Biclique Coverings, Rectifier Networks and the Cost of ϵ-Removal

Szabolcs Iván, Ádám Lelkes, Judit Nagy-György, Balázs Szörényi, György Turán

▶ **To cite this version:**

Szabolcs Iván, Ádám Lelkes, Judit Nagy-György, Balázs Szörényi, György Turán. Biclique Coverings, Rectifier Networks and the Cost of ϵ-Removal. 16th International Workshop on Descriptional Complexity of Formal Systems, Proceedings, Aug 2014, Turku, Finland. pp.174 - 185, 10.1007/978-3-319-09704-6_16 . hal-01079368

HAL Id: hal-01079368
https://inria.hal.science/hal-01079368

Submitted on 1 Nov 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:1406.0017v1 [cs.FL] 30 May 2014

Biclique coverings, rectifier networks and the cost of ε-removal

Szabolcs Iván[1⋆], Ádám D. Lelkes[2], Judit Nagy-György[1⋆⋆], Balázs Szörényi[3,4], and György Turán[2,3⋆⋆⋆]

[1] University of Szeged, Hungary
[2] University of Illinois at Chicago
[3] MTA-SZTE Research Group on Artificial Intelligence
[4] INRIA Lille, SequeL project, France

Abstract. We relate two complexity notions of bipartite graphs: the *minimal weight biclique covering number* $\mathrm{Cov}(G)$ and the *minimal rectifier network size* $\mathrm{Rect}(G)$ of a bipartite graph G. We show that there exist graphs with $\mathrm{Cov}(G) \geq \mathrm{Rect}(G)^{3/2-\epsilon}$. As a corollary, we establish that there exist nondeterministic finite automata (NFAs) with ε-transitions, having n transitions total such that the smallest equivalent ε-free NFA has $\Omega(n^{3/2-\epsilon})$ transitions. We also formulate a version of previous bounds for the weighted set cover problem and discuss its connections to giving upper bounds for the possible blow-up.

1 Introduction

In the world of descriptive complexity, questions involving the possible blow-up when transforming a description of some mathematical object from a formalism to another is a central topic, with one of the first papers dating back to 1971 [13]. We are primarily interested in the cost of chain rule removals from context-free grammars (CFGs). That is, how large a chain-rule free CFG has to be in the worst case which is equivalent to an input CFG of size n, having chain rules? The obvious upper bound resulting from the standard transformation is $O(n^2)$. The best known lower bound is $\Omega(n^{3/2-\epsilon})$ [2]. The question is interesting since chain rule elimination is the bottleneck part of the transformation to Chomsky Normal Form. Despite the question being well-motivated, we have no knowledge of progress in the last three decades; the gap is still there.

The maximal possible blow-up is not known even in the special case of regular languages. When a *regular* language is given (e.g. by a nondeterministic

⋆ This research was supported by the European Union and the State of Hungary, co-financed by the European Social Fund in the framework of TÁMOP 4.2.4.A/2-11-1-2012-0001 National Excellence Program.

⋆⋆ Supported by the European Union and co-funded by the European Social Fund under the project "Telemedicine-focused research activities on the field of Mathematics, Informatics and Medical sciences" of project number "TÁMOP-4.2.2.A-11/1/KONV-2012-0073"

⋆⋆⋆ Partially supported by NSF grant CCF-0916708

automaton or NFA, possibly having ε-transitions), an equivalent "chain-rule-free" *regular* grammar corresponds to a nondeterministic automaton with no ε-transitions. In order to define the "blow-up", we have to choose a notion for measuring the *size* of an NFA – we say that the size of an NFA is the number of its *transitions*. Regular languages can be represented by a variety of different formalisms, some of which are more concise than the others. For example, transforming a regular expression (RE) to an equivalent NFA can be done within linear bounds, i.e. the cost of this direction is worst-case $\Theta(n)$. From RE to ε-free NFA the worst-case cost is $\Theta(n \log^2 n)$, by the upper bound result of [4] and the matching lower bound of [16]. The lower bound is achieved with a language possessing a linear-size RE as well, thus it is recognized by an NFA of size $O(n)$, hence the cost of the NFA $\to$ ε-free NFA transformation is $\Omega(n \log^2 n)$. However, the gap between $\Omega(n \log^2 n)$ and $O(n^2)$ has not been reduced since 2006. It is also known that from ε-free NFA to RE an exponential blow-up can occur and Kleene's algorithm produces an RE of exponential size from an NFA.

One of the main results of the paper is that the NFA $\to$ ε-free NFA transformation has worst-case cost $\Omega(n^{3/2-\epsilon})$ for any $\epsilon > 0$. It is interesting that this bound (as well as the upper bound $O(n^2)$) coincides with that of [2] for the seemingly more general problem of chain rule elimination. The methods (as well as the models) are very different but there is also a similarity: for the lower bound of [2], languages consisting of words of length 3 were defined. In our case, we consider languages consisting of words of length 2. Such languages $L \subseteq \Sigma\Delta$ can be viewed as bipartite graphs $G_L = (\Sigma, \Delta, E_L)$ with (a, b) being an edge in the graph iff the word ab belongs to L. When the language is viewed this way, ε-free NFAs recognizing L correspond to biclique coverings [7] of G_L with the size of an NFA corresponding to the weight of the associated biclique covering. Also, NFAs recognizing L correspond to rectifier networks [7] realizing G_L; again, with the size of an NFA corresponding to the size of the associated network.

Hence, proving worst-case lower bounds for the minimum-weight covering of a bipartite graph having a rectifier network of size n, we get as byproduct worst-case lower bounds for the NFA $\to$ ε-free NFA transformation. Thus the bulk of the paper discusses biclique coverings and rectifier networks. These have also been studied for a long time in various contexts, see Sections 2 and 3.

