Bounded expansion in web graphs

SILVIA GAGO, DIRK SCHLATTER

Abstract. In this paper we study various models for web graphs with respect to bounded expansion. All the deterministic models even have constant expansion, whereas the copying model has unbounded expansion. The most interesting case turns out to be the preferential attachment model — which we conjecture to have unbounded expansion, too.

Keywords: graph minors, bounded expansion, webgraphs
Classification: 05C83, 94C15, 90B10, 90B15

1. Introduction

Many complex networks which underlie phenomena in the real world as diverse as the World Wide Web, energy infrastructures, or biological systems share some characteristics like a small diameter, high local clustering, or a power-law degree distribution. For the past decade, there has been a growing interest in finding suitable models for such networks. A model typically consists of (an algorithm which produces) an infinite sequence of graphs of increasing order.

In this paper, we will only consider models which produce undirected graphs with a small diameter and whose degree distribution follows a power law. ‘Small’ means that the diameter grows much slower than the order of the graph, e.g. like $O(\log |V(G)|)$. In the context of social networks this property implies that any two persons are linked via a relatively small number of other people — therefore the term small-world is used to describe this property. We note that some authors use the average distance between two vertices instead of the maximum one. The degree distribution of a graph $G$ follows a power law if there are positive constants $c$ and $\alpha$ such that

$$\frac{|\{v \in V(G) \mid \deg_G(v) = k\}|}{|V(G)|} \approx c \cdot k^{-\alpha}$$

for a large range of $k$. When this happens, we also say that $G$ is scale-free. Moreover, we often abbreviate small-world, scale-free graphs as web graphs. For further information on web graphs in general we refer to the surveys of Bollobás and Riordan [5] and Bonato [8] and the references therein.

Some of the models for web graphs are defined deterministically and others use randomness, but they all produce sparse graphs, i.e. graphs where the number

Research supported by the European project Aeolus.
of edges is only linear in the number of vertices. It is therefore natural to ask whether these graphs fall into some of the well-known sparse graph families like minor-closed graph families or families of graphs with bounded degree.

Nešetřil and Ossona de Mendez [19], [20], [21] generalized such families by defining a sequence of graph parameters as follows. Let $G$ and $H$ be graphs with $V(H) = \{v_1, \ldots, v_h}\}$. We say that $H$ is a minor of $G$ of depth at most $r$ (and write $H \preceq_r G$) if there are disjoint subsets $V_1, \ldots, V_h$ of $V(G)$ such that

(i) each subgraph $G[V_i]$ is a connected subgraph of radius at most $r$
(ii) if $v_i v_j \in E(H)$ then there exist $u \in V_i$ and $w \in V_j$ with $uw \in E(G)$.

We can then define the greatest reduced average density of rank $r$ of $G$ as

$$\nabla_r(G) := \max_{H \preceq_r G} \frac{|E(H)|}{|V(H)|}.$$ 

If the (non-decreasing) sequence $(\nabla_r(G))_{r \geq 0}$ has a uniform upper bound $f(r)$ for all graphs in a certain family, then we say that this family has bounded expansion. In other words, a graph family $\mathcal{G}$ has bounded expansion if there exists a function $f : \mathbb{N} \cup \{0\} \to \mathbb{R}$ such that $\nabla_r(G) \leq f(r)$ for every graph $G \in \mathcal{G}$ and non-negative integer $r$.

Naturally, the question arises whether web graphs have bounded expansion or not. This is not only of theoretical importance; Nešetřil and Ossona de Mendez [19], [22] showed that graph families with bounded expansion also have low tree-width and low tree-depth colourings — which implies that many algorithmic graph problems which are difficult in general become feasible (cf. [23]).

Using the fact that every graph in a proper minor-closed graph family is $d$-degenerate for some constant $d$, the following observation is straightforward.

