+\epsilon}}{p\cdot n} = f(n) \cdot (\log n)^{1+\epsilon}$. Again, Lemma 1 requires the validity of

$$\frac{p_S n_S}{\log n_S} \to \infty \tag{19}$$

and

$$\frac{\log n_S}{\log(n_S p_S)} \to \infty \tag{20}$$

As before $p_S = s^2 \cdot p$. We first analyze the condition given in equation (19):

$$\begin{aligned} \frac{p_S n_S}{\log n_S} &= s^2 \cdot p \cdot \frac{n}{s} \cdot \frac{1}{\log n_S} && (21)\\ &> s \cdot p \cdot \frac{n}{\log n} && (22)\\ &= (\log n)^{\epsilon} && (23) \end{aligned}$$

which approaches infinity for increasing $n$. The second condition (20) simplifies to

$$\begin{aligned} \frac{\log n_S}{\log(n_S p_S)} &= \frac{\log\left(\frac{n}{s}\right)}{\log\left(\frac{n}{s} \cdot s^2 \cdot p\right)} && (24)\\ &= \frac{\log\left(\frac{\frac{n}{f(n)}}{(\log n)^{1+\epsilon}}\right)}{\log(\log n)^{1+\epsilon}} && (25)\\ &= \frac{\log\left(\frac{n}{f(n)}\right)}{\log(\log n)^{1+\epsilon}} - 1 && (26) \end{aligned}$$

which tends to infinity for $f(n) \leq n^{1-\delta}, \delta > 0$. Therefore both conditions are met and Theorem 1 can be applied to $G_S$. If $f(n)$ is chosen to be lower than $(\log n)^{1+(\epsilon)}$ then the random graph component will be connected and Theorem 1 can be directly applied. Such, $f(n)$ is also restricted from below to be larger than $(\log n)^{1+(\epsilon)}$.

We summarize this result in

**Lemma 2** *For any function $(\log n)^{1+(\epsilon/2)} \leq f(n) \leq n^{1-\delta}, \epsilon, \delta > 0$ and $p = \frac{1}{f(n)\cdot n}$, we can partition the grid graph within a $G_d(n,p)$ graph into $n_S = \frac{n}{s}$ subgraphs $S_i$ of size $s = f(n) \cdot (\log n)^{1+\epsilon}$ such that $G_d(n,p)$ shows a diameter of asymptotically at most (Eq. 26)*

$$\underbrace{d \cdot \left( \left\lceil \sqrt[d]{c \cdot (\log n)^{1+\epsilon}} \right\rceil - 1 \right)}_{D(S_i)} \cdot \underbrace{\left( \frac{\log(n/f(n))}{\log(\log n)^{1+\epsilon}} \right)}_{D(G_S)} \tag{27}$$

Now we will discuss the second possible generalization.

*2. Possible replacements of the regular grid graph*

In the general proof pattern shown above, the following two properties of regular grid graphs are needed: First, regular grid graphs are partitionable for every $n$ into $\Theta(n/s(n))$ subgraphs of size $s(n)$ for any function $s(n) \leq n$ such that each of these subgraphs is a connected graph. The second property used is that for any pair of vertices $v, w$ there must be at least one partition such that $v$ and $w$ are contained in any of the subgraphs.

To abstract from this special graph family to all graph families with these two properties we introduce the following definition:

**Definition 2** *Let $G_L(n)$ be a graph family with the following two properties:*

1. *$G_L(n)$ is partitionable for every $n$ into $\Theta(n/s(n))$ subgraphs of size $s(n)$ for any function $s(n) \leq n$ such that each of these subgraphs is a connected graph.*

2. *For any pair of vertices $v, w$ and every $n$ there must be at least one partition as described such that $v$ and $w$ are contained in proper subgraphs.*

*$G_L(n)$ is called a regularly partitionable graph family. For graph families with a stochastic generating process, it is enough to show that such a partition is existing whp.*

Note that every graph family $G_L(n)$ is *restricted regularly decomposable* for at least $s(n) = 1$. Let $s_{max}(n)$ be that function $s'(n)$ that has the fastest growth of all functions $s(n)$ for which $G_L(n)$ is restricted regularly decomposable. If now $s_{max}(n) = k,\ k \in \mathbb{N}$ for $G_L(n)$ and $G_L(n)$ replaces the regular grid then it is clear that the size of the subgraphs is also at most $k$ to obey $\Theta(n/s)$. This implies that $p$ of the added random graph must be at least $O\left(\frac{(\log n)^{1+\epsilon}}{n}\right)$ in order to achieve a supergraph that obeys Theorem 1. It follows that the diameter is reduced to the diameter of a random graph because we add a random graph with the wanted diameter. In this case the definition of the small–world effect would exclude this kind of combination.