The paper is organized as follows. In Section 2 we give the notations we use for graphs and automata. In Section 3 we give lower bounds for the possible blow-up between rectifier network size and biclique covering weight. In Section 4 we give upper bounds for this blow-up and consider the biclique covering problem as a weighted set cover problem. An approximation bound for the greedy algorithm given by Lovász [10] for the unweighted case is generalized to the weighted case. We discuss the connection of this bound to possible upper bounds for the blow-up. In Section 5 we relate these graph-theoretic results to automata theory and prove the aforementioned lower bound of $\Omega(n^{3/2-\epsilon})$ for ε-removal.

2 Notations

Graphs, biclique coverings and rectifier networks Let $[n]$ stand for the set $\{1, \ldots, n\}$. For sets A and B, $K_{A,B}$ stands for the complete bipartite graph $(A, B, A \times B)$. When only the cardinalities a and b of the sets A and B matter, we write $K_{a,b}$ for $K_{A,B}$. When $G = (A, B, E)$ is a bipartite graph, a *biclique* of G is a complete bipartite subgraph of G and the *weight of a biclique* is the number of its vertices. A *biclique covering* of G is a collection $\mathcal{C}$ of its bicliques such that each edge of G belongs to at least one member of $\mathcal{C}$, the *weight of a covering* is the sum of the weights of the bicliques present in the covering and $\mathrm{Cov}(G)$ is the minimum possible weight of a biclique covering of G.

A biclique $K_{a,b}$ has weight $a + b$ while it covers ab edges of G. In our investigations we will frequently use the inverse $\frac{ab}{a+b}$ of the relative cost of covering the edges by $K_{a,b}$. We introduce the shorthand $\mathrm{H}(a, b)$ to denote the quantity $\frac{ab}{a+b}$.

For a bipartite graph $G = (A, B, E)$, a *rectifier network realizing* G is a directed acyclic graph (DAG) $R = (V, E')$ with A being the set of source nodes of R and B being the set of sink nodes of R, satisfying the property that $(a, b) \in E$ if and only if b is reachable from a in R. The *size* of a rectifier network is the number of its edges. The *depth* of a network is the length of its longest path. We let $\mathrm{Rect}(G)$ stand for the size of the smallest rectifier network realizing G and $\mathrm{Rect}_k(G)$ for the size of the smallest rectifier network of depth at most k realizing G. We may assume *w.l.o.g.* that there are no isolated vertices.

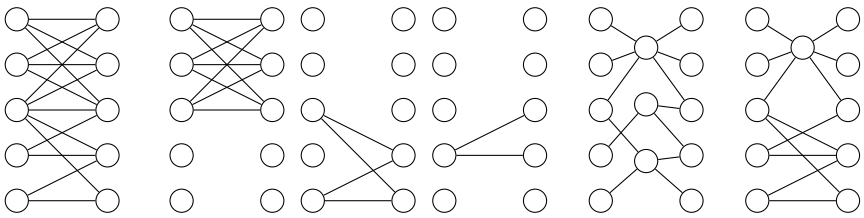

Fig. 1: From left to right: a graph G, three bicliques showing $\mathrm{Cov}(G) \leq 13$, a depth-2 network corresponding to the bicliques having size 13, and another network showing $\mathrm{Rect}_2(G), \mathrm{Rect}(G) \leq 12$. In the networks, edges are directed from left to right.

There are constructions of graphs for which only large rectifier networks exist (i.e. having large Rect value), the dates of the results ranging from 1956 till 1996, e.g. graphs G on n vertices with $\mathrm{Rect}(G)$ being $\Omega(n^{3/2})$ [14], $\Omega(n^{5/3})$ [12,15,18] and $\Omega(n^{2-\epsilon})$ [8]. Also, it is known that $\mathrm{Rect}(G) \leq \frac{n^2}{\log n}$ [11].

In this paper we are interested in the largest possible gap between Cov and Rect, thus we seek graph classes having a *small* Rect and a *large* Cov value.

For Cov, a related notion is that of Steiner 2-transitive-closure-spanners [1] (Steiner-2-TC-Spanners), which is a more general notion for realizing general graphs. The two notions coincide when we look for spanners of bipartite graphs, viewed as 2-level layered directed graphs. The authors of [1] show a lower bound

for the minimal Steiner-2-TC-Spanner a bipartite graph can have. Applying these results to our problem, we get that there exist graphs with $\mathrm{Rect}(G) = O(n)$ and $\mathrm{Cov}(G) = \Omega(n\mathrm{polylog}(n))$ which is exactly the type of result we seek to achieve. We use the asymptotic behaviour operators O, Ω and Θ as well as their "up to a polylogarithmic factor" variants $\tilde{O}$, $\tilde{\Theta}$, e.g. $f(n) = \tilde{O}(g(n))$ is a shorthand for "$f = O(g(n)\log^k g(n))$ for some constant $k \geq 0$".

Automata A nondeterministic finite automaton, or NFA for short, is a tuple $M = (Q, \Sigma, \delta, q_0, F)$ with Q being an alphabet of states, Σ being the input alphabet, $\delta \subseteq Q \times \Sigma_\varepsilon \times Q$ a transition relation where Σ_ε denotes the set $\Sigma \cup \{\varepsilon\}$, $q_0 \in Q$ being the start state and $F \subseteq Q$ being the set of accepting states. The automaton is ε-free if there is no transition of the form $(p, \varepsilon, q) \in \delta$.

A *run* of the above M is a sequence $(p_1, a_1, r_1) \dots (p_t, a_t, r_t) \in \delta^*$ such that for each $1 \leq i < t$, $r_i = p_{i+1}$, and $p_1 = q_0$. The run is accepting if $r_t \in F$. The label of the run is the Σ-word $a_0 a_1 \dots a_t$. The *language* recognized by M is $L(M) = \{w \in \Sigma^* : \text{there is an accepting run of } M \text{ with label } w\}$.