**Proposition 1.** Let $\mathcal{G}$ be a proper minor-closed graph family. Then there exists a constant $c$ such that $\nabla_r(G) \leq c$ for every graph $G \in \mathcal{G}$ and non-negative integer $r$.

We note that the converse is also easy to see: every graph family with constant expansion is contained in a proper minor-closed graph family.

A trivial lower bound for $\nabla_r(G)$ follows easily from its definition. Recall that a $\leq 2r$-subdivision of a graph $H$ is the graph from $H$ by subdividing each edge with at most $2r$ vertices.

**Proposition 2.** If a graph $G$ contains a $\leq 2r$-subdivision of another graph $H$ with minimum degree $\delta(H)$, then

$$\nabla_r(G) \geq \frac{|E(H)|}{|V(H)|} \geq \frac{\delta(H)}{2}.$$ 

Again, we note that there is a converse: Dvořák [13] proved that, for a graph $G$ and $r, d \in \mathbb{N}$, $\nabla_r(G) \geq 4(4d)^{(r+1)^2}$ implies that $G$ contains a $\leq 2r$-subdivision of a graph with minimum degree $d$. 

2. Deterministic models

2.1 Recursive clique-trees. Recursive $d$-clique-trees (for an integer $d \geq 2$) are constructed as follows: starting with $G_0 := K_d$, we obtain $G_{t+1}$ from $G_t$ by adding a new vertex for each clique of size $d$ in $G_t$ and joining this vertex to all vertices in the respective clique. The case $d = 2$ has been considered in [11] and the general case in [10].

We will denote the family of recursive $d$-clique-trees by $\mathcal{P}_d$. If we introduce only a single new vertex in each step (and connect it to all vertices in some $d$-clique), we obviously get a larger graph family $\mathcal{P}_d'$ which contains $\mathcal{P}_d$. Denote the closure of $\mathcal{P}_d'$ under taking subgraphs by $\mathcal{P}_d''$.

Lemma 3. $\mathcal{P}_d''$ is a proper minor-closed graph family.

The proof of this lemma is not necessary as $\mathcal{P}_d''$ is the class of partial $d$-trees (i.e. of graphs with tree-width at most $d$), and it is well known that it is minor closed. Due to Proposition 1, we thus get the following corollary.

Corollary 4. For each $d \geq 2$, the family of recursive $d$-clique-trees has constant expansion.

2.2 Apollonian networks. The construction of a $d$-dimensional Apollonian network is very similar to the construction of a recursive $d$-clique-tree, only that we now introduce new vertices for each clique of size $d$ in $G_t$ which does not already lie in $G_{t-1}$. The case $d = 2$ has been considered in [27], the case $d = 3$ in [2], [12], and the general case in [26]. Since the family $\mathcal{Q}_d$ of all such networks is a subset of $\mathcal{P}_d$, Corollary 4 implies that $\mathcal{Q}_d$ also has constant expansion.

2.3 Hierarchical networks. Again, we construct graphs inductively and start with $G_0 := K_d$, for an integer $d \geq 2$. Select a root $r$ in $G_0$ and let $N_0$ be the set of non-root vertices. We construct $G_{t+1}$ from $G_t$ as follows: Add $d-1$ disjoint copies $G_t^{(1)}, \ldots, G_t^{(d-1)}$ of $G_t$ to $G_t$ and connect all the vertices in $\bigcup_{i=1}^{d-1} N_t^{(i)}$ to $r$. Finally, set $N_{t+1} := N_t \cup \bigcup_{i=1}^{d-1} N_t^{(i)}$. This model was introduced in [4] and further studied in [15], [24], [25].

We denote the family of all such hierarchical networks by $\mathcal{R}_d$. If we connect all vertices in the copies of $G_t$ to $r$, we get another graph family $\mathcal{S}_d$. The closure of $\mathcal{S}_d$ under taking subgraphs is obviously closed under deletion and contraction of edges. $\mathcal{R}_d$ is thus contained in a proper minor-closed graph family and by Proposition 1 we get the following result.