We conclude this section with a theorem on the diameter of generalized small–world models combining a locally clustered graph family with a thin random graph:

**Theorem 2** *Let $G_L(n,p)$ denote the combination of instances of a regularly decomposable graph family $G_L(n)$ and a $G(n,p)$ graph where $p = \frac{1}{f(n)\cdot n}$, $\frac{1}{(\log n)^{1+(\epsilon/2)}} \leq f(n) \leq \frac{1}{n^{1-\delta}}, \epsilon > 0, \delta > 0$. $D(s(n,p))$ denotes the maximal diameter of any subgraph of $G_L(n,p)$ with size $s(n,p) = \frac{(\log n)^{1+\epsilon}}{p\cdot n},\ \epsilon > 0$, the diameter of $G_L(n,p)$ is asymptotically at most:*

$$D\left(s\left(n, \frac{1}{f(n)\cdot n}\right)\right) \cdot \underbrace{\left(\frac{\log(n/f(n)}{\log(\log n)^{1+\epsilon}}\right)}_{D(G_S)} \tag{28}$$

## IV. PROVING REGULAR DECOMPOSABILITY

We now want to discuss the connection between regularly decomposable graph families and locally clustered graph families. As indicated before, the classic small–world models are based on locally clustered graph families but our proof pattern is based on the notion of regularly decomposable graph families. It is easy to see in the special cases of chordal rings and d-dimensional lattices that these are at the same time locally clustered and regularly decomposable.

We will now show under which conditions general locally clustered graph families are also regularly decomposable. The main idea of the proof is to use the embedding of the vertices in a space and to divide this space into areas that contain connected subgraphs with at least a given number of vertices. Thus, the partition of space induces a partition of the graph. The

task that has to be tailored for every specific graph family, is showing that there is always a partition of space that yields $\Theta(n/s(n)))$ connected subgraphs with at least the wanted number $s(n)$ of vertices. Graph families that are locally clustered and show this property are also regularly decomposable. Note that this is not a necessary property of locally clustered graph families. It is possible to construct artificial locally clustered graph families that lack this property, e.g., a graph family $G_5(n)$ which consists of unconnected 5–cliques $C_1, C_2, ..., C_{\lfloor n/5 \rfloor}$ where all members of clique $C_i$ are located at position $i$ in a 1–dimensional space. This family is clearly locally clustered for the partition $\{C_1, C_2, \ldots, C_{\lfloor n/5 \rfloor}\}$ but it is not regularly decomposable because it is not possible to partition the graph into subgraphs with a size $s(n) > 5$.

In the following we will analyze the regular decomposability of a more sophisticated locally clustered graph family, namely that of $k$–next neighborhood graphs (knn–graphs): Given a set of vertices distributed in any $d$–dimensional space, let $E$ be the set of directed edges such that every vertex $v$ is connected to its $k$ next neighbors. If the set of $k$ next neighbors is ambiguous for any vertex, then choose any of the possible sets uniformly at random. Note that the edge relation is not symmetric and therefore the *knn*–graph is a directed graph and thus we differentiate between the outgoing and ingoing edges of a vertex. This kind of graph family certainly matches the common intuition about a *local* graph family and It is also locally clustered for any non-trivial partitioning of the space. We will now prove the following theorem:

**Theorem 3** *The $k$–next neighborhood graph family is regularly decomposable if the vertices of a given instance are distributed uniformly at random in a 2–dimensional unit–square.*

In order to prove the given property we will proceed in the following steps:

1. Bound the expected distance to next neighbors from above.
2. Prove that a knn–graph is connected with high probability for $k > \log n$.
3. Show that a generic partition procedure yields the required $\Theta(n/s(n))$ subgraphs for a given size function $s(n) < n$.