The *size* of an NFA M is the cardinality $|M|$ of its set δ of transitions. It is well-known that for each NFA M there exists an equivalent ε-free automaton M' with $|M'| = \mathcal{O}(|M|^2)$, i.e. ε-elimination can be achieved via a quadratic blow-up. However, no explicit lower bounds are stated in the literature.

3 Lower bounds for the blow-up

It is clear that $\mathrm{Rect}_{k+1}(G) \leq \mathrm{Rect}_k(G)$ for each $k \geq 0$, and that there exists some $k \geq 0$ with $\mathrm{Rect}_k(G) = \mathrm{Rect}(G)$ and $\mathrm{Rect}_k(G) = \mathrm{Rect}_{k'}(G)$ for every $k' > k$. Moreover, $\mathrm{Rect}_2(G) \leq \mathrm{Cov}(G) \leq 2 \cdot \mathrm{Rect}_2(G)$: for any collection $\mathcal{C}$ of bicliques one can construct a rectifier network $R = (A \uplus \mathcal{C} \uplus B, E')$ with $(a, K_{A',B'})$ and $(K_{A',B'}, b)$ being an edge iff $a \in A'$ and $b \in B'$, respectively, showing $\mathrm{Rect}_2(G) \leq \mathrm{Cov}(G)$. For $\mathrm{Cov}(G) \leq 2 \cdot \mathrm{Rect}_2(G)$, let $R = (A \uplus X \uplus B, E')$ be a depth-2 rectifier network realizing G. Then, edges of E' are directed from A to X, from X to B and also "jump edges" from A directly to B are allowed. First, subdividing each such jump edge and adding the intermediate node to X eliminates jump edges and the resulting network $R' = (A \uplus X' \uplus B, E'')$ still realizes G in depth 2 and due to the subdividing, $|E''| \leq 2 \cdot |E'|$. For a node $x \in X'$, let $A(x)$ be the set of its ancestors (in A) and $B(x)$ be the set of its descendants (in B). Note that if R is minimal, then neither of these sets is empty. Then in G, each member of $B(x)$ is reachable from any member of $A(x)$, hence $K_{A(x),B(x)}$ is a biclique of G and the collection $\mathcal{C} = \{K_{A(x),B(x)} : x \in X'\}$ is a biclique cover of G of size $|E''| \leq 2 \cdot |E'| = 2 \cdot \mathrm{Rect}_2(G)$. Observe that the factor of 2 is tight e.g. in the case of complete matchings.

Since adding or removing isolated nodes to G does not affect either $\mathrm{Cov}(G)$ or $\mathrm{Rect}(G)$, from now on we assume that G has no isolated vertices.

It is also clear that

$$n \leq \mathrm{Rect}(G) \leq \mathrm{Cov}(G) \leq 2|E(G)|$$

where n stands for the number of vertices[5] of G: in any rectifier network the outdegree of each node $a \in A$ is at least one, and the collection $\{K_{\{a\},\{b\}} : (a,b) \in E\}$ of bicliques is a covering of weight $2|E(G)|$. Hence, $\mathrm{Cov}(G) = O(\mathrm{Rect}^2(G))$. However, it is not known whether the quadratic gap is attainable: in the rest of the article we seek an $\alpha > 1$, being as high as possible, such that there exist graphs with arbitrary large $\mathrm{Rect}(G)$ and with $\mathrm{Cov}(G) = \Omega(\mathrm{Rect}^\alpha(G))$.

To this end, we have to construct graph families having *small* Rect and *large* Cov. To show Rect is small (usually it will be $\tilde{O}(n)$ in our candidates) it suffices to give a small realizing network. On the other side, to see that Cov is large, we should have good lower bound methods.

For providing lower bounds, we define the following parameter $\kappa(G)$ of a bipartite graph G: let

$$\kappa(G) \quad = \quad \max\{\mathrm{H}(|A'|,|B'|) : K_{A',B'} \text{ is a biclique of } G\}$$

Observe that by monotonicity, it suffices to take *maximal* bicliques of G into account.

This graph parameter provides lower bounds not only for $\mathrm{Cov}(G)$ but for $\mathrm{Rect}(G)$:

Proposition 1 (See e.g. [7], Lemma 1.10. and Theorem 1.72.). *For any bipartite graph $G = (A, B, E)$, it holds that $\frac{|E|}{\kappa(G)} \leq \mathrm{Cov}(G)$ and $\frac{|E|}{\kappa(G)^2} \leq \mathrm{Rect}(G)$.*

By a similar argument, we can obtain the following inequality as well:

Proposition 2. *For any bipartite graph G, it holds that $\mathrm{Cov}(G) \leq \mathrm{Rect}(G) \cdot 2\kappa(G)$.*

Proof. Claim 1.73. in [7] states the following. Let k be the maximum integer with $K_{k,k}$ being a biclique of G. For any rectifier network $R = (V, E')$ realizing G, call an edge $(u,v) \in E'$ *eligible* iff $|A(u)| \leq k$ and $|B(v)| \leq k$. Then for any edge $(a,b) \in E$ there is a path from a to b in R containing an eligible edge.

In that case $\{K_{A(u),B(v)} : (u,v) \in E' \text{ is eligible}\}$ is a covering of G, consisting of at most $|E'| = \mathrm{Rect}(G)$ bicliques. Each biclique has weight at most $2k$ which in turn is at most $2\kappa(G)$ since $H(a,b) \leq \min\{a,b\}$ holds for any $a, b > 0$. □

It is also worth observing that $k = \Theta(\kappa)$ since $\min\{a,b\} \leq 2H(a,b)$.

Our first result considers the bipartite graph corresponding to the mod 2 inner product function.

