On the Convergence of Some Biclique Operators on Multipartite Graphs\textsuperscript{1, 2}

Christophe Crespelle\textsuperscript{3}, Matthieu Latapy\textsuperscript{4} and Thi Ha Duong Phan\textsuperscript{5}

Abstract

We introduce a new graph operator, called the weak factor graph, which is close to the well-known clique-graph operator but which rather operates in terms of bicliques in a multipartite graph, and we address the problem of the convergence of the series of graphs obtained by the iterative application of this operator. As for the clique-graph operator, it is easy to find graphs for which the series of weak factor graphs does not converge. Here, we show that we can slightly modify the weak-factor-graph operator, into an operator called clean factor graph, so that it converges for all graphs. Moreover, we show that the multipartite graph to which the series converge is a decomposition of a well-known combinatorial object: the inclusion order of the intersections of maximal cliques of the initial graph.

1 Introduction

The clique-graph operator [2] is a well-known graph operator which, given a graph $G$, undirected, simple and loopless, consists in building the graph $G'$ whose vertices are maximal cliques of $G$ and such that there is an edge between two distinct vertices of $G'$ iff the corresponding cliques of $G$ share at least one common vertex. The clique-graph series, obtained by iteratively applying the clique-graph operator to the graphs obtained starting from $G$, has been widely studied (see e.g. [3, 4]). This series is said to be convergent if one of the graph of the series is the graph with one single vertex. Then, all the graphs obtained in the following iterations are the same (i.e. reduced to a single vertex).

We call biclique-graph operator the natural bipartite generalisation of the clique-graph operator. Given a bipartite graph $G = (V_0, V_1, E)$, the biclique graph $G'$ of $G$ is the incidence bipartite graph between the non-trivial maximal bicliques of $G$ (those maximal bicliques that have at least two vertices in each part $V_0, V_1$ of the bipartition) and the vertices of $G$. More explicitly, $G'$ is a bipartite graph where the bottom vertices are the vertices $x$ of $G$, the upper vertices are the bicliques $B$ of $G$, and there is an edge between $x$ and $B$ in $G'$ iff $x \in B$. Starting from any bipartite graph $G$ and iteratively applying the biclique-graph operator on the previously obtained graph gives a series of bipartite graphs which is said to be convergent if one of the graph of the series has no non-trivial bicliques. It must be clear that the condition requiring that the bicliques used in the biclique-graph operator are non trivial is necessary for the series to converge for some graph. Indeed, otherwise, if we allowed to use all maximal bicliques, the series would not converge for any initial bipartite graph (except the bipartite graph with no edges) since any vertex defining a maximal biclique together with its neighbourhood would be replicated at each step of the series.

\textsuperscript{1}This work was partially supported by the PICS program of CNRS (France) and by the Vietnam National Fondation for Science and Technology Developement (NAFOSTED).
\textsuperscript{2}This article is a complete and improved version of the extended abstract appeared in [1].
\textsuperscript{3}Université Claude Bernard Lyon 1, DNET/INRIA, LIP UMR CNRS 5668, ENS de Lyon, Université de Lyon – christophe.crespelle@inria.fr
\textsuperscript{4}LIP6, CNRS and Université Pierre et Marie Curie (UPMC - Paris 6), 104 avenue du Président Kennedy, 75016 Paris, France – Firstname.Lastname@lip6.fr
\textsuperscript{5}Institute of Mathematics, 18 Hoang Quoc Viet, Hanoi, Vietnam – phanhaduong@math.ac.vn
The weak-factor graph operator, which is the subject of this paper, is a simple rewriting of the biclique-graph operator in terms of multipartite graphs. This rewriting has a strong interest: it keeps an explicit and complete track of the original graph. More precisely, in the weak-factor graph, instead of placing all the vertices of $G$ at the bottom level, as in the biclique graph, we leave them on two different levels, we keep the edges between them and we place the vertices corresponding to maximal bicliques on a new upper level, thereby obtaining a tripartite graph. Then, at the next iteration, we consider the maximal bicliques in the bipartite graph induced by the edges between the upper level and all the other levels. The vertices corresponding to these maximal bicliques form a new upper level, and we obtain a 4-partite graph, and so on. Thus, the general definition of the weak-factor graph operates on multipartite graphs rather than bipartite graphs. Given a $k$-partite graph $G$, where $k \geq 2$, together with a $k$-partition $(V_0, \ldots, V_{k-1})$ of its vertices such that there are no edges between vertices of $V_i$ for all $i$ between 0 and $k-1$, the weak-factor graph $G'$ of $G$ is defined as follows. $G'$ is the graph $G$ augmented with a new level $V_k$ of vertices, each of which corresponds to one non-trivial maximal biclique of the bipartite graph induced by the edges between the upper level of $G$, i.e. $V_{k-1}$, and the rest of the vertices of $G$, i.e. $\bigcup_{0 \leq i < k-2} V_i$. As in the biclique-graph operator defined above, we do not add a new vertex at level $V_k$ for all maximal bicliques but only for those which have at least two vertices in $V_{k-1}$ and two vertices in $\bigcup_{0 \leq i < k-2} V_i$ (i.e. non-trivial bicliques). For each new vertex $x$ at level $V_k$ corresponding to such a maximal biclique $B$, we define its neighbourhood in $G'$ as being the vertices of $B$. Note that, opposite to the case of the biclique-graph series, we do not apply the weak factor graph operator starting from an arbitrary bipartite graph, but only starting from the vertex-clique-incidence bipartite graph of some graph $G$. This allows us to define the series of weak factor graphs of a graph $G$ as the series obtained by iteratively applying the weak-factor-graph operator starting from the vertex-clique-incidence bipartite graph of $G$ (see Figure 1), where the vertices of $G$ are at level $V_0$ and the maximal cliques of $G$ are at level $V_1$. This series is said to be convergent iff, at some point, no new vertices are created. Then, all the following graphs are the same. For example, the series depicted on Figure 1 is convergent since no new vertices are created when applying the weak factor operator on graph $G_3$ of the series.

It is important to notice that the fact we start from the vertex-clique-incidence bipartite graph of some graph $G$ instead of an arbitrary bipartite graph is not a real restriction. Indeed, for an arbitrary bipartite graph $H = (V_0, V_1, E)$ it is straightforward to build a vertex-clique-incidence bipartite graph $H' = (V'_0, V_1, E')$ generating exactly the same series: simply add to $V_0$, for each vertex $y \in V_1$, a particularising vertex $x \in V'_0$ linked to $y$. Clearly, $H'$ is the vertex-clique-incidence bipartite graph of some graph and the series of $H'$ is exactly identical to the one of $B$, from level $V_1$ and above. In other words, the set of series generated by arbitrary bipartite graphs and the set of those generated by vertex-clique-incidence bipartite graphs are the same. Thus, in particular, the convergence results we obtain here for vertex-clique-incidence bipartite graph are actually more general and hold for arbitrary bipartite graphs.

Clearly, from their definitions, the weak-factor series converges for some graph $G$ iff the biclique-graph series is convergent for $G$. As we will show in Section 2, the weak-factor series is not always convergent. Then, the interest of the multipartite structure of the weak-factor graph is that it will allow us to restrict the definition of the non-trivial bicliques we use in the weak-factor operator, taking into account the different levels of the multipartite graphs. In this way, we will be able to devise a refined version of the weak-factor graph, which we call the clean factor graph, whose associated series converges for all graphs.