Proposition 5. For each $d \geq 2$, the family of hierarchical networks with parameter $d$ has constant expansion.

3. Stochastic models

If we want to discuss whether stochastic models also have bounded expansion, we first have to clarify what we mean by that. For a given stochastic model, we could of course ask whether the family of all graphs which might occur as the
outcome, i.e. which have positive probability, has bounded expansion. However, this approach somehow seems to neglect the randomness and in most cases it trivially leads to the result that the expansion is unbounded. Instead, we will adopt the following

**Definition 6.** A random graph process \((G_t)_{t \geq 0}\) has *bounded expansion* if there exists a class with bounded expansion \(\mathcal{C}\) such that \(G_t\) belongs to \(\mathcal{C}\) asymptotically almost surely.

We will discuss the two well-known stochastic models for web graphs, the copying model of Kumar, Raghavan, Rajagopalan, Sivakumar, Tomkins, and Upfal [17] and the preferential attachment model of Barabási and Albert [3].

### 3.1 Copying.

The linear growth copying model was first introduced in [16] and rigorously studied in [17]. Their original model generates directed graphs: Let \(G_0\) be a directed graph in which every vertex has out-degree \(d\). To get from \(G_t\) to \(G_{t+1}\) we first add a vertex \(v_{t+1}\) and choose a ‘prototype’ vertex \(v\) from \(V(G_t)\) uniformly at random. Let \(u_1, \ldots, u_d\) be the vertices at which edges from \(v\) arrive and let \(p \in (0, 1)\) be a constant. For \(1 \leq i \leq d\) we then add an edge \((v_{t+1}, u_i)\) with probability \(p\); else, we add the edge \((v_{t+1}, w)\), where \(w\) is again chosen uniformly at random from \(V(G_t)\).

The intuition behind this model can be best explained in the context of the www graph. A new website is likely to be concerned with a certain topic and its author will probably — consciously or unconsciously — copy some of the links of an already existing website concerned with the same topic. As the choice of the prototype vertex is uniform, popular topics are more likely to attract new websites. And the ‘error’ case, when we add \((v_{t+1}, w)\) instead of \((v_{t+1}, u_i)\), reflects that a new website might also add a new perspective on the topic, linking to another website not previously related to the topic.

There are various variants and generalisations of this model. For the sake of simplicity, we will work with undirected graphs and start with \(K_2\). Moreover, we use a copying model without error, i.e. without the ‘else’ case from above. Thus, we define a random graph process \((G_t)_{t \geq 0}\) inductively as follows.

1. Start with a graph \(G_0\) consisting of two vertices \(u, w\) and the edge \(uw\).
2. Given \(G_t\), choose a vertex \(v \in V(G_t)\) uniformly at random. Add a new vertex \(v_{t+1}\) and join it to each neighbour of \(v\) (independently) with some constant probability \(p \in (0, 1)\).

As noted in [17], the linear growth copying model contains many large complete bipartite graphs (as does the www graph, see [18]). Hence the following result is not surprising.

**Proposition 7.** For all \(d \in \mathbb{N}\),

\[
\mathbb{P}[G_t \text{ contains } K_{d,d}] \longrightarrow 1 \quad \text{as} \quad t \longrightarrow \infty.
\]

**Proof:** For \(1 \leq i < t^{1/2}\) we define the events
A_i: \( w \) is the prototype chosen for \( v_i \) and \( vw_i \in E(G_i) \)
and for \( 1 \leq j \leq \log t/2 \) we set

\[
B_j := \bigcup_{i=2^{j-1}}^{2^j-1} A_i.
\]

Because we clearly have \( \mathbb{P}[A_i] = p/(i+1) \), we get

\[
\mathbb{P}[B_j] = 1 - \prod_{i=2^{j-1}}^{2^j-1} (1 - \mathbb{P}[A_i]) \geq 1 - \left(1 - \frac{p}{2^j}\right)^{2^j-1} \geq 1 - e^{-p/2}.
\]