Theorem 1. *Let $d > 0$ be an even integer and $G^d_\perp = (A, B, E)$ be the bipartite graph with $A = B = \{0,1\}^d$ and $(\boldsymbol{u}, \boldsymbol{v}) \in E$ for the vectors $\boldsymbol{u}, \boldsymbol{v} \in \{0,1\}^d$ iff $\boldsymbol{u} \perp \boldsymbol{v}$ in $\mathbb{Z}_2^d$, i.e. iff $\sum_{i \in [d]} u_i v_i = 0$ where sum is taken modulo 2.*

Then $\mathrm{Rect}(G^d_\perp) = \tilde{O}(n)$ and $\mathrm{Cov}(G^d_\perp) = \Omega(n^{3/2})$ where $n = 2^d$ is the number of vertices of $G^d_\perp$.

[5] At times n will denote the size of *one* of the two classes of G, introducting a factor of 2 but never causing differences in the growth order.

The proof is broken into two parts. The lower bound follows from the first inequality of Proposition 1 and a special case of Lindsey's lemma [6].

Proposition 3. $\kappa(G_\perp^d) = \frac{\sqrt{n}}{2}$. *Thus* $\mathrm{Cov}(G_\perp^d) = \Omega(n^{3/2})$.

At the same time, $\mathrm{Rect}(G_\perp^d)$ is small enough. To see this, we show $\mathrm{Rect}(G) = \tilde{O}(n)$ for a specific family of bipartite graphs, which we call *permutation invariant* graphs. A bipartite graph $G = (\{0,1\}^d, \{0,1\}^d, E)$ is permutation invariant if $(\boldsymbol{u}, \boldsymbol{v}) \in E$ implies $(\pi(\boldsymbol{u}), \pi(\boldsymbol{v})) \in E$ for any permutation $\pi : [d] \to [d]$ of the coordinate index set. Here $\pi(u_1, \ldots, u_d)$ is defined to be $(u_{\pi(1)}, \ldots, u_{\pi(d)})$. It is clear that the graphs $G_\perp^d$ are permutation invariant.

For such graphs the following holds (which also state that within this class of graphs, the bound $\mathrm{Cov}(G) = \Omega(\mathrm{Rect}(G)^{3/2})$ is optimal):

Theorem 2. *For permutation invariant graphs* Rect *is* $\tilde{O}(n)$ *and* Cov *is* $\tilde{O}(n^{3/2})$.

Proof. Suppose $G = (A, B, E)$ is permutation invariant with $A = B = \{0,1\}^d$. Let $c : \{0,1\}^d \times \{0,1\}^d \to \{0, \ldots, d\}^{\{0,1\}\times\{0,1\}}$ be the function defined as

$$c((u_1, \ldots, u_d), (v_1, \ldots, v_d))(a, b) \;=\; |\{i \in [d] : u_i = a, v_i = b\}|.$$

That is, $c(\boldsymbol{u}, \boldsymbol{v})(a, b)$ is the number of positions i on which $\boldsymbol{u}$ is a and $\boldsymbol{v}$ is b.

Then, G factors through c in the following sense: if $c(\boldsymbol{u}, \boldsymbol{v}) = c(\boldsymbol{u}', \boldsymbol{v}')$, then $(\boldsymbol{u}, \boldsymbol{v}) \in E$ iff $(\boldsymbol{u}', \boldsymbol{v}') \in E$. Indeed, $c(\boldsymbol{u}, \boldsymbol{v}) = c(\boldsymbol{u}', \boldsymbol{v}')$ if and only if there exists a permutation $\pi : [d] \to [d]$ such that $u_i = u'_{\pi(i)}$ and $v_i = v'_{\pi(i)}$ for each $i \in [d]$, yielding $(\boldsymbol{u}, \boldsymbol{v}) \in E$ if and only if $(\boldsymbol{u}', \boldsymbol{v}') \in E$.

Hence there exists a subset C of the finite set $\{0, \ldots, d\}^{\{0,1\}\times\{0,1\}}$ such that $(\boldsymbol{u}, \boldsymbol{v}) \in E$ iff $c(\boldsymbol{u}, \boldsymbol{v}) \in C$.

We define a rectifier network $R = (\{0,1\}^d \times \{0, \ldots, d\}^{\{0,1\}\times\{0,1\}} \times \{0, \ldots, d\})$: the pair $((u_1, \ldots, u_d), f, \ell), ((v_1, \ldots, v_d), f', \ell')$ is an edge of R iff the following conditions hold: $\ell' = \ell + 1$ (so that R is a DAG of depth $d + 1$); for each $i \neq \ell'$, $u_i = v_i$ holds; finally, $f'(u_{\ell'}, v_{\ell'}) = f(u_{\ell'}, v_{\ell'}) + 1$ and for any other $(a, b) \in \{0,1\} \times \{0,1\}$, $f'(a, b) = f(a, b)$.

Then by induction on $\ell' - \ell$ we get that there is a path from $((u_1, \ldots, u_d), f, \ell)$ to $((v_1, \ldots, v_d), f', \ell')$ iff the following conditions hold: $\ell < \ell'$; for each $i \leq \ell$ and $i > \ell'$, $u_i = v_i$; finally, $f'(a, b) = f(a, b) + |\{\ell < i \leq \ell' : u_i = a, v_i = b\}|$.

Now let $R' = (V(R) \uplus \{0,1\}^d, E(R) \uplus E')$ with E' consisting of the edges of the form $(\boldsymbol{v}, f, d) \to \boldsymbol{v}$ with $f \in C$. Then R' realizes G by identifying each $\boldsymbol{u} \in A$ with $(\boldsymbol{u}, \boldsymbol{0}, 0)$ and each $\boldsymbol{v} \in B$ with the element $\boldsymbol{v}$ of this last layer of R' (here $\boldsymbol{0}$ stands for the constant zero function $\boldsymbol{0} : (a, b) \mapsto 0$). Since in R', there are at most $2^d(\cdot d \cdot \{d+1\}^4) \cdot 2$ edges (each node not belonging to layer d has outdegree 2 in R and in the last step, $2^d \times |C| \leq 2^d \cdot \{d+1\}^4$ edges are added), which is $O(n \log^5 n)$, showing $\mathrm{Rect}(G) = \tilde{O}(n)$.