### A. A bound for the maximum distance of nearest neighbors

Let the *knn*–disk of any vertex $v$ be defined as the minimal disk which contains all of its $k$ next neighbors. Note that the disc's radius is equal to the maximum distance of any connected nearest neighbor to $v$.

The probability for any vertex $v$ to be placed in some area of size $A \leq 1$ within the unit square is exactly $A$. Thus, the placement of vertices into a given area is a Bernoulli trial with $p = A$ and $q = 1 - A$. The radius of a disk with expectedly $k = n \cdot A$ vertices is thus clearly given by $\bar{r} = \sqrt{\frac{k}{\pi \cdot n}}$. Now, the Chernoff bound may be applied to yield an upper bound for the diameter of any disk within the unit square that contains at least $k$ vertices:

**Lemma 3** *Let $\hat{r} = \sqrt{\hat{c}} \cdot \bar{r}$ denote a knn–disk radius with $\hat{c} = 3 + \sqrt{8}$. Further let $k \geq \log n$. With high probability, no disk with radius $\hat{r}$ around any vertex $v$ exists that does not contain at least $k$ vertices.*

**Proof 1** *Let $D_v$ denote a knn–disk around $v$ with an expected number of vertices lying in that disk equal to $\bar{k} = c \cdot k$. $X_k$ denotes the number of vertices lying inside of $D_v$. Now we apply a relaxed version of the Chernoff inequality for independent Bernoulli trials. With $\mu = c \cdot k$ and $\delta = 1 - \frac{1}{c}$*

$$Pr[X_k < (1-\delta)\mu] < e^{-\frac{1}{2}\mu\delta^2} = e^{-\frac{ck}{2}(1-\frac{1}{c})^2} < n^{-\frac{c}{2}(1-\frac{1}{c})^2} \tag{29}$$

*The latter inequation is only valid for $k > \log n$. For $c = \hat{c} = 3 + \sqrt{8}$ we yield*

$$Pr[X_k < (1-\delta)\mu] < \frac{1}{n^2} \tag{30}$$

*Hence, the probability that there is a knn–disk with radius larger than $\hat{r} = \sqrt{3 + \sqrt{8}} \cdot \bar{r}$ in a knn–graph with $k > \log n$ is $< 1/n$.*

The interpretation of this result is that it is almost impossible for $n \to \infty$ that any knn–disk exists with a radius larger than $\hat{r}$. Therefore in our following theorems, we consider the radius of knn–discs to be bound by $\hat{r} = \sqrt{\hat{c}} \cdot \bar{r}$.

Note that these equations are only valid for disks that do not intersect with the unit squares border. If a vertex $v_c$ is positioned in a corner of the unit square, a factor of 2 has to be applied to $\hat{r}$ to yield the correct upper bound on the radius of its *knn*–disk.

FIG. 2: $v$ and $w$ are two vertices from different connected components of a *knn*–graph having minimal euclidian distance to each other. Two circles are drawn around $v$ and $w$, respectively, with a radius that equals the euclidian distance between $v$ and $w$. The figure shows that none of the $k$–next neighbors of neither $v$ nor $w$ can exist in the intersection of these circles without contradicting the condition that $v$ and $w$ are the pair of vertices from different connected components with minimal euclidian distance.

### B. Connectedness of knn–graphs

We will now show that a knn–graph is connected w.h.p.

The proof for the following lemma will be ommited. Here we will just sketch it shortly: As can be seen in Fig. 2, in every unconnected *knn*–graph there is one pair of closest vertices lying in different components. It can easily be shown that there must be a an angle of at least 120° in which none of either's $k$ next neighbors is placed. A simple stochastic argument shows that the probability for the existence of any vertex with this property is given by $(2/3)^k$. Equating this with the probability bound of $1/n$ and solving the equation yields the needed $k$ such that with high probability not even one vertex with the above mentioned property exists. This leads to the following lemma:

**Lemma 4** *A knn–graph is connected with high probability for* $k \geq \frac{\log n}{\log(3/2)}$

Note that the probability for an unconnected *knn*–graph is smaller than $1/n$ since the existence of at least one vertex with the above given property is only necessary for an unconnected graph but certainly not sufficient.

