The Homogeneous Set Sandwich Problem

Vinícius Gusmão Pereira de Sá[1*]

[1]COPPE, Universidade Federal do Rio de Janeiro

vgusmao@cos.ufrj.br

Abstract. *Homogeneous sets have proved useful for the construction of graph decomposition procedures. Sandwich graphs are obtained from two pre-defined graphs which provide them with both mandatory and optional edges. Given such a pair of graphs, the Homogeneous Set Sandwich Problem (HSSP) searches for a sandwich graph which contains a homogeneous set. This thesis presents the development of techniques for solving the HSSP and culminates in the correction of the established upper bounds for its time complexity. Such results comprise the description of thoroughly depicted counterexamples which prove the incorrectness of the (so far) best published HSSP algorithm, along with the proposal of a new, faster algorithm.*

1. Introduction

A homogeneous set is a non-trivial module of a graph, i.e. a non-unitary, proper subset H of a graph's vertices such that all vertices in H have the same neighborhood outside H. Given two graphs $G_1(V, E_1), G_2(V, E_2)$, the Homogeneous Set Sandwich Problem (HSSP) asks whether there exists a sandwich graph $G_S(V, E_S)$, with $E_1 \subseteq E_S \subseteq E_2$, which has a homogeneous set.

Sandwich-graph problems were first defined in the context of Computational Biology as a relaxation of recognition problems. Their applications abound in the literature [Golumbic et al., 1995, Golumbic, 1998, Golumbic and Wassermann, 1998, Kaplan and Shamir, 1999, de Figueiredo et al., 2002, Dantas et al., 2002], whereas the importance of homogeneous sets in the context of graph decomposition has been well acknowledged, specially in the perfect graphs field [Lovász, 1972].

Notwithstanding the existence of linear-time algorithms for solving the recognition problem of finding homogeneous sets in a single graph, the HSSP cannot benefit from them in any straightforward manner and still remains a subject for research.

The first polynomial-time algorithm for this problem was presented in [Cerioli et al., 1998], setting its upper bound at $O(n^4)$. We refer to this algorithm as the *Exhaustive Bias Envelopment Algorithm* (*EBE algorithm*, for short). A few years later, [Tang et al., 2001] introduced a brand new algorithm, based on a quite beautiful idea, which would have largely decreased HSSP's upper bound. In this thesis, we show that this latter algorithm, which we refer to as the *Bias Graph Components*

*Supported by CAPES, M. Sc. Dissertation presented at COPPE/UFRJ under the supervision of Celina M. H. de Figueiredo.

Algorithm (*BGC algorithm*, for short), is unfortunately not correct. As a consequence, the most efficient algorithm that correctly solves the HSSP would turn back to be the former EBE algorithm presented in [Cerioli et al., 1998]. This thesis brings, however, a faster algorithm, which establishes a new upper bound to the problem at $O(m_1 n^2)$.

In Section 2, we summarize the EBE algorithm. In Section 3, we give a brief description of the BGC algorithm and point out where its basic flaw lies. Finally, Section 4 introduces the new algorithm.

Throughout this paper, we denote the number of vertices in the input graphs by n and the number of edges in graph G_i by m_i. Also, $\triangle_i$ stands for G_i's maximum vertex degree.

2. The Exhaustive Bias Envelopment algorithm

Let $G_S(V, E_S)$ be a sandwich graph of graphs $G_1(V, E_1)$, $G_2(V, E_2)$. The edges in E_1 are called *mandatory edges*, once each and every sandwich graph of (G_1, G_2) has to contain them. On the other hand, the edges *not* in E_2 are said to be *forbidden edges*, meaning that no sandwich graph of (G_1, G_2) is allowed to contain them. A vertex $b \in V$ is called a *bias vertex* of a vertex set $S \subseteq V \setminus \{b\}$ if there exists at least one mandatory edge $(b, v) \in E_1$ and at least one forbidden edge $(b, w) \notin E_2$, for some $v, w \in S$. The set $B(S)$ contains all bias vertices of S and is thereby called the *bias set* of S.