For $\mathrm{Cov}(G) = \tilde{O}(n^{3/2})$, let X be the set of vertices of R' of the form $(\boldsymbol{u}, f, d/2)$. (That is, nodes of the middle layer of R'.) As before, let $A(x) \subseteq A$, $x \in X$ stand for the set of nodes from which x is reachable in R' and let $B(x) \subseteq B$ stand for the set of those nodes which are reachable from x in R'. Then, since

each node of R has indegree at most 2, we have that $|A(x)| \leq 2^{d/2}$. For $|B(x)|$, since each outdegree in R is 2, we get that there are at most $2^{d/2}$ nodes of the form $(\boldsymbol{v}, f, d)$ reachable from x. In E', the outdegree of these nodes is $|C|$ which is at most $(d+1)^4$, hence $|B(x)| \leq 2^{d/2}(d+1)^4 = \tilde{O}(\sqrt{n})$. Thus, the covering $\mathcal{C} = \{K_{A(x),B(x)} : x \in X\}$ has size $\sum_{x \in X}(|A(x)| + |B(x)|)$ which is at most $2^d \cdot (d+1)^4 \cdot (2^{d/2} + 2^{d/2}(d+1)^4) = \tilde{O}(n^{3/2})$. Note that due to the layered structure of R', each $\boldsymbol{u} \to \boldsymbol{v}$ path contains a node belonging to X, so $\mathcal{C}$ is indeed a covering. □

Thus we have showed that for an arbitrarily $\epsilon > 0$ there are graphs $G = G_{\perp}^d$ having arbitrarily large $\mathrm{Rect}(G) = \tilde{O}(n)$ and with $\mathrm{Cov}(G) = \Omega(\mathrm{Rect}^{3/2-\epsilon}(G))$. (Observe that any permutation invariant graph with $\kappa = \Theta(\sqrt{n})$ and $\Theta(n^2)$ edges meets this condition.)

As an interesting corollary, we get that $\mathrm{Cov}(G_{\perp}^d)$ is $\tilde{\Theta}(n^{3/2})$ which is $\tilde{\Theta}(\frac{|E|}{\kappa})$ so in this case the bound of Proposition 1 is optimal up to a log factor.

A general construction for constructing a biclique covering of a graph $G = (A, B, E)$ is the following: starting from a rectifier network $R = (V, E')$ first one chooses a *cut* $E_0 \subseteq E'$ of the edges of R (so that each $a \to b$, $a \in A$, $b \in B$ path contains an edge from E_0), in which case a covering is $\mathcal{C}(E_0) = \{K_{A(x),B(y)} : (x, y) \in E_0\}$. We call coverings of this form *cut-coverings of* R. (In the proof of Theorem 2 we employ a similar construction, choosing a subset X of vertices instead of a subset E_0.) In the following we state without proof that this construction is not optimal, not even up to a polylogarithmic factor, even when R is optimal up to a polylogarithmic factor.

Theorem 3. *Consider the graph $G_{\Delta}^n = (A, B, E)$ with $A = B = [n]$ and $(i, j) \in E$ iff $d(i, j) \leq \frac{n}{4}$ where $d(i, j)$ is the modulo n distance* $\min\{|i-j|, |n+i-j|\}$. *(That is, distance on the circle graph C_n.) Then:*

1. *There exists a rectifier network R_n realizing G_{Δ}^n with $\tilde{O}(n)$ edges.*
2. *Any cut-covering of R_n has size $\Omega(n^2)$.*
3. *At the same time, $\mathrm{Cov}(G_{\Delta}^n)$ is $O(n^{1+\epsilon})$ for any $\epsilon > 0$ where the O notation hides a constant depending only on ϵ.*

Note that for this graph we have $\kappa = \Theta(n)$ since $K_{[n/4],[n/4]}$ is a biclique. Hence also for this class of graphs, $\frac{|E|}{\kappa} = \Theta(n)$ approximates $\mathrm{Cov} = O(n^{1+\epsilon})$ relatively well. In the next section we show that a closely related formula gives an upper bound for $\mathrm{Cov}(G)$.

4 Upper bounds for the blow-up

In this section we will show that, under certain assumptions, $\mathrm{Cov}(G) = o(\mathrm{Rect}(G)^2)$ or even $\mathrm{Cov}(G) = O(\mathrm{Rect}(G)^{3/2})$ holds. Proposition 2 implies the following result:

Theorem 4. *For any bipartite graph G and $0 < \alpha \leq 1$ with $\mathrm{Cov}(G) \leq \frac{|E|}{\kappa^{\alpha}}$ we have $\mathrm{Cov}(G) \leq 2\mathrm{Rect}(G)^{\beta}$ for some $\beta \leq 1 + \frac{1}{1+\alpha} \in [3/2, 2)$. Hence if $\mathrm{Cov}(G) \leq \frac{|E|}{\kappa^{\alpha}}$ holds for a family of graphs G, then $\mathrm{Cov}(G) = O(\mathrm{Rect}(G)^{2-\frac{\alpha}{1+\alpha}})$.*

Proof. Let us introduce the following notation: $|E| = n^\delta$ for $n = |V(G)|$, $\mathrm{Rect}(G) = n^r$ and $\mathrm{Cov}(G) = \frac{|E|}{\kappa^\alpha}$, $0 < \alpha \leq 1$. We will show that choosing $\beta = \frac{\delta+\alpha\cdot r}{r(1+\alpha)}$ suffices. (Note that since $\delta \leq 2$ and $r \geq 1$, β is indeed at most $1 + \frac{1}{1+\alpha}$.)

