

Hankel Matrices: From Words to Graphs (Extended Abstract)

J.A. MAKOWSKY * and N. LABAI **

Department of Computer Science
Technion - Israel Institute of Technology, Haifa, Israel
e-mail: {nadia,janos}@cs.technion.ac.il,

Abstract. We survey recent work on the use of Hankel matrices $H(f, \square)$ for real-valued graph parameters f and a binary sum-like operation \square on labeled graphs such as the disjoint union and various gluing operations of pairs of labeled graphs. Special cases deal with real-valued word functions. We start with graph parameters definable in Monadic Second Order Logic MSOL and show how MSOL-definability can be replaced by the assumption that $H(f, \square)$ has finite rank. In contrast to MSOL-definable graph parameters, there are uncountably many graph parameters f with Hankel matrices of finite rank. We also discuss how real-valued graph parameters can be replaced by graph parameters with values in commutative semirings.

In this talk we survey recent work done together with the first author's former and current graduate students B. Godlin, E. Katz, T. Kotek, E.V. Ravve, and the second author on the definability of word functions and graph parameters and their Hankel matrix. There are three pervasive themes.

- Definability of word functions and graph parameters f in some logical formalism \mathcal{L} which is a fragment of Second Order Logic SOL, preferably Monadic Second Order Logic MSOL, or CMSOL, i.e., *MSOL* possibly augmented with modular counting quantifiers;
- Replacing the definability of f by the assumption that certain Hankel matrices have finite rank; and
- Replacing the field of real numbers \mathbb{R} by arbitrary commutative rings or semirings \mathcal{S} .

1 Hankel Matrices

In linear algebra, a *Hankel matrix*, named after Hermann Hankel, is a square matrix with constant skew-diagonals. In automata theory, a *Hankel matrix* $H(f, \circ)$ is an infinite matrix where the rows and columns are labeled with words w over a fixed alphabet Σ , and the entry $H(f, \circ)_{u,v}$ is given by $f(u \circ v)$. Here $f : \Sigma^* \rightarrow \mathbb{R}$

* Partially supported by a grant of Technion Research Authority.

** Partially supported by a grant of the Graduate School of the Technion.

is a real-valued word function and \circ denotes concatenation. A classical result of G.W. Carlyle and A. Paz [3] in automata theory characterizes real-valued word functions f recognizable by weighted (aka multiplicity) automata (WA-recognizable) in algebraic terms:

Theorem 1 (G.W. Carlyle and A. Paz, 1971).

A word function is WA-recognizable iff its Hankel matrix has finite rank.

Hankel matrices for graph parameters (aka connection matrices) were introduced by L. Lovász [33] and used in [14, 34] to study real-valued partition functions of graphs. In [14, 34] the role of concatenation is played by k -connections of k -graphs, i.e., graphs with v_1, \dots, v_k distinguished vertices. Given two k -graphs G, G' the k -connections $G \sqcup_k G'$ is defined by taking first the disjoint union of G and G' and then identifying corresponding labeled vertices. The Hankel matrix $H(f, \sqcup_k)$ is the infinite matrix where rows and columns are labeled by k -graphs and the entry $H(f, \sqcup_k)_{G, G'}$ is given by $f(G \sqcup_k G')$. We say that f has *finite connection rank* if all the matrices $H(f, \sqcup_k)$ have finite rank.

Partition functions are defined by counting weighted homomorphisms, which in some way generalize weighted automata. Let $H = (V(H), E(H))$ be a fixed graph, and let $\alpha : V(H) \rightarrow \mathbb{R}$ and $\beta : E(H) \rightarrow \mathbb{R}$ be real-valued functions (weights). For a graph $G = (V(G), E(G))$ we define

$$Z_{H, \alpha, \beta}(G) = \sum_{h: G \rightarrow H} \prod_{(v) \in V(G)} \alpha(h(v)) \cdot \prod_{(u, v) \in E(G)} \beta(h(u), h(v))$$

M. Freedman, L. Lovász and A. Schrijver [14] give the following characterization of partition functions:

Theorem 2 (M. Freedman, L. Lovász and A. Schrijver, 2007).