We will now show that connected commensurate partitions can be found.

### C. Constructing the partition

The following procedure constructs partitions as required by the definition for *regularly decomposable* graph families. The needed size $s$ of the subgraphs is depending on the probability $p = \frac{1}{f(n)\cdot n}$ of the added random graph $G(n,p)$:

$$s(n) = f(n) \cdot \log n^{1+\epsilon} \tag{31}$$

FIG. 3: This figure shows the result of the partitioning procedure for $s = 2$ as described in IV C. Each square contains more than $4/\pi \cdot s \approx 2.5$ vertices. Each circle within any quadratic region contains at least $s = 2$ vertices that must form a connected subgraph. Note that the distribution of points is only schematic.

The partitioning algorithm must be capable of finding for each pair of vertices a partition into $\Theta(n/s)$ subgraphs such that both vertices are included in some full subgraph of size $s$. To guarantee this, we construct slightly different partitions for each pair $v, w$ of vertices. For each of them, we start with a geometric partition, based on squares containing at least $4/\pi \cdot s$ vertices. The exact positions for the squares are chosen such that both vertices are contained in full subgraphs. Beside this requirement the positions of the squares are arbitrary as long as the number of squares placed completely inside the unit square is maximal. Note, that a constant relative fraction of vertices may exist, that is not contained in any subgraph. Each of the squares covers an area $A_s$ so that with high probability at least $4/\pi \cdot s$ vertices are geometrically contained in each of them. The area is given by $A_s > \frac{4}{\pi} \cdot \pi \cdot \hat{r}^2$, where $\hat{r}$ denotes the maximal expected knn–disc radius (Lemma 3).
The maximal (centered) circle (Fig. 3) within each $A_s$ contains only vertices from the same connected component. Otherwise at least one vertex would have an arc of more than 120° without any *knn*–edge which is highly unlikely as was already shown in Lemma 4. The area of this circle covers $\pi/4$ of $A_s$. We expect therefore that at least $\pi/4 \cdot 4/\pi \cdot s = s$ vertices from the same connected component for each $A_s$. As we explained before, for each constructed partition a constant fraction of vertices can be disregarded.

Figure 3 shows an example for a partition for $s = 2$.

Note that the expected diameter $D(S_i)$ of the subgraphs is expectedly scaling with $O(\sqrt[d]{s})$ as it is the case with grid graphs.

With this we have shown that a locally clustered graph family is also regularly decomposable if it is possible to divide the space in which it is located into $\Theta(n/s(n))$ areas that contain connected subgraphs of a wanted size $s(n)$.

## V. DISCUSSION

In this paper we have proposed a general framework for the design of hybrid graph families with a small–world effect: We have shown that certain combinations of a regularly decomposable graph family and a random graph component $G(n, p)$ belong to that family. Furthermore, we have given a generalized theorem that describes an upper bound for the diameter of these hybrid graphs in dependence of the structure of both, the regularly decomposable and the random graph component. We just want to note here, that the family of graphs with a small–world effect is not restricted to the hybrid graphs analyzed in depth here, but also includes all the classical small–world models [2, 9, 11–13].

The last years have seen new variants of small–world models that try to capture different network properties in addition to the sheer smallness of the network, e.g., the *navigability* approach of Kleinberg [11, 12], or the hybrid power law graphs of Chung and Lu [13]. To find models for real–world networks that show different additional properties seems to be an important task for the next years. This framework provides high flexibility in tuning a model to simulate a given real–world complex system because every wanted graph property, as for example, a scale–free degree distribution [22], a certain assortativity [23], or a wanted clustering coefficient, can now be put into the non-random graph component and will thus also appear in the hybrid graph. As long as these wanted properties do not hinder the regular decomposability of the graph family, the small–world effect will be maintained and, simultaneously, the wanted graph property will also show up in the hybrid graph since the random graph component is so sparse that it will not affect most properties, i.e., degree distributions, clustering coefficient, or assortativity, the one exception being — of course— the diameter and related distance measures.