Theorem 1. *[Cerioli et al., 1998] The set $S \subseteq V$ is a sandwich homogeneous set of a pair (G_1, G_2) if and only if its bias set $B(S)$ is the empty set.*

The EBE algorithm starts by choosing a sandwich homogeneous set *candidate* $\{x, y\}$. Then it successively determines the candidate's bias vertices and adds all of them to the current candidate. We refer to this procedure as *Bias Envelopment*. The Bias Envelopment continues until either a candidate with an empty bias set has been found, whereby the algorithm stops with an *yes* answer, or else the candidate set has become equal to the input vertex set V, in which case the algorithm restarts the process with another initial pair of vertices. If no sandwich homogeneous set has been found by the time all possible pairs have been investigated, the algorithm answers *no*. The worst-case time complexity of this algorithm is $O(n^4)$.

3. The Bias Graph Components algorithm

The main idea of the BGC algorithm, presented in [Tang et al., 2001], is to use the bias relation introduced in Section 2 to construct a directed graph, called *bias graph*. The bias graph $G_B(V_B, E_B)$ of a pair of graphs $G_1(V, E_1), G_2(V, E_2)$ has vertex set $V_B = \{[x, y] \mid x, y \in V, x \neq y\}$ and there are two outgoing edges from vertex $[u, v]$ to vertices $[u, w]$ and $[v, w]$ in G_B if and only if vertex w is a bias vertex of vertex set $\{u, v\}$ with respect to the pair (G_1, G_2).

Once the bias graph has been constructed, the algorithm proceeds to finding all its strongly connected components (SCC). Then it looks for an *end* strongly connected component (ESCC) among them, i.e. a SCC with no outgoing edges. If only one ESCC

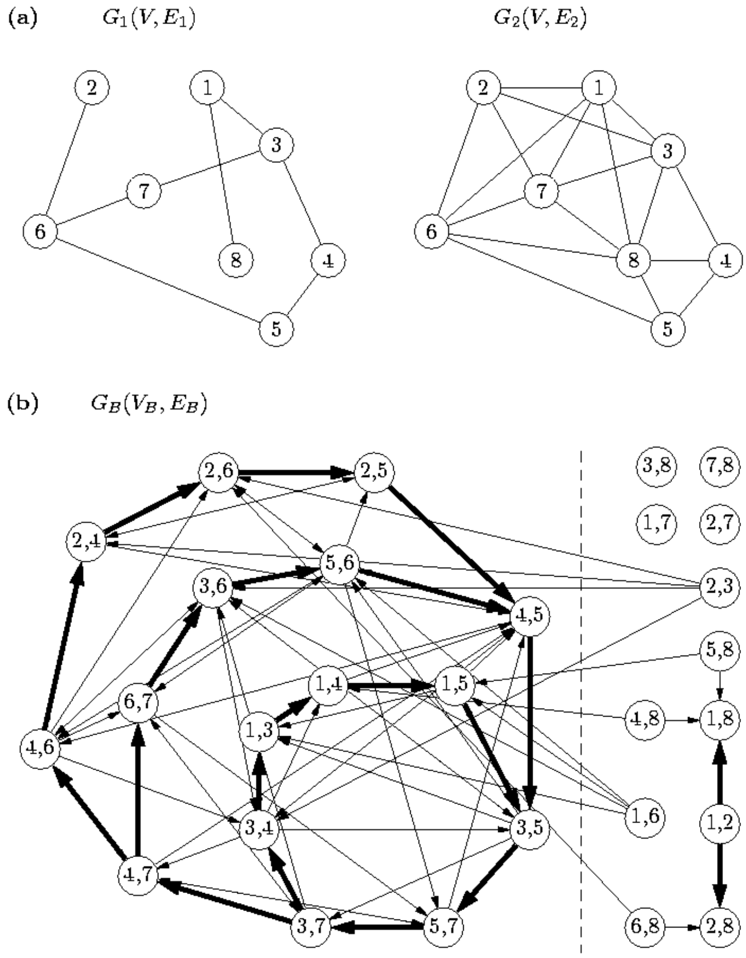

Figure 1: Counterexample 1 (to Claim 2 [Tang et al., 2001])