A real-valued graph parameter f can be presented as a partition function

$$f(G) = Z_{H, \alpha, \beta}(G)$$

for some H, α, β , iff all its connection matrices $H(f, \sqcup_k)$ have finite rank and are positive definite.

In [34] many variations of this theorem are discussed using different notions of connections of labeled graphs.

2 Definability in MSOL via Guiding Examples

Second Order Logic SOL allows quantification over vertices, edges and relations thereof. Monadic Second Order Logic MSOL allows quantification only over unary relations over the universe. In the case of graphs we have to distinguish between graphs G as structures where the universe $V(G)$ consists of vertices only and edges are given by the edge relation $E(G)$, and hypergraphs, where the universe consists of vertices and edges, and the hyperedges are given by an incidence relation $R(G)$. The notion of *definability of graph parameters and*

graph polynomials in SOL and MSOL was first introduced in [7] and extensively studied in [18, 12, 26, 22, 17, 25, 36, 24]. Later these studies also included MSOL augmented by modular counting quantifiers $D_{m,k}x \Phi(x)$ which assert that there are, modulo m , exactly k elements satisfying ϕ . This logic is denoted by CMSOL.

The set of real-valued graph parameters definable in SOL (MSOL, CMSOL) is denoted by SOLEVAL (MSOLEVAL, CMSOLEVAL), as they are evaluations of SOL-definable (MSOL, CMSOL-definable) graph polynomials. Words are treated here as special cases of labeled graphs.

Our first examples use *small*, i.e., polynomial sized sums and products:

- (i) The cardinality $|V|$ of V is FOL-definable by

$$|V| = \sum_{v \in V} 1$$

- (ii) The number of connected components of a graph G , $k(G)$ is MSOL-definable by

$$k(G) = \sum_{C \subseteq V: \text{component}(C)} 1$$

where $\text{component}(C)$ says that C is a connected component.

- (iii) The graph polynomial $X^{k(G)}$ is MSOL-definable by

$$X^{k(G)} = \prod_{c \in V: \text{first-in-comp}(c)} X$$

if, in addition, we have a linear order on the vertices and $\text{first-in-comp}(c)$ says that c is a first element in a connected component.

Our next examples use possibly *large*, i.e., exponential sized sums:

- (iv) The number of cliques $\#\text{Clique}(G)$ in a graph is MSOL-definable by

$$\#\text{Clique}(G) = \sum_{C \subseteq V: \text{clique}(C)} 1$$

where $\text{clique}(C)$ says that C induces a complete graph.

- (v) Similarly “the number of maximal cliques” $\#\text{MClique}(G)$ is MSOL-definable by

$$\#\text{MClique}(G) = \sum_{C \subseteq V: \text{maxclique}(C)} 1$$

where $\text{maxclique}(C)$ says that C induces a maximal complete graph.

- (vi) The clique number of G , $\omega(G)$ is SOL-definable by

$$\omega(G) = \sum_{C \subseteq V: \text{largest-clique}(C)} 1$$

where $\text{largest-clique}(C)$ says that C induces a maximal complete graph of largest size.

An inductive definition of a fragment \mathcal{L} of SOLEVAL can be sketched as follows:

Definition 3 Let \mathcal{R} be a (polynomial) ring. A numeric graph parameter $p : \text{Graphs} \rightarrow \mathcal{R}$ is \mathcal{L} -definable if it can be defined inductively as follows:

- Monomials are of the form $\prod_{\bar{v}:\phi(\bar{v})} t$ where t is an element of the ring \mathcal{R} and ϕ is a formula in \mathcal{L} with first order variables \bar{v} .
- Polynomials are obtained by closing under small products, small sums, and large sums.

Usually, *summation* is allowed over *second order variables*, whereas *products* are over *first order variables* only.

Our definition of SOLEVAL is somewhat reminiscent to the definition of *Skolem's Lower Elementary Functions*, [39, 37, 38].

3 The Finite Rank Theorem

In [18] the following Finite Rank Theorem is proved:

Theorem 4 (Finite Rank Theorem).

Let f be a real-valued graph parameter definable in CMSOL. Then f has finite connection rank.

The same holds for a wider class of Hankel matrices arising from sum-like binary operations on labeled graphs. A binary operation on labeled graphs is *sum-like* if it can be obtained from the disjoint union of two graphs by applying a quantifier-free scalar transduction, see e.g. [35, 4].