By $\mathrm{Cov} = \frac{|E|}{\kappa^\alpha}$ we have $\log_n \mathrm{Cov} = \delta - k\alpha$ where $k = \log_n \kappa$. Now assuming for contradiction that $2^{\frac{\alpha}{1+\alpha}}\mathrm{Rect}^\beta < 2\mathrm{Rect}^\beta < \mathrm{Cov}$ we get

$$r\beta = \frac{\delta + \alpha \cdot r}{1+\alpha} < \delta - k \cdot \alpha - \frac{\alpha}{1+\alpha}\log_n 2.$$

Then direct computation shows that $r < \delta - k(1+\alpha) - \log_n 2$ which is a contradiction, since by $\mathrm{Rect} \geq \frac{\mathrm{Cov}}{2\kappa}$ we have $r \geq \delta - k(1+\alpha) - \log_n 2$. □

Simple examples show that the assumption of the theorem does not hold for all graphs. A similar argument gives a similar, but somewhat weaker, bound $\mathrm{Cov}(G) = O(\mathrm{Rect}(G)^{2-\varepsilon})$ for some $\varepsilon > 0$ if the condition $\mathrm{Cov}(G) \leq \frac{|E|}{\kappa^\alpha}$ is replaced by $\mathrm{Cov}(G) \leq \mathrm{polylog}n \max \frac{|E(G')|}{\kappa(G')}$, where the maximum ranges over induced subgraphs G' of G. Thus an affirmative answer to the following open problem would imply $\mathrm{Cov}(G) = O(\mathrm{Rect}(G)^{2-\varepsilon})$ for all bipartite graphs.

Problem 1 *Is it true that for any bipartite graph G on n vertices,*

$$\mathrm{Cov}(G) \leq \mathrm{polylog}n \max \frac{|E(G')|}{\kappa(G')}$$

where the maximum ranges over induced subgraphs G' of G?

4.1 The set cover problem

Now we apply the weighted set cover problem to our setting. For a detailed discussion of this problem, and an introduction to approximation methods see [20].

The *weighted set cover* problem is the following: we are given a collection $\mathcal{S} = \{S_1, \ldots, S_t\}$ of subsets of some finite universe A of n elements with $\cup\mathcal{S} = A$, and to each S_i, a cost $c(S_i) > 0$ is associated. The goal is to find a subset $\mathcal{C}$ of $\mathcal{S}$ such that $\cup\mathcal{C} = A$ and the total cost $\sum_{S\in\mathcal{C}} c(S)$ is minimized. The problem is well-known to be NP-complete already for the uniform setting when $c(S_i) = 1$; however, the following greedy algorithm returns a fair enough approximation:

Let $U := A$ and $\mathcal{C} := \emptyset$.
while $U \neq \emptyset$ **do**
  Choose $S \in \mathcal{S}$ such that $\frac{c(S)}{|S\cap U|}$ is the minimum possible value.
  Let $U := U - S$ and $\mathcal{C} := \mathcal{C} \cup \{S\}$.
return $\mathcal{C}$.

The following linear program is the standard relaxation of weighted set cover:

$$\text{minimize } \sum_{i=1}^{t} c(S_i)x_i \qquad \text{subject to } \sum_{i:a\in S_i} x_i \geq 1 \quad \forall a \in A, \quad x_i \geq 0$$

Denote by OPT the optimal solution of the weighted set cover problem. It is well-known [3,5,10,17] that the value of the solution returned by the above algorithm is bounded by $\ln n \cdot \mathrm{OPT}$, where $n = |A|$, and even by $\ln n \cdot Z^*_{LP}$, where Z^*_{LP} denotes the value of an optimal solution to the LP relaxation.

Now we define a related combinatorial quantity. For a subset B of A, let $\eta(B)$ stand for the value $\min_{S \in \mathcal{S}} \frac{c(S)}{|S \cap B|}$ which is present inside the loop of the greedy algorithm. Note that this value is positive and finite for any $B \subseteq A$. Also, let η^* stand for $\max_{B \subseteq A} |B| \cdot \eta(B)$. Then we have:

Proposition 4. $\eta^* \leq Z^*_{LP} \leq \mathrm{OPT}$.

Proof. Consider any feasible solution $\boldsymbol{x}$ and a subset B of A. Then,

$$\begin{aligned}\sum_{i \in [t]} c(S_i) x_i &\geq \sum_{i : S_i \cap B \neq \emptyset} c(S_i) x_i = \sum_{i : S_i \cap B \neq \emptyset} \sum_{a \in S_i \cap B} \tfrac{c(S_i)}{|S_i \cap B|} x_i \\ &\geq \sum_{i : S_i \cap B \neq \emptyset} \sum_{a \in S_i \cap B} \eta(B) x_i = \eta(B) \sum_{a \in B} \sum_{S_i \ni a} x_i \\ &\geq \eta(B) \sum_{a \in B} 1 = \eta(B) \cdot |B|.\end{aligned}$$

□

On the other hand, this quantity can be used to give an upper bound for OPT as well. The following bound is proven in [9] for the unweighted case.

Proposition 5. *Let* Greedy *stand for the cost of the solution returned by the greedy algorithm. Then* $\textsc{Greedy} \leq H_n \eta^*$, *where* H_n *is* $\sum_{i \in [n]} \frac{1}{i} \approx \ln n$.