It is worth to mention that we did not come to the study of the convergence of the weak-factor series only for purely theoretic motivations: this question is of key interest in complex network modelling. Complex networks are those graphs encountered in practice in various domains such
as computer science, biology, social sciences and others. In the last decade, they were shown to share some non-trivial common properties [5, 6], independently from the context they come from. A lot of efforts have been done to design models able to capture these properties while staying general enough. One of the difficulty of the domain is to encompass in a same model the two major properties of these networks, namely their heterogeneous degree distribution and their high local density (clustering coefficient, see [7] for a formal definition). Among the most promising approaches, [8, 9] propose to model complex networks based on the properties of their clique incidence bipartite graph. Their idea is to use prescribed-degree-graph generation, which is a well understood technique since the works of [10, 11], for the clique-incidence graph instead of the graph itself, in order to generate the graphs by its cliques rather than generating it by its edges. In this way, they show that one obtains graphs having a high local density (thanks to the clique structure) and a heterogeneous degree distribution that is controlled by the degrees of the vertices in the clique incidence bipartite graph. However, the bipartite model suffers from a severe limitation: when generating the edges of the bipartite graph at random, the obtained neighbourhoods of the upper vertices intersect only on one (or zero) vertex with a very high probability (see [8, 9]). This is not the case in real world networks, where maximal cliques have non trivial overlaps (i.e. overlaps of cardinality at least two). Thus, even though it gives the desired properties concerning degree distribution and local density, the bipartite model results in graphs having a caricaturistic structure. The weak factor graph (see Section 2) is introduced in order to correct this drawback. The idea is to generate an object that encodes the non trivial intersections of maximal cliques of the network. Moreover, we wish to encode this structure by edges of a graph so that we can still use the prescribed-degree generation technique of [10, 11]. One can proceeds as follows. As previously, we denote by $V_0$ and $V_1$ the sets of the bipartition of the clique incidence graph, where $V_0$ is the set of vertices of $G$ and $V_1$ is the set of maximal cliques of $G$. If two or more maximal cliques $K_1, \ldots, K_i$ of $G$ share at least two vertices in common (that is we have a non trivial biclique $B$ in the clique incidence bipartite graph) we introduce a new vertex $x$ which is placed on a new upper level, denoted $V_2$, and we define the neighbourhood of $x$ as being the vertices of biclique $B$. Then, we can delete the edges of biclique $B$ since they are now encoded by the presence of $x$. After this operation, the neighbourhoods at level $V_0$ of the vertices of $V_1$ representing cliques $K_1, \ldots, K_i$ have less intersection than they previously had. If we proceed this way simultaneously for all the maximal non-trivial bicliques, we obtain a tripartite graph in which the neighbourhoods of vertices at level $V_1$ have no non-trivial intersection at level $V_0$, that is their intersection is
at most one. But still, the neighbourhoods of the new introduced vertices, at level $V_2$, may have some non-trivial intersections (i.e. intersections of cardinality at least two). Then, we can repeat the operation by considering the maximal bicliques between vertices at level $V_2$ and vertices on the other levels of the tripartite graph. We can repeat the factorising step until we hopefully obtain a multipartite graph without any non-trivial intersection of neighbourhoods. If the process stops, we obtain a multipartite graph that encodes the original graph: the latter can be retrieved from the multipartite graph obtained since the factorising operation is reversible. The interest of proceeding in this way is that we obtain an encoding of $G$ containing no non-trivial neighbourhood intersections. Thus, we can generate similar structures at random using the prescribed-degree generation method without bumping on the problem raised by [8, 9].

These are the reasons why we came to study the convergence of the weak factor graph series obtained from an arbitrary graph $G$. Note that, opposite to the process described in the explanation of our motivations in the previous paragraph, in the definition of the weak-factor-graph operator we do not delete edges of the bicliques involved in one factorisation step. The reason is that they are not involved in the further factorising steps. Thus, keeping those edges does not affect the structure of the obtained graphs: the set of nodes created at each step of the series is the same. In particular it does not modify the convergence of the series. On the other hand, keeping those edges helps to describe the structure of the graphs of the weak factor series, and this is precisely why we keep those edges in the formal definition given further (Definition 1) and in the rest of the paper.

Related works

As we mentioned previously, the weak-factor operator we introduce here is a bipartite version of the well-known clique graph operator which has received a lot of attention [2]. Most of the efforts that have been done on clique graph theory have been placed in studying the convergence of the clique-graph series of a graph [3, 4], which is precisely the question we address in this paper for the weak-factor operator. It is easy to find examples of graphs whose clique graph series does not converge: the cycles are such examples. Thus, many works focussed on obtaining convergence results, or divergence results, for some particular graphs or graph classes [12, 13, 14, 15]. Some other works have studied the graphs invariant under the clique-graph operator [16], algebraic properties of the clique graph operator [17], operations on graph $G$ that preserve the behaviour of the clique-graph series of $G$ [18], and the relation between convergence of the clique graph series and the fixed point properties on posets [19]. Among the related questions that have been studied is the problem of recognizing graphs that are the clique graph of some graph [20], which has been shown to be NP-complete [21]. Let us mention, that another graph operator called edge-clique-graph operator, which is closely related to clique graphs, has been studied (see e.g. [22, 23]). Finally, we note that very recently [24] showed the interest of clique graphs to study communities in complex networks. However, their approach and results are not equivalent to ours. In particular, they do not consider the series obtained by iterating the operator, which is our main concern here in the case of the weak-factor operator.

Our contribution

In this paper, we introduce a new graph operator called the weak-factor operator which is a bipartite version of the clique-graph operator. We show that there are graphs for which the weak-factor series, obtained by iteratively applying the weak-factor-graph operator, does not converge, which is the same situation as for the clique-graph operator. This led research works in clique-graph theory to focus on determining which graphs give rise to a convergent series for the clique-graph operator. Here, we follow an orthogonal and complementary approach.
that consists in slightly modifying the definition of the weak-factor operator in order to obtain convergence for all graphs.

There are two main contributions in this paper. The first one is to design a variation of the weak-factor operator, that we name the clean factor, for which we prove that the series terminates for any arbitrary graph. We note that the possibility to do so strongly relies on the multipartite nature of the graphs produced by the weak-factor operator: the levels of the multipartite graph keep tracks of the history of factorisation (each level corresponds to a factorisation step in the construction of the series) and the clean-factor-graph operator takes advantage of this information in order to avoid redundant operations that would lead to divergence of the series. A striking property of our result is that we can obtain convergence for all graphs while imposing constraints only a bounded number (namely 3) of levels in the definition of the clean-factor graph. In other words, examining only a bounded part of the factorisation history we can avoid to produce redundant factorisations that would threaten the convergence of the series.

The second contribution, which is maybe more important to us, is that we show that the multipartite graph on which the clean-factor series terminates is strongly related to a fundamental combinatorial object. Namely, its vertices are in bijection with the chains of the inclusion order of the non-trivial intersections of maximal cliques of the graph (Theorem 1), denoted $\mathcal{L}$ in the rest of the paper. More precisely, the vertices produced at level $V_k$ in the clean factor series correspond to the chains of $\mathcal{L}$ having length $k - 2$ (see Figure 2). In other words, each step of factorisation in the clean-factor series correspond to a refinement in the decomposition of $\mathcal{L}$ with respect to its chains: a new vertex created at level $V_k$ is a chain of length $k - 2$ in $\mathcal{L}$ obtained by refining a set of chains of length $k - 3$ (i.e. vertices at level $V_{k-1}$ in the clean-factor series). The set of chains of length $k - 3$ used in this refinement, as well as the way this refinement operates, is completely and explicitly characterised by recursion hypothesis $H_N$ in the proof of Theorem 1.