If we consider words instead of graphs, also the converse holds for the Hankel matrix of concatenation, [28, 29]:

Theorem 5 (NL and JAM, 2013). *A real-valued word function f is definable in MSOL iff its Hankel matrix for concatenation has finite rank.*

These results are reminiscent to results by [10], but their logical formalism differs from ours and was introduced later than MSOL-definability of graph parameters, [7].

The Finite Rank Theorem can also be used to show non-definability, [23, 24], which gives a more convenient and versatile tool than the usual methods involving Ehrenfeucht-Fraïssé games.

4 Meta-theorems Using Logic

The notions of path-width, tree-width and clique-width are the most used notions of width of graphs, cf. [20]. Widths are graph parameters with non-negative integer values. The exact definition is not needed here. What matters is that

graphs can have unbounded width of either kind. Classes of bounded path-width have bounded tree-width, which in turn have bounded clique-width, but not conversely.

B. Courcelle’s celebrated theorem for graph properties and graph classes of bounded tree-width [9, 4] says that on graph classes of bounded tree-width, MSOL-definable graph properties can be decided in linear time.

In [5], [6, Theorem 4], [7, Theorem 31], this is extended to graph parameters and bounded clique-width:

Theorem 6 (B. Courcelle, JAM, and U. Rotics, 1998). *Let f be a CMSOL-definable graph parameter with values in a ring \mathcal{R} . Then f can be computed in polynomial time¹ on graph classes of bounded clique-width.*

As a generalization of graph classes of given tree-width or clique-width, the notion of CMSOL-inductive classes of graphs was introduced in [35]. Special cases of CMSOL-inductive classes are the *sum-like inductive classes*.

Definition 7 (Sum-like inductive) \mathcal{C} is sum-like inductive if it is inductively defined using a finite set of basic labeled graphs $G_j, j \leq J$ and a finite set of sum-like binary operations $\square_i, i \leq I$. In other words, each $G_j, j \leq J$ is in \mathcal{C} , and whenever $H_1, H_2 \in \mathcal{C}$ then also $\square_i(H_1, H_2) \in \mathcal{C}$ for all $i \leq I$.

The classes of graphs of fixed tree-width (path-width, clique-width) are all sum-like inductive, cf. [35]. Other examples of sum-like inductive classes of labeled graphs can be found using various graph grammars, cf. [15, 16, 35]. In the framework of sum-like inductive classes, Theorem 6 can be stated in model theoretic terms, [35, Theorem 6.6].

Theorem 8 (JAM, 2004/14).

Let \mathcal{C} be sum-like inductive², and f be a graph parameter in CMSOLEVAL. Then the computation of $f(G)$ is Fixed Parameter Tractable³ in the size of the parse tree witnessing that $G \in \mathcal{C}$.

5 Eliminating Logic

L. Lovász, in [34], also noted that Hankel matrices can be used to make Courcelle’s Theorem *logic-free* for the case of bounded tree-width by replacing MSOL-definability by a finiteness condition on the rank of its connection matrices. In addition, graph parameters are allowed to take values in an arbitrary field \mathcal{K} .

¹ For real-valued graph parameters we have to be careful about the model of computation. Either we work in a Turing computable subfield of \mathbb{R} , or we use the computational model of Blum-Shub-Smale BSS, cf. [1].

² Originally the theorem was stated for MSOL-smooth operations. The proof I had in mind in [35] only works for sum-like operations. However, it is not known whether there are MSOL-smooth operations which are not sum-like.

³ A graph parameter is Fixed Parameter Tractable (FPT), if it can be computed in time $O(c(k) \cdot n^{d(k)})$ where n is the size of the graph, and $c(k), d(k)$ are functions depending on the parameter k , but independent of the size of the graph, cf. [9, 13]. Here the parameter is hidden in the fact that \mathcal{C} is CMSOL-inductive.

Theorem 9 (L. Lovász, 2007).

Let \mathcal{K} be a field and let f be a \mathcal{K} -valued graph parameter with finite connection rank. Then f can be computed in linear time on graph classes of bounded tree-width.