Let us denote the indicator variable of \( B_j \) by \( X_j \) and set
\[
X := \sum_{j=1}^{\log t/2} X_j.
\]

Observe that a binomially distributed random variable \( Y \) with parameters \( \log t/2 \)
and \( 1 - e^{-p/2} \) is concentrated around its expectation \( (1 - e^{-p/2}) \log t/2 \),
we obtain that \( X > (1 - e^{-p/2}) \log t/4 \) with high probability. In other words, we can be
almost sure that \( w \) has \( O(\log t) \) neighbours in \( G_{t^{1/2}-1} \).

For \( t^{1/2} \leq i < t \) we then define the events

\[ C_i: \ w \text{ is the prototype chosen for } v_i \text{ and } v_i \text{ is connected to the}
\text{d oldest neighbours of } w
\]
and for \( \log t/2 < j \leq \log t \) we set

\[ D_j := \bigcup_{i=2^{j-1}}^{2^j-1} C_i. \]

Now we have \( \mathbb{P}[C_i] = p^d/(i+1), \) and as above we can deduce that with high
probability there are \( O(\log t) \) vertices in \( G_t \) which together with the \( d \) oldest
neighbours of \( w \) form a complete bipartite graph. \( \square \)

We note that Proposition 7 also follows from a result of Bonato and Janssen [9]
about the limit of such graph sequences. Together with Proposition 2, it implies
that \( (G_t)_{t \geq 0} \) has unbounded expansion in the sense of Definition 6.

**3.2 Preferential attachment.** The key idea of the preferential attachment
model is simple and yet intriguing. As in the copying model, in each step we
add a new vertex into the graph. In order to connect it to the rest of the graph
we randomly choose \( m \) vertices with probability \textit{proportional to their degrees}
and connect them to the new vertex. One of the reasons for the popularity of this
model seems to be the plausibility of its construction: for example, when a new
website goes online, it seems to be more likely to link to a website which is already
popular than to one which is not.

Preferential attachment was first suggested as a model for web graphs by
Barabási and Albert [3] and in recent years, various variants and generalisations
have been studied. We will adopt the following rigorous definition of Bollobás
and Riordan [6]. First we generate a sequence of graphs \( (G^1_t)_{t \geq 0} \) as follows.
(1) Start with a graph $G_0^{(1)}$ consisting of a single vertex $v_1$ and the loop $v_1v_1$.

(2) Given $G_t^{(1)}$, add a new vertex $v_{t+1}$ and an edge $vv_{t+1}$, where $v$ is chosen randomly with

$$P[v = v_i] = \begin{cases} \frac{d_{G_t^{(1)}}(v_i)}{(2t - 1)} & \text{if } 1 \leq i \leq t, \\ \frac{1}{(2t - 1)} & \text{if } i = t + 1. \end{cases}$$

To get $(G_t^{(m)})_{t \geq 0}$ from $(G_t^{(1)})_{t \geq 0}$ for some fixed integer $m \geq 2$, we take those graphs from the latter sequence for which $m$ divides $t$ and contract $v_1, \ldots, v_m$ to a new vertex $v_m$ (deleting multiple edges), $v_{m+1}, \ldots, v_{2m}$ to a new vertex $v_{2m}$, and so forth. In order to simplify notation, we will assume from now on that $V(G_t^{(1)}) = \{1, \ldots, t\}$.

Bollobás and Riordan [6] proved that the graph sequence thus constructed is indeed small-world. More precisely, they proved that, with high probability, $\text{diam}(G_t^{(m)}) \sim \ln n / \ln \ln n$. Likewise, Bollobás, Riordan, Spencer, and Tusnády [7] showed that $G_t^{(m)}$ is scale-free and determined the exponent. Their results hold for vertices up to degree $t^{1/15}$; Flaxman, Frieze, and Fenner [14] showed that the $k$ largest vertices (for some constant $k$) are distributed around $t^{1/2}$ (and separated from each other).