This link between iterative biclique (or even clique) factorisation and the chains of the inclusion order of the non-trivial intersections of maximal cliques of the graph (namely order $\mathcal{L}$) is quite natural. The reason why it was not highlighted before, despite of the great developments achieved in clique graph theory, is that the most general definition of the factorisation operation (i.e. the weak-factor graph in the context of bicliques) produces much more than the chains of $\mathcal{L}$. Our contribution is to give the exact conditions of restriction of the general definition (Definition 5) so that the iterative factorisation steps produce exactly the chains of $\mathcal{L}$. We believe this link to be fundamental. It may serve as a reference point to understand why the general definition of the factorisation operation converges or not on such class of graphs and why such refined definition converges or not for all graphs. Moreover, we note that order $\mathcal{L}$ is not defined in terms of bicliques but in terms of cliques. Then, transferring our results to the context of classic clique graph theory would be of high interest and may lead to significant developments in this context as well.

Finally, we give an upper bound on the size and computation time of the graph on which the iterated clean-factor series of $G$ terminates, under reasonable hypothesis on the degree distributions of the clique-incidence-bipartite graph of $G$ (which hold for most real-world complex networks), therefore showing that this multipartite graph can be used in practice for complex network modelling.

Let us mention that this work is an improved and complete version of the extended abstract appeared in [1]. In [1], the notion of clean factor (introduced further) is slightly different from the one we use here. As a consequence, we could not prove that imposing constraints only a bounded number of history steps is enough to guarantee convergence of the series, and we did not have a real bijection between the multipartite graph obtained at convergence and the vertices of $\mathcal{L}$. 
Figure 2: Top left: a graph $G$ and its maximal cliques. Top right: the inclusion order $\mathcal{L}$ of the non-trivial intersections of maximal cliques of $G$. Bottom: the multipartite graph $M$ obtained at convergence of the clean factor series of $G$. $M$ has 5 levels. Level $V_0$ is for the vertices of $G$ and level $V_1$ for the maximal cliques of $G$. The bijection between the vertices of $M$ and the chains of order $\mathcal{L}$ appear in the rest of the levels: each vertex $x$ of $M$ in levels $V_2$ to $V_4$ has been labeled with the corresponding chain of $\mathcal{L}$, that is its characterising sequence $S(x)$ (see Definition 8). Level $V_2$ is for the non-trivial intersections of maximal cliques of $G$ (i.e. the chains of $\mathcal{L}$ of length 0), level $V_3$ is for the chains of $\mathcal{L}$ of length 1, and level $V_4$ is for the chains of $\mathcal{L}$ of length 3, which is precisely the height of $\mathcal{L}$. For sake of clarity, only a few links of $M$ have been drawn on the figure, namely all the links between two consecutive levels higher than level $V_2$. The black bold lines are for the link between a vertex $x$ at level $V_i$, for $i \geq 3$, and the unique vertex of $V_{i-1}$ whose characterising sequence is a prefix of the sequence of the characterising sequence of $x$ (necessarily of length $\text{length}(S(x)) - 1$). In this way, we obtain a clearer representation of the bijection with the chains of $\mathcal{L}$, as one can clearly see the four prefix trees of the chains starting at a given element of $\mathcal{L}$ (i.e. rooted at level $V_2$).

Outline of the paper

In the rest of this introduction section, we give a few notations and basic definitions, useful in the whole paper, including the definition of a fundamental notion, the *factorisation*, which will play a key role in the following. In Section 2, we formally define the weak-factor operator and the associated series, and we show that there are some graphs for which the series is not convergent. Based on this example of divergent graph, we introduce a natural variation of this operator, called the factor operator, which seems to converge but for which the question remains open.
In Section 3 we propose a deeper refinement of the operator, called the clean-factor operator, for which we prove convergence for all graphs. We also prove that the multipartite graph on which the clean-factor series terminates is a decomposition of the inclusion order $L$ of the non-trivial intersections of maximal cliques of the initial graph. Finally, in Section 4, we address the practical problem posed by the computation of the clean-factor series, and we give some evidence of the fact that, under reasonable assumptions on the degree distribution of the initial clique-incidence-bipartite graph, the clean-factor decomposition of a graph can be efficiently computed and stored. Section 5 gives some conclusions and perspectives of our work.

Notations and preliminary definitions

All graphs considered here are finite, undirected and simple (no loops and no multiple edges). A graph $G$ having vertex set $V$ and edge set $E$ will be denoted by $G = (V, E)$. We also denote by $V(G)$ the vertex set of $G$. The edge between vertices $x$ and $y$ will be indifferently denoted by $xy$ or $yx$. $K(G)$ denotes the set of maximal cliques of a graph $G$, and $N(x)$ the neighbourhood of a vertex $x$ in $G$.

A $k$-partite graph $G$ is a graph whose vertex set is partitioned into $k$ parts, with edges between vertices of different parts only (a bipartite graph is a 2-partite graph, a tripartite graph a 3-partite graph, etc): $G = (V_0, \ldots, V_{k-1}, E)$, where the $V_i$’s are pairwise disjoint, and with $E \subseteq \{uv \mid u \in V_i, v \in V_j, i \neq j\}$. The vertices of $V_i$, for any $i$, are called the $i$-th level of $G$, and the vertices of $V_{k-1}$ are called its upper vertices.

When $G = (V_0, \ldots, V_{k-1}, E)$ is $k$-partite, we denote by $N_i(x)$, where $0 \leq i \leq k-1$, the set of neighbours of $x$ at level $i$: $N_i(x) = N(x) \cap V_i$. A biclique of a graph is a set of vertices of the graph inducing a complete bipartite graph. We denote by $B(G)$ the clique incidence graph of $G = (V, E)$: $B(G) = (V, K(G), E')$ where $E' = \{vc \mid c \in K(G), v \in c\}$. A non-trivial biclique of a bipartite graph is a biclique having at least two vertices in the upper level and at least two vertices on the bottom level. Two sets have a non-trivial intersection if they share at least two elements. In all the paper, we denote $L$ the inclusion order of the non-trivial intersections of maximal cliques of a graph $G$ (there will be no confusion on the graph $G$ referred to when we use this notation).

For two non-negative integers $a, b \in \mathbb{N}$, we use the notation $[a, b]$ for the set $\{p \in \mathbb{N} \mid a \leq p \leq b\}$, with the convention $[a, b] = \emptyset$ if $a > b$.

In all the paper, an operation will play a key role, we name it factorisation and define it generically as follows.

**Definition 1 (factorisation with respect to $V'_k$)** Given a $k$-partite graph $G = (V_0, \ldots, V_{k-1}, E)$ with $k \geq 2$ and a set $V'_k$ of subsets of $V(G)$, we define the factorisation of $G$ with respect to $V'_k$ as the $(k+1)$-partite graph $G' = (V_0, \ldots, V_k, E \cup E'_+)$ where:

- $V_k$ is the set of maximal (with respect to inclusion) elements of $V'_k$,
- $E_+ = \{XY \mid X \in V_k \text{ and } y \in X\}$.

When $V_k \neq \emptyset$, the factorisation is said to be effective.

**Definition 2 (series associated to a factorisation operation)** Given a factorisation operation that associates any $k$-partite graph $G = (V_0, \ldots, V_{k-1}, E)$ with $k \geq 2$ to a $k+1$-partite graph $G'$ obtained by factorisation of $G$ with respect to some set $V'_k$ (see Definition 1), we define the series of multipartite graphs $(G_i)_{i \geq 1}$, associated to this factorisation operation and generated by a graph $G_0 = (V_0, E_0)$, by: $G_1 = B(G_0)$ is the vertex-clique-incidence bipartite graph of $G_0$ (in which the cliques are on the upper level of $B(G_0)$) and, for all $i \geq 1$, $G_{i+1} = G'_i$. If for some $i \geq 1$ the factorisation is not effective then we say that the series is convergent.
In the rest of the paper, we will refine the notion of factorisation by using different sets $V'_k$ on which is based the factorisation operation, and we will study termination of the graph series resulting from each of these refinements.