In [31, 30] this is extended to make Theorem 6 logic-free for the case of bounded clique-width. To do this one defines a suitable sum-like *binary* operation $\eta_{P,Q}$ on graphs with additional unary predicates $P(G), Q(G)$ on the vertices $V(G)$. $\eta_{P,Q}(G_1, G_2)$ is the disjoint union of G_1 and G_2 augmented with all the edges from

$$E_{P,Q} = \{(u, v) \in (V(G_1) \sqcup V(G_2))^2 : u \in P(G_1) \sqcup P(G_2) \text{ and } v \in Q(G_1) \sqcup Q(G_2)\}$$

In words $\eta_{P,Q}(G_1, G_2)$ is the disjoint union of G_1 and G_2 augmented with all the edges with one vertex in $P(G_1 \cup G_2)$ and one vertex in $Q(G_1 \cup G_2)$.

Theorem 10 (NL and JAM, 2014). Let f be a real-valued graph parameter with $H(f, \eta_{P,Q})$ of finite rank. Then f can be computed in polynomial time on graph classes of bounded clique-width.

In [27] this is further extended to make Theorem 8 also logic-free. A detailed discussion will appear in [32]. For this extension we introduce the notion of *linearly linked Hankel matrices*.

Definition 11 (Linearly linked Hankel matrices) Let $\square_i, i \leq I$ be finitely many binary operations on labeled graphs, and let $G_j, j \leq J$ be a finite set of basic graphs. $p_k, k \leq K$ be finitely many real-valued graph parameters. For a labeled graph H let $\bar{p}(H)$ denote the vector $(p_1(H), \dots, p_K(H))$.

1. \mathcal{C} is inductively defined using $G_j, j \in J$ and $\square_i, i \leq I$ if each $G_j, j \in J$ is in \mathcal{C} , and whenever $H_1, H_2 \in \mathcal{C}$ then also $\square_i(H_1, H_2) \in \mathcal{C}$. Here \square_i does not have to be sum-like.
2. The Hankel matrices $H(p_k, \square_i, i \leq I, j \leq J)$ are linearly linked if the following hold:
 - (a) For each $p_k, k \leq K$ and $\square_i, i \leq I$ the Hankel matrices $H(p_k, \square_i)$ are of finite rank.
 - (b) For each $i \leq I$ there is a matrix P_i such that for all graphs H_1, H_2

$$\bar{p}(\square_i(H_1, H_2)) = P_i \cdot \bar{p}(\square_1(H_1, H_2))$$

Theorem 12 (NL and JAM, 2014).

Let \mathcal{C} be inductively defined using $G_j, j \in J$ and $\square_i, i \leq I$, and let $p_k, k \leq K$ be finitely many graph parameters, such that the Hankel matrices $H(p_k, \square_i), i \leq I, j \leq J$ are linearly linked. Then for graphs $H \in \mathcal{C}$ with parse-tree $pt(H)$, all the graph parameters $p_k, k \leq K$ can be computed in polynomial time in the size of $pt(H)$.

Theorem 12 is a proper generalization of Theorem 8:

Proposition 13 *If \mathcal{C} is sum-like inductive using $G_j, j \in J$ and $\square_i, i \leq I$, and $f \in \text{CMSOLEVAL}$, there are finitely many graph parameters p_1, \dots, p_K such that all the Hankel matrices $H(p_k, \square_i), i \leq I, j \leq J$ are linearly linked.*

To prove Proposition 13 one uses the Bilinear Reduction Theorem from [35], which is proven in full detail as [24, Theorem 8.7].

In the logical versions of these theorems there are only countably many CMSOL-definable graph parameters. However, there are uncountably many graph parameters with finite rank Hankel matrices even for the disjoint union of graphs. Hence, in contrast to the case of word functions, the finiteness assumption on the rank does not imply MSOL-definability. Furthermore, eliminating logic from these theorems allows us to separate the algebraic character of the proof from its logical part given by the Finite Rank Theorem for sum-like operations.

6 From Fields to Semirings

Finally, we discuss how to formulate Theorem 8 both logic-free and for graph parameters with values in a commutative semiring. A motivating example for this shift of perspective is the clique number $\omega(G)$ of a graph G , which has infinite connection rank over the reals, but finite row-rank in the tropical semiring \mathcal{T}_{\max} , the max-plus algebra defined over the reals. There are several notions of rank for matrices over commutative semirings. All of them coincide in the case of a field, and some of them coincide in the tropical case, [2, 19, 8].