Proof. Let U_k denote the set of uncovered elements at the beginning of the kth iteration of the loop of the greedy algorithm and let $n_k = |U_k|$. Then

$$\begin{aligned}\min_{S_i} \tfrac{c(S_i)}{|S_i \cap U_k|} &= \min_{S_i} \tfrac{c(S_i)}{|S_i \cap U_k|} \tfrac{|U_k|}{n_k} \leq \max_{B \subseteq A} \min_{S_i} \tfrac{c(S_i)}{|S_i \cap B|} \tfrac{|B|}{n_k} \\ &= \max_{B \subseteq A} \tfrac{|B|}{n_k} \min_{S_i} \tfrac{c(S_i)}{|S_i \cap B|} = \max_{B \subseteq A} \tfrac{|B|}{n_k} \eta(B) = \tfrac{1}{n_k} \eta^*.\end{aligned}$$

Thus, the covering of the elements covered in the kth iteration costs at most $\frac{1}{n_k}\eta^*$ for each such element. Since $n_1 = |A| = n$ and at each iteration, n_k is strictly decreasing, the total cost is at most $\sum_{i \in [n]} \frac{1}{i} \eta^* = H_n \eta^*$. □

Thus we have the following chain of inequalities:

$$\eta^* \leq Z^*_{LP} \leq \mathrm{OPT} \leq \textsc{Greedy} \leq H_n \cdot \eta^*.$$

4.2 Application to biclique coverings

Determining $\mathrm{Cov}(G)$ for $G = (A, B, E)$ can be viewed as a set cover problem: the universe is E, the allowed sets are bicliques of G, and the cost of a biclique $K_{A',B'}$ is $|A'| + |B'|$.

In this problem, η is the following: given a subset E' of E (that is, a *subgraph* $G' = (A', B', E')$ of G), $\eta(E')$ is defined as

$$\min_{K_{A'',B''} \subseteq E} \frac{|A''|+|B''|}{|E' \cap (A'' \times B'')|}$$

and η^* is the maximal possible value of $|E'|\eta(E')$. By Proposition 5 we have that $\mathrm{Cov}(G) \leq \eta^* \cdot H_n$.

Observe that for the biclique $K_{A_0,B_0} = \arg\min_{K_{A',B'} \subseteq E} \frac{|A'|+|B'|}{|E' \cap (A' \times B')|}$ we have $A_0 \subseteq A'$ and $B_0 \subseteq B'$. Indeed, otherwise $K_{A_0 \cap A', B_0 \cap B'}$ would be a better biclique. Thus, the minimizer biclique K_{A_0,B_0} is a biclique of the subgraph of G *induced* by $A_0 \cup B_0$. It is also clear that for induced subgraphs G' we have $\frac{1}{\kappa(G')} = \eta(E')$ and $|E'|\eta(E') = \frac{|E(G')|}{\kappa(G')}$.

Hence, there are two cases: either η^* takes its value on some *induced* subgraph of G up to a polylogarithmic factor, in which case the bound $\mathrm{Cov}(G) \geq \max \frac{|E(G')|}{\kappa(G')}$ is essentially optimal up to a polylog factor, or not, in which case there are graphs having much larger η^* than $\frac{|E|}{\kappa}$. In the first case, the remarks following Theorem 4 imply a subquadratic upper bound for the blow-up.

Problem 2 *Determine the gap possible between η^* and $\max \frac{|E(G')|}{\kappa(G')}$, where the maximum is taken over all induced subgraphs.*

5 Application: the cost of ε-removal

Let A and B be disjoint alphabets (nonempty finite sets) and $L \subseteq AB$ a (finite) language consisting of two-letter words. Then L can also be viewed as a bipartite graph $G_L = (A, B, L)$ where the notation for L is slightly abused (i.e. (a, b) is an edge iff the word ab belongs to the language). Without loss of generality we may assume that for each $a \in A$ ($b \in B$, resp.) there exists a $b \in B$ ($a \in A$, resp.) such that ab is in L.

Proposition 6. *There is some NFA M recognizing L with $|M| = O(\mathrm{Rect}(G_L))$.*

Proof. Let $R = (V, E)$ be a rectifier network for G_L with $|E| = \mathrm{Rect}(G_L)$. Then the automaton $M = (V \uplus \{q_0, q_f\}, A \cup B, \delta, q_0, \{q_f\})$ with

$$\delta = \{(q_0, a, a) : a \in A\} \cup \{(b, b, q_f) : b \in B\} \cup \{(p, \varepsilon, q) : p \to q \in E\}$$

recognizes L with $|M| = |E| + |A| + |B| = O(\mathrm{Rect}(G_L))$. □

Proposition 7. *For any ε-free NFA M recognizing L, $\mathrm{Cov}(G_L) \leq |M|$. Moreover, there exists a ε-free NFA M recognizing L with $|M| = \mathrm{Cov}(G_L)$.*

Proof. Let $M = (Q, A \cup B, \delta, q_0, F)$ be an ε-free NFA recognizing L of minimal size. Since L is prefix-free and M is minimal, $F = \{q_f\}$ is a singleton set. Also, M is *trim*, i.e. for each state $p \in Q$ there exist words x, y with $p \in q_0 x$ and

$q_f \in py$. Since every word in L has the same length 2, to each state p there is an integer $0 \leq n_p \leq 2$ such that whenever $p \in q_0 x$ for some word x, then $|x| = n_p$. Otherwise if $p \in q_0 x_1$ and $p \in q_0 x_2$ for words x_1, x_2 of different length, then $x_1 y$ and $x_2 y$ are members of L of different length for any word y with $q_f \in py$, a contradiction. Also, it is clear that $n_p = 0$ only for $p = q_0$ and $n_p = 2$ only for $p = q_f$. Thus if X stands for $Q - \{q_0, q_f\}$, we get that M is a layered automaton with transitions of the form (q_0, a, p) for $a \in A$ and $p \in X$ and (p, b, q_f) for $b \in B$ and $p \in X$. Hence, letting $A(p)$ to stand for the set $\{a \in A : (q_0, a, p) \in \delta\}$ and $B(p)$ stand for the set $\{b \in B : (p, b, q_f) \in \delta\}$ we get that there is an associated biclique covering $\mathcal{C}_M$ of L to M consisting of the bicliques $K_{A(p),B(p)}$, $p \in X$ that has the same size as M.