2 Weak factor series and factor series

Weak factor series

We will now formally define the factorisation process and show that it may result in an infinite sequence of graphs. In the following sections, we will restrict the definition of the factorising step in order to always obtain a finite representation of the graph.

**Definition 3 (V^+_k and weak factor graph)** Given a $k$-partite graph $G = (V_0, \ldots, V_{k-1}, E)$ with $k \geq 2$, we define the set $V^+_k$ as:

$$V^+_k = \{\{x_1, \ldots, x_l\} \cup \bigcap_{1 \leq i \leq l} N(x_i) \mid l \geq 2, \forall i \in [1, l], x_i \in V_{k-1} \text{ and } |\bigcap_{1 \leq i \leq l} N(x_i)| \geq 2\}.$$  

The weak factor graph $G^+$ of $G$ is the factorisation of $G$ with respect to $V^+_k$ and the weak factor series $\text{WFS}(G) = (G_i)_{i \geq 1}$ of a graph $G_0 = (V_0, E_0)$ is the series of multipartite graphs associated to the weak factor graph operation and generated by $G_0$ (see Definition 2).

Figure 3: An example graph for which the weak factorising process is infinite. From left to right: the original graph $G$, its bipartite decomposition $B(G)$, and its tripartite decomposition $B(G)^+$. The shaded edges are the ones involving vertex $e$, which plays a special role: all the vertices of the upper level of the decompositions are linked to $e$. The structure of the tripartite decomposition is very similar to the one of the bipartite decomposition, revealing that the process will not terminate.

Figure 1 gives an illustration for this definition. In this case, the weak factor series is finite. However, this is not true in general; see Figure 3. Intuitively, this is due to the fact that a vertex may be the base for an infinite number of factorisation steps (like vertex $e$ in the example of Figure 3). The aim of the next sections is to avoid this case by giving more restrictive definitions.

Factor series

We have introduced above the weak-factor series, which is the most general definition of the biclique factorisation process, but we showed that the series is not necessarily finite. In this section, we introduce a slightly more restricted definition that forbids the repeated use of a same vertex to produce infinitely many factorisations (as observed on the example of Figure 3). However, we have no proof that it necessarily gives finite series, which remains an open question.

**Definition 4 (V^0_k and factor graph)** Given a $k$-partite graph $G = (V_0, \ldots, V_{k-1}, E)$ with $k \geq 2$, we define the set $V^0_k$ as:

$$V^0_k = \{X \in V^+_k \text{ such that } |\bigcap_{y \in X \cap V_{k-1}} N_{k-2}(y)| \geq 2\}.$$  

The factor graph $G^k$ of $G$ is the factorisation of $G$ with respect to $V^k$ and the factor series $\mathcal{F}(G) = (G_i)_{i \geq 1}$ of a graph $G_0 = (V_0, E_0)$ is the series of multipartite graphs associated to the factor graph operation and generated by $G_0$ (see Definition 2).

This new definition results from the restriction of the weak factor definition by considering only sets $X \in V^k$ such that the vertices of $X \cap V^{k-1}$ have at least two common neighbours at level $k-2$. The reason is that, in this way, a vertex cannot contribute to more than two factorising steps: once when it is on the upper level of the multipartite graph, once when it is on the level just below. Indeed, even if the vertices of levels lower than the two upper levels may be involved in a factorisation step, they are not responsible for the creation of a new vertex. Such a creation depends only the edges between the two upper levels of the multipartite graph. This restriction also plays a key role in the convergence proof of the clean factor series, defined in next section. That is why we think it may be possible that it is sufficient to guarantee the convergence of the factor series. But we could not prove it with this sole hypothesis and neither we could build a counterexample for which the factor series does not converge.

**Open question 1** Does the factor series converge for all graphs?

One may think of a natural candidate to be a counterexample to the convergence of the factor series: the graph whose clique-incidence-bipartite graph is the anti-matching. This graph is the graph on $n$ vertices having $n$ maximal cliques that contain all vertices of the graph except one (see Figure 4 for the example of the anti-matching of size 5). Indeed, the anti-matching of size $n$ presents a combinatorial explosion of the number of vertices in the first levels that may threaten the convergence of the series\(^6\). Each subset of vertices of level $V_1$ containing between 2 and $n-2$ vertices gives rise to a vertex on level $V_2$, which gives $2^n - 2n - 2$ vertices on level $V_2$. The vertices on level $V_3$ are in bijection with the set of couples $(P, Q) \in 2^{V_1} \times 2^{V_0}$ such that $2 \leq |P| \leq n - 3$ and $Q \subseteq P'$, where $P'$ is the set of vertices on level $V_0$ that are not adjacent to some vertex of $\bar{P}$, the complement of $P$ in $V_1$. This gives $\sum_{i=2}^{n-3} \binom{n}{i} (2^{n-i} - 1)$ vertices on level $V_3$. The factor series of the anti-matching of size 5 is given in extension on Figure 4. One can observe that despite the combinatorial explosion of the number of vertices in the first levels, the series terminates. Actually, we could prove that this result also holds for the anti-matching of size $n$, with an arbitrary $n \in \mathbb{N}$. Our proof is not straightforward and has been omitted here, let us simply mention that it uses a technique different from the one we will use in the next section to prove convergence of the clean-factor series.

Thus, the question of determining whether the factor series converges for all graphs or not (Open question 1) is challenging and it would be of great interest either to provide a proof or a counterexample. As we could not prove it we introduce, in the next section, a supplementary condition on the factor operation: we do not only require that the neighbourhoods of vertices at level $V_{k-1}$ involved in the creation of a new vertex at level $V_k$ share at least two vertices on level $V_{k-2}$ but we also require that the vertices at level $V_{k-1}$ involved have the same neighbourhood at level $V_{k-3}$ (see Definition 5 of the clean-factor graph). This supplementary condition is not only a technical condition used to guarantee convergence: we will show that the clean-factor series is a fundamental combinatorial object that shed new light on biclique (and clique) factorisation operators.