In view of Proposition 2, we might ask whether $G_t^{(m)}$ is likely to contain a $\leq 2r$-subdivision of $K_d$, say. We conjecture an even stronger result.

**Conjecture 8.** Let $H$ be a 1-subdivision of $K_d$ for some $d \in \mathbb{N}$. Then

$$P[G_t \text{ contains a copy of } H] \longrightarrow 1 \quad \text{as} \quad t \longrightarrow \infty.$$ 

In order to justify our conjecture, we will use a result of Bollobás and Riordan [5] about the containment of a fixed subgraph in $G_t^{(1)}$. Let us introduce some definitions first. We will temporarily orient every edge $uv$ in $G_t^{(1)}$ from $v$ to $u$ whenever $u \leq v$. Let $S$ be a feasible subgraph of $G_t^{(1)}$, i.e. an oriented graph of fixed size such that $P[S \subseteq G_t^{(1)}] > 0$ for any $t$ such that $V(S) \subseteq \{1, \ldots, t\}$. Then $\deg_{S}(\cdot)$, $V^{-}(S)$, and $V^{+}(S)$ have their obvious meaning:

$$\deg_{S}(v) := |\{w \in V(S) \mid (v, w) \in E(S)\}|,$$

$$V^{-}(S) := \{v \in V(S) \mid \exists (v, w) \in E(S)\} = \left\{v \in V(S) \mid \deg_{S}(v) > 0\right\},$$

$$V^{+}(S) := \{v \in V(S) \mid \exists (u, v) \in E(S)\}.$$ 

Furthermore, we count the edges ‘crossing’ a vertex $v$ of $S$:

$$C_S(v) := |\{(u, w) \mid u \leq v \leq w\}|.$$
Theorem 9 ([5]). Let $S$ be a feasible subgraph of $G_t^{(1)}$. Then

$$\mathbb{P}[S \subseteq G_t^{(1)}] = \prod_{v \in V^-(S)} \deg_{S}^{\text{in}}(v)! \prod_{v \in V^+(S)} \frac{1}{2v-1} \prod_{v \notin V^+(S)} \left(1 + \frac{C_S(v)}{2v-1}\right).$$

Furthermore,

$$\mathbb{P}[S \subseteq G_t^{(1)}] = \prod_{v \in V^-(S)} \deg_{S}^{\text{in}}(v)! \prod_{uv \in E(S)} \frac{1}{2\sqrt{uv}} \cdot \exp \left(\sum_{v \in V(S)} O\left(\frac{C_S(v)^2}{v}\right)\right).$$

This allows us to prove the following result.

Lemma 10. Let $X$ count the number of subgraphs in $G_t^{(m)}$ which are (isomorphic to) a $1$-subdivision of $K_d$ such that the vertices of $K_r$ correspond to the first $d$ vertices of $G_t^{(m)}$, for integers $d \geq 2$ and $m \geq 2$. Then

$$\mathbb{E}[X] \rightarrow \infty \quad \text{as} \quad t \rightarrow \infty.$$

**Proof:** We may assume without loss of generality that we count only subgraphs where the $D := \binom{d}{2}$ subdividing vertices all lie in $(t^{1/2}, t]$. Observe that there are $h := (t-t^{1/2})/(t-t^{1/2} - D)!$ such subgraphs which could appear in $G_t^{(m)}$. We will denote them by $H_1, \ldots, H_h$ and the subdividing vertices of $H_i$ by $v_1^{(i)}, \ldots, v_D^{(i)}$. For every $1 \leq i \leq h$, we fix a feasible subgraph $H_i'$ of $G_{mt}^{(1)}$ such that, when we perform the contractions to get $G_t^{(m)}$ from $G_{mt}^{(1)}$ as described above, we get $H_i$ from $H_i'$.