is found and it embraces all input vertices (as part of its vertices' labels), the algorithm returns *no*. Otherwise, the algorithm translates one of the bias graph's ESCCs, say component C, into the set $H \subset V$ of input vertices that are used to label C's vertices. In this case it returns *yes* and H, for H would allegedly be a sandwich homogeneous set. The time complexity of the BGC algorithm is shown to be $O(\triangle_1 n^2)$.

Claim 2. [Tang et al., 2001] *The BGC algorithm correctly solves the HSSP.*

To begin with, Figure 1(a) shows a simple refutation. It presents a pair of graphs (G_1, G_2) that produce the bias graph $G_B(V_B, E_B)$ in Figure 1(b). The vertex set $S \subset V_B$ on the left of the dashed line constitutes an ESCC. The set $H = \{1, 2, \ldots, 7\} \subset V$ that labels the vertices in S, however, is *not* a sandwich homogeneous set of (G_1, G_2).

By its turn, Figure 2 gives the pair $G_1(V, E_1)$, $G_2(V, E_2)$, which has sandwich homogeneous set $H = \{1, 2, \ldots, 9, 1', 2', \ldots, 9'\}$ (and no other). However, the BGC algorithm would give it a *no* answer, for its one and only sandwich homogeneous set is not associated with any ESCCs in its bias graph.

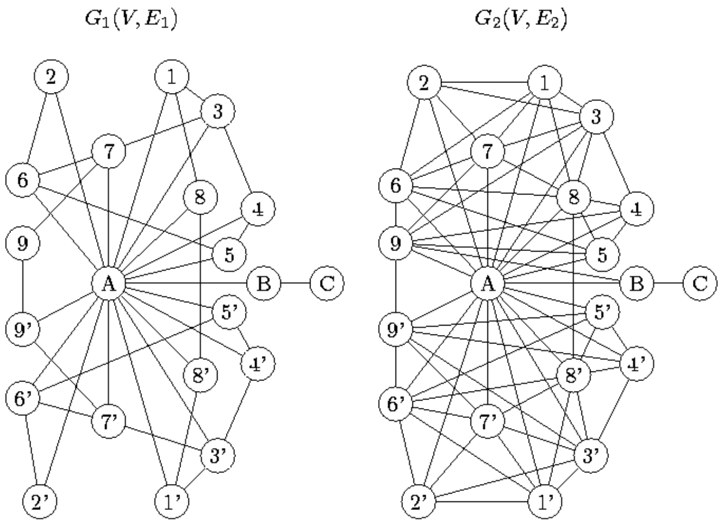

Figure 2: Counterexample 2 (to Claim 2 [Tang et al., 2001])

4. A new $O(m_1 n^2)$ upper bound: The Two-Phase algorithm

Let $G_B(V_B, E_B)$ be the bias graph of input graphs $G_1(V, E_1), G_2(V, E_2)$. A subset $K \subseteq V_B$ is said to be a *pair-closed* set if and only if there do not exist two vertices $x, y \in V$, among those which label K's vertices, such that vertex $[x, y]$ is not an element of K. The set $A = \{[1,2],[1,3],[2,3]\}$ is a pair-closed set. The set $B = \{[1,2],[1,3],[1,4],[2,3],[2,4]\}$ is *not* pair-closed, for vertices 3 and 4 appear in the label of some vertices in B but $[3,4] \notin B$.