\(^6\)The anti-matching of size $n$ is known to be the bipartite graph on $2n$ vertices that maximizes the number of maximal bicliques.
Figure 4: The factor series converges for the anti-matching of size 5. The picture shows part of the multipartite graph generated by the factor series. The series stops at level $V_6$. There are nine other parts that have been omitted on the picture, and that are strictly similar to the one which is depicted. Levels $V_0$ and $V_1$ are respectively the vertices and the maximal cliques of the considered graph. The dashed lines stand for the non-edges of the bipartite graph between level $V_0$ and $V_1$, called anti-matching. More precisely, each vertex at level $V_1$ is adjacent to all the vertices on level $V_0$, except one of them, its corresponding vertex. For sake of clearness, the edges of the multipartite graph are not represented. Instead, for each vertex in the levels above $V_2$, the list of its neighbours in the lower levels appears below its name. There are 20 vertices at level $V_2$, each one corresponding to a subset of vertices at level $V_1$ having cardinality 2 or 3. At level $V_3$, and above, the multipartite graph splits into ten disjoint parts that have no edges between them. The picture shows only the part of vertices having neighbours $A$ and $B$ at level $V_1$. There is a similar part for each pair $\{\alpha, \beta\} \subseteq \{A, B, C, D, E\}$ of vertices at level $V_1$. Let us explain the reason of this separation into ten disjoint parts. By definition of the factor operator, the vertices $u$ at level $V_3$ must have at least two neighbours at level $V_1$. But they cannot have 3 such neighbours since only one vertex at level $V_2$ is adjacent to all of these three neighbours and may then be used to build $u$. Consequently, $u$ must have exactly two neighbours at level $V_1$, say $A, B$, and $u$ can be built only using those vertices at level $V_2$ having both $A$ and $B$ as neighbours (namely $x, y, z, t$ in the case of the pair $\{A, B\}$). Now consider two distinct pairs of vertices at level $V_1$, say $A, B$ and $A, C$ for example. For two such pairs, since the vertices at level $V_2$ have at most three neighbours at level $V_1$, there is at most one vertex at level $V_2$ having as neighbours all the vertices of the two pairs (there is none if the two pairs are disjoint). Thus, two vertices at level $V_3$ corresponding to two different pairs of vertices at level $V_1$ may share at most one common neighbour at level $V_2$, and consequently they cannot be involved together in the creation of a new vertex at level $V_4$. This is the reason why the multipartite graph splits into ten disjoint parts from level $V_3$ and above. The figure entirely shows the part corresponding to the pair $A, B$. One can see that the series terminates at level $V_6$ as there are no possible factorisations at this step. The whole multipartite graph (including the ten disjoint parts) have 20 vertices at level $V_2$, $10 \times 7$ vertices at level $V_3$, $10 \times 12$ vertices at level $V_4$, $10 \times 15$ vertices at level $V_5$ and $10 \times 6$ vertices at level $V_6$, where the series stops despite the combinatorial explosion on the first four levels ($V_2$ to $V_5$).
3 Clean factor series

In the two previous sections, we studied two series of multipartite graphs based on two different factorisation operations. The first one, the weak-factor operation, is very natural but its associated series is not always convergent, even for very simple graphs (see Figure 3). The second one, the factor operation, is a refinement of the weak factor which remains very general but for which we were unable to prove whether its associated series is always convergent or not.

In this section, we introduce a more restricted refinement of the two factorisation operations above, which we call the clean-factor operator, for which we prove that the associated series is always convergent. The interest of our results essentially lies in two points. First, it shows that it is possible to design a non-trivial refinement of the weak-factor operator, which is the more general one, in order to obtain the convergence of the operator on all graphs. Moreover, we show that it is not needed to keep the whole history of factorisation to do so: in the definition of the clean factor, it is sufficient to impose restricted conditions on three levels of the multipartite graph, namely $V_{k-2}$, $V_{k-3}$ and $V_1$ (it is even possible to remove the conditions imposed on $V_1$ and still guarantee the convergence for all graphs).

Second, the multipartite graph obtained at termination of the clean factor series is a decomposition of a fundamental combinatorial object: the inclusion order $\mathcal{L}$ of the non-trivial intersections of maximal cliques of the initial graph $G$. This shows strong links between biclique factorisation processes on the vertex-clique-incidence bipartite graph of $G$ and the structure of the set family made of the non-trivial intersections of maximal cliques of $G$. At last, the clean factor operator introduced in this section may also help in designing a proof for the convergence of the factor operator, or lead to the construction of a counter-example.

We now give the formal definition of the clean factor of a multipartite graph: the general factorisation step is the case where $k \geq 5$, the construction of levels $V_2, V_3$ and $V_4$ are subject to particular conditions. This definition is rather involved, mainly because of the particular conditions imposed on the first levels. We would like to emphasize on the fact that some of these conditions are not necessary to convergence of the series. But, on the other hand, omitting them would result in the creation of other vertices (not only in the lower levels) and we would not obtain the exact bijection with the chains of $\mathcal{L}$, which is our main concern here.

**Definition 5 ($V^*_k$ and clean factor graph)** Given a $k$-partite graph $G = (V_0, \ldots, V_{k-1}, E)$ with $k \geq 2$, we define the set $V^*_k$ as:

- If $k = 2$, $V^*_2 = V^+_2$.
- If $k = 3$, $V^*_3 = \{ X \in V^+_3 | | \cap_{x \in X \cap V_2} N_1(x) | \geq 2 \ and | \cap_{x \in X \cap V_2} N_0(x) | \geq 2 \}.$
- If $k = 4$, $V^*_4 = \{ X \in V^+_4 | | \cap_{x \in X \cap V_3} N_2(x) | \geq 2 \ and | \cap_{x \in X \cap V_3} N_1(x) | \geq 2 \ and \ \forall x, y \in X \cap V_3, N_0(x) = N_0(y) \}.$
- If $k \geq 5$, $V^*_k = \{ X \in V^+_k | | \cap_{x \in X \cap V_{k-1}} N_{k-2}(x) | \geq 2 \ and \ \forall x, y \in X \cap V_{k-1}, N_{k-3}(x) = N_{k-3}(y) and | \cap_{x \in X \cap V_{k-1}} N_1(x) | \geq 2 \}.$

The clean factor graph $G^*$ of $G$ is the factorisation of $G$ with respect to $V^*_k$ and the clean factor series $\mathcal{CFS}(G) = (G_i)_{i\geq 1}$ of a graph $G_0 = (V_0, E_0)$ is the series of multipartite graphs associated to the clean factor graph operation and generated by $G_0$ (see Definition 2).

The restricted conditions of the clean factor operator apply on the neighbourhoods on the lower levels (i.e. $V_0, \ldots, V_{k-2}$) of the set $X$ of vertices of the upper level $V_{k-1}$ involved in the creation of a new vertex $x$ at level $V_k$. There are two types of conditions: cardinality condition:
the neighbourhoods of vertices of \( X \) must share at least two vertices in common at a given level; equality condition: the neighbourhoods of vertices of \( X \) must be the same at a given level.

More precisely, the general conditions, for \( k \geq 5 \), require the cardinality condition at level \( V_{k-2} \), as in the factor operation (see Definition 4), the equality condition at level \( V_{k-3} \), which is the supplementary condition that allows us to prove the convergence of the series and the bijection between vertices of the obtained multipartite graph and the chains of \( L \). The cardinality condition required at level \( V_1 \) is not essential to the convergence result, which remains valid without it, but is necessary in order to obtain a clean bijection with \( L \). Actually, the role of level \( V_1 \) is quite singular. As we will see in the proof of Theorem 1 (recursion hypothesis \( H_E \)), the equality condition at level \( V_{k-3} \) implies the equality condition at all levels \( V_i \), with \( i \leq k-3 \), except at level \( V_1 \). This is a key property of the clean factor which allows to obtain convergence by imposing restrictions on only three levels: the control of the equality of neighbourhoods at level \( V_3 \) induces a control on all the lower levels, except \( V_1 \).

Because of side constraints, the general conditions for \( k \geq 5 \) must be adapted for \( k \in \{2, 3, 4\} \) in the following way. For \( k = 2 \), the conditions required for \( k \geq 5 \) on levels \( V_{k-3} \) and \( V_1 \) have no meaning. They are simply omitted and the only condition remaining is the one on \( V_{k-2} = V_0 \), which gives \( V_2^* = V_2^+ \). For \( k = 3 \), the conditions on \( V_1 \) and \( V_{k-2} = V_1 \) are identical, and are unchanged compared to the general case where \( k \geq 5 \). Moreover, the equality condition on level \( V_{k-3} = V_0 \) is too strong and would lead to no creations of vertices at level 3. Instead, this equality condition is replaced by the cardinality condition, which is weaker. This latter condition is necessary in order to get a clean bijection with \( L \) but could be omitted to guarantee convergence. At last, for \( k = 4 \), the cardinality condition on level \( V_{k-2} = V_2 \) is unchanged. The equality condition on level \( V_{k-3} = V_1 \) and the cardinality condition on \( V_1 \) are concurrent: in order to get the clean bijection, we keep the cardinality condition, which is the weaker one and which respects the specific role of level \( V_1 \). Finally, we add a supplementary equality condition on level \( V_0 \), which is implied, in the general case where \( k \geq 5 \), by the equality condition on level \( V_{k-3} \), which is absent in the case \( k = 4 \).