In [31, 30] Lovász's Theorem is generalized to graph parameters with values in the tropical semirings rather than a field, and graph classes of bounded clique-width. There we work with two specific notions: row-rank in the tropical case, and a finiteness condition introduced by G. Jacob [21], which we call J-finiteness, in the case of arbitrary commutative semirings.

Theorem 14 (NL and JAM, 2014). *Let f be a graph parameter with values in \mathcal{T}_{\max} with $H(f, \eta_{P,Q})$ of finite row rank. Then f can be computed in polynomial time on graph classes of bounded clique-width.*

In the case of graph parameters with values in arbitrary commutative semirings, this remains true for graph classes of bounded linear clique-width, cf. [11]. Linear clique-width relates to clique-width like path-width relates to tree-width.

References

1. L. Blum, F. Cucker, M. Shub, and S. Smale. *Complexity and Real Computation*. Springer Verlag, 1998.
2. P. Butkovič. *Max-linear Systems: Theory and Algorithms*. Springer Monographs in Mathematics. Springer, 2010.
3. J.W. Carlyle and A. Paz. Realizations by stochastic finite automata. *J. Comp. Syst. Sc.*, 5:26–40, 1971.
4. B. Courcelle and J. Engelfriet. *Graph Structure and Monadic Second-order Logic, a Language Theoretic Approach*. Cambridge University Press, 2012.

5. B. Courcelle, J.A. Makowsky, and U. Rotics. Linear time solvable optimization problems on graph of bounded clique width, extended abstract. In J. Hromkovic and O. Sykora, editors, *Graph Theoretic Concepts in Computer Science, 24th International Workshop, WG'98*, volume 1517 of *Lecture Notes in Computer Science*, pages 1–16. Springer Verlag, 1998.
6. B. Courcelle, J.A. Makowsky, and U. Rotics. Linear time solvable optimization problems on graphs of bounded clique-width. *Theory of Computing Systems*, 33.2:125–150, 2000.
7. B. Courcelle, J.A. Makowsky, and U. Rotics. On the fixed parameter complexity of graph enumeration problems definable in monadic second order logic. *Discrete Applied Mathematics*, 108(1-2):23–52, 2001.
8. R.A Cuninghame-Green and P Butkovič. Bases in max-algebra. *Linear Algebra and its Applications*, 389(0):107 – 120, 2004.
9. R.G. Downey and M.F Fellows. *Parametrized Complexity*. Springer, 1999.
10. M. Droste and P. Gastin. Weighted automata and weighted logics. *Theor. Comput. Sci.*, 380(1-2):69–86, 2007.
11. M.R. Fellows, F.A. Rosamond, U. Rotics, and S. Szeider. Proving NP-hardness for clique width i: Non-approximability of linear clique-width. *Electronic Colloquium on Computational Complexity*, 2005.
12. E. Fischer, T. Kotek, and J.A. Makowsky. Application of logic to combinatorial sequences and their recurrence relations. In M. Grohe and J.A. Makowsky, editors, *Model Theoretic Methods in Finite Combinatorics*, volume 558 of *Contemporary Mathematics*, pages 1–42. American Mathematical Society, 2011.
13. J. Flum and M. Grohe. *Parameterized complexity theory*. Springer, 2006.
14. M. Freedman, László Lovász, and A. Schrijver. Reflection positivity, rank connectivity, and homomorphisms of graphs. *Journal of AMS*, 20:37–51, 2007.
15. A. Glikson. Verification of generally intractable graph properties on graphs generated by graph grammars. Master's thesis, Technion - Israel Institute of Technology, Haifa, Israel, 2004.
16. A. Glikson and J.A. Makowsky. NCE graph grammars and clique-width. In H.L. Bodlaender, editor, *Proceedings of the 29th International Workshop on Graph-Theoretic Concepts in Computer Science (WG 2003), Elspeet (The Netherlands)*, volume 2880 of *Lecture Notes in Computer Science*, pages 237–248. Springer, 2003.
17. B. Godlin, E. Katz, and J.A. Makowsky. Graph polynomials: From recursive definitions to subset expansion formulas. *Journal of Logic and Computation*, 22(2):237–265, 2012.
18. B. Godlin, T. Kotek, and J.A. Makowsky. Evaluation of graph polynomials. In *34th International Workshop on Graph-Theoretic Concepts in Computer Science, WG08*, volume 5344 of *Lecture Notes in Computer Science*, pages 183–194, 2008.
19. A.E. Guterman. Matrix invariants over semirings. In M. Hazewinkel, editor, *Handbook of Algebra Volume 6*, volume 6 of *Handbook of Algebra*, pages 3 – 33. North-Holland, 2009.
20. P. Hlinený, S. Oum, D. Seese, and G. Gottlob. Width parameters beyond tree-width and their applications. *Comput. J.*, 51(3):326–362, 2008.
21. G. Jacob. *Représentations et substitutions matricielles dans la théorie algébrique des transductions*. PhD thesis, Université de Paris, VII, 1975.
22. T. Kotek. *Definability of combinatorial functions*. PhD thesis, Technion - Israel Institute of Technology, Haifa, Israel, March 2012.
23. T. Kotek and J.A. Makowsky. Connection matrices and the definability of graph parameters. In *CSL 2012*, pages 411–425, 2012.