Theorem 3. *A set $H \subset V$ is a sandwich homogeneous set of graphs $G_1(V, E_1), G_2(V, E_2)$ if and only if the pair-closed set $K = \{[x, y] \mid x, y \in H\} \subset V_B$ induce an* end *subgraph in bias graph G_B of (G_1, G_2).*

Theorem 3 does not lead directly to an efficient algorithm for the HSSP, as there is no quick means of finding pair-closed sets which induce end subgraphs. Corollary 4, however, brings about the central inspiration for the algorithm that follows.

Corollary 4. *If $H \subset V$ is a sandwich homogeneous set of graphs $G_1(V, E_1)$, $G_2(V, E_2)$, then either the subgraph $G_B\langle K\rangle$, induced by the pair-closed set $K = \{[x, y] \mid x, y \in H\} \subset V_B$ in the bias graph $G_B(V_B, E_B)$ of (G_1, G_2), is itself an end strongly connected component or else it contains, properly, some end strongly connected component of G_B.*

The first phase of the Two-Phase algorithm builds the bias graph of the input instance and locates all its ESCCs $G_B\langle C_i\rangle$. Each of these ESCCs is then used to determine a subset H_i of the input vertices such that H_i contains all vertices which appear in the labels of the bias graph's vertices that belong to $G_B\langle C_i\rangle$.

Its second phase runs the Bias Envelopment procedure on each of those subsets H_i only (due to Corollary 4), returning *yes* and a sandwich homogeneous set H that contains

The Two-Phase algorithm $(G_1(V, E_1), G_2(V, E_2))$

1. Construct the bias graph G_B of (G_1, G_2).
2. Find all end strongly connected component $G_B\langle C_i\rangle$ of G_B.
3. Let H_i be the set of vertices in V that label the vertices in $G_B\langle C_i\rangle$.
4. For each set $H_i \subset V$ do
4.1. $H \leftarrow H_i$.
4.2. Find the bias set $B(H)$.
4.3. While $H \neq V$ do
4.3.1. If $B(H) = \varnothing$ then return *yes* and H. End.
4.3.2. $H \leftarrow H \cup B(H)$.
4.3.3. Update $B(H)$.
5. Return *no*.

Figure 3: The Two-Phase algorithm

H_i, in case there exists one, or *no* in case none of the subsets H_i happen to be contained in any sandwich homogeneous sets of the input instance. Notice that it clearly saves time as compared to the EBE algorithm, which runs the Bias Envelopment procedure on all vertex pairs of the input instance.

Figure 3 presents the pseudo-code for the Two-Phase algorithm, whose time complexity is $O\left(m_1 n^2\right)$.

References

Cerioli, M. R., Everett, H., de Figueiredo, C. M. H., and Klein, S. (1998). The homogeneous set sandwich problem. *Information Processing Letters*, 67:31–35.

Dantas, S., Faria, L., and de Figueiredo, C. M. H. (2002). On the complexity of (k,l)-graph sandwich problems. *Lecture Notes in Computer Science*, Proceedings of WG 2002(2573):92–101.

de Figueiredo, C. M. H., Klein, S., and Vušković, K. (2002). The graph sandwich problem for 1-join composition is NP-complete. *Discrete Appl. Math.*, 121:73–82.

Golumbic, M. C. (1998). Matrix sandwich problems. *Linear Algebra Appl.*, 277:239–251.

Golumbic, M. C., Kaplan, H., and Shamir, R. (1995). Graph sandwich problems. *Journal of Algorithms*, 19:449–473.

Golumbic, M. C. and Wassermann, A. (1998). Complexity and algorithms for graph and hypergraph sandwich problems. *Graphs Combin.*, 14:223–239.

Kaplan, H. and Shamir, R. (1999). Bounded degree interval sandwich problems. *Algorithmica*, 24:96–104.

Lovász, L. (1972). Normal hypergraphs and the perfect graph conjecture. *Discrete Math.*, 2:253–267.

Tang, S., Yeh, F., and Wang, Y. (2001). An efficient algorithm for solving the homogeneous set sandwich problem. *Information Processing Letters*, 77:17–22.