The rest of this section is devoted to proving the convergence of the clean factor series \( (G_i)_{i \geq 1} \) generated by any graph \( G \) (Theorem 2) and the bijection between vertices of the multipartite graph \( G_i \), \( i \geq 2 \), and the chains of length \( i-2 \) of \( L \) (Theorem 1). We start by proving Theorem 1 since Theorem 2 will be obtained as a direct corollary from it.

Theorem 1 gives a characterisation of \( V_i \), \( i \geq 2 \) by associating to each of its nodes a chain of length \( i-2 \) in the inclusion order \( L \) of the non-trivial intersections of maximal cliques of \( G \). Formally, we associate to a node \( x \) of \( V_i \) a sequence \( S(x) \) of subsets of \( V(G) \) which are precisely the elements of \( L \) defining the chain associated to \( x \). Before formally defining \( S(x) \) (Definition 8) and stating Theorem 1 we need to state some basic definitions, notations and properties of the non-trivial intersections of the maximal cliques of a graph.

**Definition 6** We denote by \( \mathcal{O}' \) the set of intersections of maximal cliques of \( G \) (possibly only one clique or none), that is \( \mathcal{O}' = \{O \subseteq V(G) \mid \exists P \subseteq K(G), O = \bigcap_{C \in P} C\} \), using the convention that \( \bigcap_{C \in \emptyset} C = V(G) \). And we denote by \( \mathcal{O} \) the subset of \( \mathcal{O}' \) formed with the elements that contain at least two vertices of \( G \) and that are obtained as the intersection of at least two distinct maximal cliques of \( G \), that is \( \mathcal{O} = \{O \in \mathcal{O}' \mid |O| \geq 2 \text{ and } \exists k \geq 2, C_1, \ldots, C_k \in K(G), (\forall j,l \in [1,k], j \neq l \Rightarrow C_j \neq C_l) \text{ and } O = \bigcap_{1 \leq i \leq k} C_i\} \).

**Definition 7** For any subset \( A \subseteq V(G) \) of vertices of \( G \), we denote by \( K(A) \) the set of maximal cliques of \( G \) containing \( A \), that is \( K(A) = \{C \in K(G) \mid A \subseteq C\} \). And we denote by \( \mathcal{C} \) the family of subsets \( Y \) of \( K(G) \) such that \( Y \) is precisely the subset of maximal cliques of \( G \) containing some subset of vertices \( O \in \mathcal{O}' \), that is \( \mathcal{C} = \{Y \subseteq K(G) \mid \exists O \in \mathcal{O}', Y = K(O)\} \).
**Remark 1** For any subsets $A, B \subseteq V(G)$, if $A \subseteq B$ then $K(B) \subseteq K(A)$. And for any subsets $A \subseteq V(G)$ and $O \in \mathcal{O}'$, if $K(O) \subseteq K(A)$ then $A \subseteq O$.

**Proof:** The first part of the remark is self-evident. For the second part, note that, by definition of $\mathcal{O}'$, $O = \bigcap_{C \in K(O)} C$. And on the other hand, we have $A \subseteq \bigcap_{C \in K(A)} C \subseteq \bigcap_{C \in K(O)} C = O$.

**Remark 2** For any $A, B \subseteq V(G)$, $K(A) \cap K(B) = K(A \cup B)$. Conversely, if $A_1, \ldots, A_n \subseteq V(G)$, with $n \geq 2$, and if $O \in \mathcal{O}'$ and if $\bigcap_{1 \leq i \leq n} K(A_i) = K(O)$, then $\bigcup_{1 \leq i \leq n} A_i \subseteq O$.

**Proof:** Let $A, B \in V(G)$. The cliques in $K(A) \cap K(B)$ are exactly the cliques that contain both $A$ and $B$, i.e., the cliques that contain $A \cup B$. Therefore $K(A) \cap K(B) = K(A \cup B)$.

Let $A_1, \ldots, A_n \subseteq V(G)$, with $n \geq 2$, and let $O \in \mathcal{O}'$ such that $\bigcap_{1 \leq i \leq n} K(A_i) = K(O)$. From what precedes, $\bigcap_{1 \leq i \leq n} K(A_i) = K(\bigcup_{1 \leq i \leq n} A_i)$. Consequently, we have $K(O) \subseteq K(\bigcup_{1 \leq i \leq n} A_i)$. And since $O \in \mathcal{O}'$, from Remark 1, we have $\bigcup_{1 \leq i \leq n} A_i \subseteq O$.

**Lemma 1** $\mathcal{O}'$ and $\mathcal{C}$ are closed under intersection.

**Proof:** The fact that $\mathcal{O}'$ is closed under intersection is clear from the definition. Let us show that $\mathcal{C}$ is closed under intersection. Let $k \geq 2$ and let $O_1, \ldots, O_k \in \mathcal{O}'$. Let us show that $\bigcap_{1 \leq i \leq k} K(O_i) \in \mathcal{C}$. To that purpose, consider a set $O \in \mathcal{O}'$ which is such that $O \supseteq \bigcup_{1 \leq i \leq k} O_i$, and which is minimal for inclusion among sets having this property, that is $O$ is such that $\forall \tilde{O} \in \mathcal{O}'$ if $\tilde{O} \supseteq \bigcup_{1 \leq i \leq k} O_i$ and $\tilde{O} \subseteq O$ then $\tilde{O} = O$. We will show that $\bigcap_{1 \leq i \leq k} K(O_i) = K(O)$. Since $O \supseteq \bigcup_{1 \leq i \leq k} O_i$, we have $K(O) \subseteq K(\bigcup_{1 \leq i \leq k} O_i) = \bigcap_{1 \leq i \leq k} K(O_i)$. Let us show the converse inclusion: $K(\bigcup_{1 \leq i \leq k} O_i) \subseteq K(O)$. Let $C \in K(\bigcup_{1 \leq i \leq k} O_i)$, by definition $\bigcap_{1 \leq i \leq k} O_i \subseteq C$. It follows that $\bigcup_{1 \leq i \leq k} O_i \cap C \subseteq O \cap C \subseteq O$. Clearly, $O \cap C \in \mathcal{O}'$ and, by minimality of $O$, it follows that $O \cap C = O$. Thus, $C \in K(O)$ and we have $K(\bigcup_{1 \leq i \leq k} O_i) \subseteq K(O)$.

Together with the previous inclusion, we obtain $\bigcap_{1 \leq i \leq k} K(O_i) = K(\bigcup_{1 \leq i \leq k} O_i) = K(O)$, which ends the proof.

The following lemma is the first step toward the bijection theorem (Theorem 1), it establishes the bijection between vertices of $V_2$ and the chains of length 0 of $\mathcal{L}$. We will use it in the initialising step of the recursion proving Theorem 1.

**Lemma 2** In the clean factor series, $V_0 = V(G)$, $V_1 = K(G)$ and $V_2 = \mathcal{O}$ in the sense that the map $\phi$ defined by $x \mapsto N_0(x)$ is a bijection from $V_2$ to $\mathcal{O}$. Moreover, $\forall x \in V_2, N_1(x) = K(N_0(x))$.