We now use equation (1) to calculate $\mathbb{P}[H_i' \subseteq G_{mt}^{(1)}]$. Clearly, the only vertices of $H_i'$ with positive indegree can be those which correspond to the vertices $1, \ldots, r$ in $H_i$. Moreover, as all the subdividing vertices are larger than $mt^{1/2}$, the contribution of the these vertices to the sum in the exponent of the last product is negligible. Therefore, we can deduce that there is a constant $c$ such that, for every $1 \leq i \leq h$,

$$\mathbb{P}[H_i' \subseteq G_{mt}^{(1)}] \sim \frac{c}{v_1^{(i)} \cdots v_D^{(i)}}.$$ 

Denoting the indicator variable of this event by $X_i$, we thus have

$$\mathbb{E}[X] > \sum_{i=1}^{h} \mathbb{E}[X_i] \sim \sum_{i=1}^{h} \frac{c}{v_1^{(i)} \cdots v_D^{(i)}} = \Theta \left((\ln n)^D\right).$$

\qed
Unfortunately, we cannot use the second moment method to prove Conjecture 8. The reason is as follows. We would need to prove that \( \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2} \to 1 \), but the dominant contribution in the numerator arises from pairs \( X_iX_j \) where the corresponding sets of subdividing vertices are disjoint. If we use Equation (1) to calculate \( \mathbb{E}[X_iX_j] \), we get a different constant \( c' \). To show that \( \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]^2} \to 1 \) we would need \( c' \leq c^2 \), but it is not difficult to show that \( c' > c^2 \).

Of course, the random graph process generated by the preferential attachment model might have unbounded expansion, nevertheless. For example, we may hope that a different approach proves that \( \mathbb{P}[X > 0] \) with high probability. But even if that statement is false, Conjecture 8 could still hold, since we were counting rather special 1-subdivisions of \( K_d \). And finally, we note that by Proposition 2 we only have to find, with high probability, a \( \leq 2r \)-subdivision of a graph with minimum degree \( d \), for some \( r \in \mathbb{N} \) and all \( d \in \mathbb{N} \).

4. Conclusion

We have seen that there is no clear answer to the question whether web graphs have bounded expansion. For the deterministic web graph models we considered the answer is affirmative, whereas for the copying model the answer is negative and we expect the same for the preferential attachment model.

On the other hand, note that it is easy to have deterministic models with unbounded expansion (for instance, generating graphs with unbounded average degree). Random models with bounded expansion also exist (for instance, Nešetřil and Ossona de Mendez announced that for each fixed \( d > 0 \) there exists a bounded expansion class \( \mathcal{C}(d) \) such that random graphs with edge probability \( d/n \) in the Erdos-Renyi model a.a.s. belong to \( \mathcal{C}(d) \)).

The question also remains whether real web graphs have bounded expansion. Of course, there can be no answer in terms of a mathematical proof: despite the fact that web graphs are typically massive and constantly growing, we will always have only a finite set of finite graphs available for analysis. But more experimental results such as [18] could at least provide a hint what the answer for web graph models should be — which would help to decide which models are more suitable than others.

Acknowledgment. The authors would like to thank Jaroslav Nešetřil for very helpful discussions. This research was supported by AEOLUS integrated project IST-015964 (Algorithmic Principles for Building Efficient Overlay Computers).

References

Bounded expansion in web graphs


Universitat Politècnica de Catalunya, Department de Matemàtica Aplicada IV, Avda. Canal Olímpic s/n, 08860 Castelldefels, Barcelona, Spain

Email: sgago@ma4.upc.edu

Charles University, Institute for Theoretical Computer Science and Department for Applied Mathematics, Malostranské nám. 25, 118 00 Praha, Czech Republic

(Received March 14, 2008, revised January 25, 2009)