24. T. Kotek and J.A. Makowsky. Connection matrices and the definability of graph parameters. *Logical Methods in Computer Science*, 10(4), 2014.
25. T. Kotek, J.A. Makowsky, and E.V. Ravve. A computational framework for the study of partition functions and graph polynomials (abstract). In V. Negru and et al. , editors, *SYNASC 2012*, Proceedings of the International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC, Timisoara, Romania, page in press. IEEE Computer Society, 2013.
26. T. Kotek, J.A. Makowsky, and B. Zilber. On counting generalized colorings. In M. Grohe and J.A. Makowsky, editors, *Model Theoretic Methods in Finite Combinatorics*, volume 558 of *Contemporary Mathematics*, pages 207–242. American Mathematical Society, 2011.
27. N. Labai. Hankel matrices and definability of graph parameters. Master’s thesis, Technion - Israel Institute of Technology, Haifa, Israel, 2015.
28. N. Labai and J.A. Makowsky. Weighted automata and monadic second order logic. In *Proceedings Fourth International Symposium on Games, Automata, Logics and Formal Verification, GandALF 2013, Borca di Cadore, Dolomites, Italy, 29-31th August 2013.*, pages 122–135, 2013.
29. N. Labai and J.A. Makowsky. Weighted automata and monadic second order logic. *arXiv preprint arXiv:1307.4472*, 2013.
30. N. Labai and J.A. Makowsky. Finiteness conditions for graph algebras over tropical semirings. *arXiv preprint arXiv:1405.2547*, 2014.
31. N. Labai and J.A. Makowsky. Tropical graph parameters. In *26th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2014)*, DMTCS Proceedings, pages 357–368, 2014.
32. N. Labai and J.A. Makowsky. to be determined. In preparation, 2015.
33. L. Lovász. Connection matrices. In G. Grimmet and C. McDiarmid, editors, *Combinatorics, Complexity and Chance, A Tribute to Dominic Welsh*, pages 179–190. Oxford University Press, 2007.
34. L. Lovász. *Large Networks and Graph Limits*, volume 60 of *Colloquium Publications*. AMS, 2012.
35. J.A. Makowsky. Algorithmic uses of the Feferman-Vaught theorem. *Annals of Pure and Applied Logic*, 126.1-3:159–213, 2004.
36. J.A. Makowsky, T. Kotek, and E.V. Ravve. A computational framework for the study of partition functions and graph polynomials. In *Proceedings of the 12th Asian Logic Conference '11*, pages 210–230. World Scientific, 2013.
37. R. Péter and I. Földes. *Recursive functions*. Academic Press New York, 1967.
38. H.E. Rose. *Subrecursion: functions and hierarchies*. Clarendon Press, Oxford, 1984.
39. T. Skolem. Proof of some theorems on recursively enumerable sets. *Notre Dame J. Formal Logic*, 3:65–74, 1963.