**Proof:** Let us start the second part of the lemma. Let $x \in V_2$. By definition of $V_2^+$, all the elements $y$ in $N_1(x)$ are such that $N_0(x) \subseteq N_0(y)$. Then, $y \in K(N_0(x))$, by identifying $V_1$ and $K(G)$. On the other hand, the maximality of $x$ in $V_2^+$ implies that all $y \in K(N_0(x))$ belong to $N_1(x)$. Thus, $N_1(x) = K(N_0(x))$.

Let us prove that the map $\phi$ is a bijection from $V_2$ to $\mathcal{O}$. First, if $x \in V_2$ then by definition, $|N_0(x)| \geq 2$, $|N_1(x)| \geq 2$, and $N_0(x) = \bigcap_{y \in N_1(x)} N_0(y)$, hence $N_0(x)$ belongs to $\mathcal{O}$, and the map $\phi$ is well defined.

Second, $\phi(x) = \phi(x')$ means $N_0(x) = N_0(x')$. But $N_1(x)$ is the set of all maximal cliques containing $N_0(x)$, and $N_1(x')$ is the same. Then, if $\phi(x) = \phi(x')$, we have $x = x'$: $\phi$ is injective.

We now prove that $\phi$ is surjective. Let $O$ be an element of $\mathcal{O}$, we show that the element $x = K(O) \cup \bigcap_{y \in K(O)} N_0(y)$ is an element of $V_2$ and $\phi(x) = O$. It is clear that $x \cap V_0 = O$, so $|x \cap V_0| \geq 2$. Since $O \in \mathcal{O}$, $|K(O)| \geq 2$, and we have $|x \cap V_1| \geq 2$. Then $x \in V_2^+$. Moreover, by definition, $K(O)$ is exactly the set of all maximal cliques containing $O$, then $x$ is maximal in $V_2^+$. It follows that $x$ is an element of $V_2$, and $\phi(x) = O$. 


We are now ready to give the definition of the sequence $S(x)$ that we associate to a vertex $x \in V_i$ with $i \geq 2$. Theorem 1 below will state that for each $i \geq 2$, the application $x \mapsto S(x)$ is a bijection from $V_i$ to the set of chains of $\mathcal{L}$ of length $i - 2$.

Definition 8 (Characterising sequence $S(x)$) Let $G$ be a graph and let $(G_i)_{i \geq 1}$ be its clean factor series. The characterising sequence $S(x) = (O_1(x), \ldots, O_{k-1}(x))$ of a vertex $x \in V_k$, with $k \geq 2$, is defined by:

- $O_1(x) = N_0(x)$, and
- for $k \geq 3$, $\forall j \in [2, k-1], O_j(x)$ is the unique element of $\mathcal{O}'$ such that $K(O_j(x)) = \bigcap_{y \in N_j(x)} N_1(y)$.

Note that $O_j$ is properly defined. Indeed, from Lemma 2, $\forall y \in V_2, V_1(y) \in \mathcal{C}$. And since $\mathcal{C}$ is closed under intersection, a simple recursion shows that for all $i \geq 3$ and for all $y \in V_i$, $N_1(y) = \bigcap_{z \in V_{i-1}} N_1(z) \in \mathcal{C}$. Then, for any $j \geq 2$, $\bigcap_{y \in N_j(x)} N_1(y)$ is in $\mathcal{C}$ and there exists some $O_j$ in $\mathcal{O}'$ satisfying the condition. The fact that such an $O_j$ in $\mathcal{O}'$ is unique comes from the fact that for any set $O \in \mathcal{O}'$, we have $O = \bigcap_{C \in K(O)} \mathcal{C}$. Then, if there exists some $O$ such that $K(O) = K(O_j)$, necessarily $O = O_j$. Consequently, $O_j$ is unique and properly defined.

We will often use the following remark in the proof of Theorem 1.

Remark 3 For any $x \in V_k$, with $k \geq 2$, $K(O_{k-1}(x)) = N_1(x)$.

Proof: For $k = 2$, the remark rewrites $K(O_1(x)) = N_1(x)$. Since $O_1(x) = N_0(x)$ and since, from Lemma 2, $K(N_0(x)) = N_1(x)$, then the result follows. For $k > 2$, the remark simply follows from the fact that $\bigcap_{y \in N_{k-1}(x)} N_1(y) = N_1(x)$.

We will now state the bijection theorem (Theorem 1) which is our main combinatorial tool for proving the convergence of the clean factor series (Theorem 2). Its proof is rather intricate, but it gives much more information than the convergence of the series. By associating a sequence of sets to each vertex in levels greater than $V_2$ in the multipartite graph, we show that each such vertex corresponds to a chain of the inclusion order $\mathcal{L}$ of the non-trivial intersections of maximal cliques of $G$. The correspondence thereby highlighted between this very natural structure and the multipartite factorisation scheme we introduced is non-trivial and of great combinatorial interest.

Theorem 1 (Bijection theorem) Let $G$ be a graph and $(G_i)_{i \geq 1}$ its clean factor series. For any $k \geq 2$, the map $\phi$ defined by $x \mapsto S(x)$ is a bijection from $V_k$ to $\{(O_1, \ldots, O_{k-1}) \in \mathcal{O}^{k-1} | O_1 \subset \ldots \subset O_{k-1}\}$ (see Figure 2).

Proof: The case $k = 2$ directly follows from Lemma 2, then, in the following, we only deal with the cases where $k \geq 3$. We prove Theorem 1 by recursion, using the five recursion hypothesis below, namely $H_{\text{tar}}, H_{\text{inj}}, H_{\text{sur}}, H_N$ and $H_E$. Actually, hypothesis $H_{\text{tar}}, H_{\text{inj}}$ and $H_{\text{sur}}$ are the targeted properties: they imply that map $\phi$ is a bijection. Hypothesis $H_N$ and $H_E$ contain the fundamental structure of the multipartite graph series. Hypothesis $H_N$ is essential, it gives a complete characterisation of the neighbourhood of a vertex on the lower levels. Hypothesis $H_E$ shows that the condition of equality of the neighbourhood at level $k - 3$ of the children of $x$ in Definition 5 actually induce a control of the neighbourhoods of the children of $x$ on all the lower levels.

$H_{\text{tar}}(k)$: If $x \in V_k$ then $O_1(x) \subset \ldots \subset O_{k-1}(x)$ and $(O_1(x), \ldots, O_{k-1}(x)) \in \mathcal{O}^{k-1}$,
we prove that $O H$

**Initialisation step.** We will prove that $H_N(k)$, $H_{tar}(3)$, $H_{inj}(3)$, and $H_{sur}(3)$ are true. We do not prove $H_E(k)$ since it is undefined. It is worth to note that we do not need $H_E(3)$ in the proof of $H_N(4)$: instead we use Definition 5 that provide us the initialisation we need.