Watts and Strogatz have provided us with the first formal model for generating small–world networks. The beauty of their model lies in its simplicity and clarity. In their model, the high clustering coefficient gives an indication whether there is a local graph component or not. Our main focus in this paper lies on describing a large family of hybrid graphs as a new way to modelling an interesting real–world network, the focus is not so much on whether a given network shows the small–world effect. Whether a network shows the small–world effect or not should — in our opinion — be decided by looking at the generating process and not so much by looking at the result of it, i.e., the network: If the network

generating process prefers building local edges and sometimes generates only few random long-distance edges, the resulting network will show the small-world effect. In the case of all social networks, including co–authorship networks, citation networks, or acquaintanceship networks, it is intuitive that most of the relationships are dominated by the environment and will thus be 'local', implying a large diameter of this component. Additionally, almost every person has these long–distance relationships to persons met somewhere, in an almost random fashion. This part of the network is certainly sparse with respect to the total number of possible relationships and should thus also show a high diameter. In summary, the generating process of social networks of all kind can be meaningfully modelled by two different graph components, a local and a random part, and thus we would argue that social networks show the small-world effect. In summary, the small–world effect directs the focus towards network generating processes and thus to families of graphs that are generated by the same process.

Of course, our definition of the small–world effect is somewhat influenced by our personal impression of what small–worlds really are. Since we are aware of that problem we want to conclude our discussion with the introductory quotation of Wiener and Rosenblueth in "The Role of Models in Science":

> *[...] the best material model for a cat is another, or preferably the same cat.* - N. Wiener and A. Rosenblueth [24]

---


[1] S. Milgram, Psychology Today **1**, 61 (1967).

[2] D. J. Watts and S. Strogatz, Nature **393**, 440 (1998).

[3] L. Adamic, in *Proceedings of ECDL'99 - Lecture Notes on Computer Science* (1999), pp. 443–452.

[4] L. Adamic and E. Adar, Social Networks **27** (2005).

[5] P. Dodds, R. Muhamad, and D. J. Watts, Science **301**, 827 (2003).

[6] A. Iamnitchi, M. Ripeanu, and I. Foster, in *Proceedings of the IEEE INFOCOM 2004* (2004).

[7] L. Lago-Fernandez, R. Huerta, F. Corbacho, and J. Siguenza, Phys. Rev. Letters **84**, 2758 (2000).

[8] M. E. J. Newman and D. J. Watts, Phys. Rev. E **60**, 7332 (1999).

[9] M. E. J. Newman and D. J. Watts, Phys. Lett. A **263**, 341 (1999).

[10] S. Dorogovtsev and J. Mendes, Europhys. Lett. **50** (2000).

[11] J. Kleinberg, in *Proceedings of the 32nd ACM Symposium on Theory of Computing* (2000).

[12] J. Kleinberg, Nature **406**, 845 (2000).

[13] F. Chung and L. Lu, in *Complex Networks (E. Ben-Naim, H. Frauenfelder, Z. Toroczkai (eds.))* (2004), pp. 91–106.

[14] R. Andersen, F. Chung, and L. Lu, in *Algorithms and Models for the Web-Graph; Third International Workshop, WAW 2004* (2004), LNCS 3243, pp. 19–30.

[15] B. Bollobás and F. Chung, SIAM Journal on Discrete Mathematics **1**, 328 (1988).

[16] E. N. Gilbert, Anual Math. Statist. **30**, 1141 (1959).

[17] M. Gaertler, *Network Analysis: Methodological Foundations* (Springer-Verlag, 2005), chap. Clustering, pp. 178–215.

[18] M. Girvan and M. Newman, Proceedings of the National Academy of Sciences **99**, 7821 (2002).

[19] B. Bollobás and O. M. Riordan, in *Handbook of Graphs and Networks (S. Bornholdt, H.G. Schuster (eds.))* (2003), pp. 1–34.

[20] S. P. Fekete, M. Kaufmann, A. Kröller, and K. A. Lehmann, in *Proceedings of the 17th Canadian Conference on Computational Geometry* (2005).

[21] P. S. Bearman, J. Moody, and K. Stovel, American Journal of Sociology **110**, 44 (2004).

[22] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[23] M. E. Newman, Physical Review Letters **89** (2002).

[24] N. Wiener and A. Rosenblueth, Philosophy of Science **12**, 316 (1945).