Observe that the transformation is invertible in the sense that to each biclique covering $\mathcal{C}$ such an automaton of the same size can be constructed, showing the second part of the claim. □

Since by Theorem 1 there exist graphs with arbitrary large $\mathrm{Rect}(G)$ and $\mathrm{Cov}(G) = \Omega(\mathrm{Rect}(G)^{3/2-\epsilon})$, we have the following as byproduct:

Theorem 5. *For any $\epsilon > 0$ and for arbitrarily large n there exist languages (consisting of two-letter words only) which are recognizable by NFAs of size n but are only recognizable by ε-free NFAs of size $\Omega(n^{3/2-\epsilon})$.*

In other words, in order to make an NFA ε-free, an $n^{3/2-\epsilon}$ blow-up in the size can be inevitable.

6 Conclusion, future directions

We proved a lower bound for the blow-up when transforming NFAs to ε-free NFAs. We showed that the cost of ε-removal from NFAs is worst-case $\Omega(n^{3/2-\epsilon})$, improving the previous bound $\Omega(n \log^2 n)$. The largest possible gap is between $\Omega(n^{3/2-\epsilon})$ and $O(n^2)$, just like in the case of going from CFGs to chain rule free CFGs. Narrowing these gaps seem to be nontrivial open problems.

We used a graph-theoretic approach by translating the problem into finding large blow-ups between two complexity measures for bipartite graphs: the rectifier network size Rect and the minimal weight biclique covering Cov. We proved that there are graphs with arbitrarily large Rect value n such that $\mathrm{Cov} = \Omega(n^{3/2-\epsilon})$ for any $\epsilon > 0$. We gave partial results for determining the largest possible blow-up between these quantities. These include a sufficient condition for a subquadratic upper bound, and the sharpness of a combinatorial bound for the minimal weight biclique covering (obtained by proving a bound for the general weighted set covering problem). We also formulated two open problems about related combinatorial bounds, which appear to be of interest in themselves. Solving these problems may also be useful for determining the largest possible blow-up. The relationship between Rect and Cov can be viewed as a size-depth trade-off problem for depth-2 and unrestricted depth circuits computing sets of Boolean disjunctions [19]. As far as we know, there are many other related open problems, such as establishing a bounded-depth hierarchy.

References

1. Piotr Berman, Arnab Bhattacharyya, Elena Grigorescu, Sofya Raskhodnikova, David P. Woodruff, and Grigory Yaroslavtsev. Steiner transitive-closure spanners of low-dimensional posets. *Combinatorica*, pages 1–24, 2014.
2. Norbert Blum. More on the power of chain rules in context-free grammars. *Theoretical Computer Science*, 27(3):287 – 295, 1983. Special Issue Ninth International Colloquium on Automata, Languages and Programming (ICALP) Aarhus, Summer 1982.
3. V. Chvatal. A greedy heuristic for the set-covering problem. *Mathematics of Operations Research*, 4(3):233–235, 1979.
4. Juraj Hromkovic, Sebastian Seibert, and Thomas Wilke. Translating regular expressions into small epsilon-free nondeterministic finite automata. In *Proceedings of the 14th Annual Symposium on Theoretical Aspects of Computer Science*, STACS '97, pages 55–66, London, UK, 1997. Springer-Verlag.
5. David S. Johnson. Approximation algorithms for combinatorial problems. In *Proceedings of the Fifth Annual ACM Symposium on Theory of Computing*, STOC '73, pages 38–49, New York, NY, USA, 1973. ACM.
6. Stasys Jukna. *Boolean Function Complexity - Advances and Frontiers*, volume 27 of *Algorithms and combinatorics*. Springer, 2012.
7. Stasys Jukna. Computational complexity of graphs. In M. Dehmer and F. Emmert-Streib, editors, *Advances in Network Complexity*, pages 99–153. Wiley, 2013.
8. János Kollár, Lajos Rónyai, and Tibor Szabó. Norm-graphs and bipartite Turán numbers. *Combinatorica*, 16(3):399–406, 1996.
9. László Lovász. A kombinatorika minimax tételeiről. *Matematikai Lapok*, 26:209–264, 1975.
10. László Lovász. On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13(4):383 – 390, 1975.
11. O. B. Lupanov. On rectifier and switching-and-rectifier schemes. *Dokl. Akad. Nauk SSSR*, 111:1171–1174, 1956.
12. Kurt Mehlhorn. Some remarks on Boolean sums. In Jirí Becvár, editor, *Mathematical Foundations of Computer Science 1979*, volume 74 of *Lecture Notes in Computer Science*, pages 375–380. Springer Berlin Heidelberg, 1979.
13. Albert R. Meyer and Michael J. Fischer. Economy of description by automata, grammars, and formal systems. In *SWAT (FOCS)*, pages 188–191. IEEE Computer Society, 1971.
14. E. I. Nechiporuk. On a Boolean matrix. *Systems Theory Res.*, 21:236–239, 1971.
15. Nicholas Pippenger and Leslie G. Valiant. Shifting graphs and their applications. *J. ACM*, 23(3):423–432, July 1976.
16. Georg Schnitger. Regular expressions and NFAs without ε-transitions. In *23th Symposium on Theoretical Aspects of Computer Science (STACS 2006), LNCS 3884*, pages 432–443, 2006.
17. S.K Stein. Two combinatorial covering theorems. *Journal of Combinatorial Theory, Series A*, 16(3):391 – 397, 1974.
18. Ingo Wegener. A new lower bound on the monotone network complexity of Boolean sums. *Acta Informatica*, 13(2):109–114, 1980.
19. Ingo Wegener. *The Complexity of Boolean Functions*. John Wiley & Sons, Inc., New York, NY, USA, 1987.
20. David P. Williamson and David B. Shmoys. *The Design of Approximation Algorithms*. Cambridge University Press, New York, NY, USA, 1st edition, 2011.