**Proof of $H_N(k)$.** Since in $H_N(k)$, $j$ varies from 2 to $k - 1$, then in $H_N(3)$, $j$ only takes the value 2. Then, to prove $H_N(3)$, we just have to prove that for all $x \in V_3$, $N_2(x)$ is equal to $W_2(x) = \{ y \in V_2 \mid O_1(x) \subseteq O_1(y) \subseteq O_2(x) \}$. Let $x \in V_3$; we first show that $N_2(x) \subseteq W_2(x)$. We denote by $a_1, \ldots, a_l$, with $l \geq 2$, the elements of the set $N_2(x)$. Clearly, for any $i \in [1, l]$, $\bigcap_{1 \leq i \leq l} N_0(a_i) \subseteq N_0(a_i) \subseteq \bigcup_{1 \leq i \leq l} N_0(a_i)$. From Definition 8, we have $N_0(a_i) = O_1(a_i)$, $O_1(x) = N_0(x) = \bigcap_{1 \leq i \leq l} N_0(a_i)$, and $K(O_2(x)) = \bigcap_{1 \leq i \leq l} K(O_1(a_i))$. Moreover, from Remark 3, $N_1(a_i) = K(O_1(a_i))$, so we have $K(O_2(x)) = \bigcap_{1 \leq i \leq l} K(O_1(a_i))$. And since, by definition, $O_2(x) \in \mathcal{O}$, then, from Remark 2, $\bigcup_{1 \leq i \leq l} O_1(a_i) \subseteq O_2(x)$. Consequently, for any $i \in [1, l]$, $O_1(x) \subseteq O_1(a_i) \subseteq O_2(x)$. That is $N_2(x) \subseteq W_2(x)$.

Conversely, we show that if $y \in W_2(x)$, then $y \in N_2(x)$. To that purpose, we show that $N_0(x) \subseteq N_0(y)$ and $N_1(x) \subseteq N_1(y)$, which implies, by maximality of $x$ in $V_3^*$ (see Definition 1), that $y \in N_2(x)$. First, we have $N_0(x) = O_1(x)$, and since $y \in W_2(x)$, we also have $O_1(x) \subseteq O_1(y) = N_0(y)$. Then, $N_0(x) \subseteq N_0(y)$. Since $O_1(y) \subseteq O_2(x)$, we have $K(O_2(x)) \subseteq K(O_1(y))$. And from Remark 3, we have $K(O_1(y)) = N_1(y)$ and $K(O_2(x)) = N_1(x)$. Thus, $N_1(x) \subseteq N_1(y)$ and we conclude that $y \in N_2(x)$. Finally, we showed that $N_2(x) = W_2(x)$, and so $H_N(3)$ is true.

**Proof of $H_{tar}(3)$.** Since, from $H_N(3)$, $N_2(x) = W_2(x)$ and since $|N_2(x)| \geq 2$, it follows that $O_1(x) \subseteq O_2(x)$, otherwise $W_2(x)$ would contain at most one element. By definition of $V_3^*$, $|N_0(x)| \geq 2$. Since $O_1(x) = |N_0(x)|$, it follows that $O_1(x)$, and so $O_2(x)$, contains at least two elements. Moreover, from Remark 3, we have $K(O_2(x)) = N_1(x)$. And from Definition 5, $|N_1(x)| \geq 2$. It follows that $K(O_2(x))$ contains at least two elements, and so does $K(O_1(x))$ since $K(O_2(x)) \subseteq K(O_1(x))$. Thus $O_1(x)$ and $O_2(x)$ both belong to $\mathcal{O}$: $H_{tar}(3)$ is true.

**Proof of $H_{inj}(3)$.** For any $x \in V_3$, $N_2(x) = W_2(x)$. Thus, for any $x, y \in V_3$, $(O_1(x), O_2(x)) = (O_1(y), O_2(y))$ implies that $N_2(x) = N_2(y)$, which implies that $x = y$. So $H_{inj}(3)$ holds.

**Proof of $H_{sur}(3)$.** Let $O_1, O_2 \in \mathcal{O}$ such that $O_1 \subseteq O_2$. We will find an element $x$ of $V_3$ such that $O_1(x) = O_1$ and $O_2(x) = O_2$. Let $Y = \{ y \in V_2 \mid O_1 \subseteq O_1(y) \subseteq O_2 \}$. Let $x = Y \cup \bigcap_{y \in Y} N(y)$, we prove that $x$ is the desired element.

First, to prove that $x \in V_3$, we must prove that $x \in V_3^*$, that is $|x \cap V_2| \geq 2$ and $|x \cap V_1| \geq 2$ and $|x \cap V_0| \geq 2$. From Lemma 2, there exists two distinct elements $y_1, y_2 \in V_2$ such that $O_1(y_1) = O_1$ and $O_1(y_2) = O_2$. Clearly, $\{ y_1, y_2 \} \subseteq Y$, which gives $|x \cap V_2| \geq 2$. Furthermore, $x \cap V_1 = \bigcap_{y \in Y} N_1(y) = \bigcap_{y \in Y} K(O_1(y))$, from Remark 3. Since, for any $y \in Y$, $O_1(y) \subseteq O_2$,
we have also \( K(O_2) \subseteq K(O_1(y)) \). It follows that \( K(O_2) \subseteq x \cap V_1 \), and \( O_2 \in \mathcal{O} \) implies that \([K(O_2)] \geq 2\), then \( x \cap V_1 \geq 2\). We also have \( x \cap V_0 = \bigcap_{y \in Y} N_0(y) = \bigcap_{y \in Y} O_1(y) \). And since \( O_1 \subseteq O_1(y) \) for all \( y \in Y \), then \( O_1 \subseteq x \cap V_0 \). It follows that \( x \cap V_0 \geq 2 \).

We now show that \( x \) is maximal in \( V_3 \). Let \( z \in V_2 \setminus Y \) i.e. \( O_1 \not\subseteq O_1(z) \) or \( O_1(z) \not\subseteq O_2 \), we prove that \( x \cap V_0 \not\subseteq N_0(z) \) or \( x \cap V_1 \not\subseteq N_1(z) \), which implies that \( \bigcap_{y \in Y} N_0(y) = \bigcap_{y \in Y} O_1(y) = O_1 \), the last equality coming from the fact that \( \forall O \in \mathcal{O}, \exists z \in V_2, O(z) = O \) (see Lemma 2). On the other hand \( x \cap V_1 = \bigcap_{y \in Y} N_1(y) = \bigcap_{y \in Y} K(O_1(y)) = K(\bigcap_{y \in Y} O_1(y)) = K(O_2) \), again using Lemma 2 for the last equality. Now, if \( O_1(z) \not\subseteq O_2 \), since \( O_1(z) \in \mathcal{O} \), then, from Remark 1, \( K(O_2) \not\subseteq K(O_1(z)) \). And since, from what precedes, \( K(O_2) = x \cap V_1 \) and, from Remark 3, \( K(O_1(z)) = N_1(z) \), we obtain \( x \cap V_1 \not\subseteq N_1(z) \). On the other hand, if \( O_1 \not\subseteq O_1(z) \), since \( O_1 = x \cap V_0 \) (see above) and \( O_1(z) = N_0(z) \) (by definition), we obtain \( x \cap V_0 \not\subseteq N_0(z) \). Consequently, \( x \) is maximal in \( V_3 \) and then \( x \in V_3 \).

The last condition we have to check for proving \( H_{\text{sur}}(3) \) is that \( O_1(x) = O_1 \) and \( O_2(x) = O_2 \). We already have \( O_1 = N_0(x) = O_1(x) \). Moreover by definition, \( O_2(x) \) is the unique element of \( \mathcal{O} \) such that \( K(O_2(x)) = \bigcap_{y \in Y} N_1(y) \), and we also have from above that \( K(O_2) = \bigcap_{y \in Y} N_1(y) \), then \( O_2(x) = O_2 \).

Finally, \( H_{\text{sur}}(3) \) is true.

**Recursion step.** Now, let us suppose that \( k \geq 4 \) and that for all \( i \) such that \( 3 \leq i < k \), \( H_{\text{tar}}(i) \), \( H_{\text{sur}}(i) \), \( H_{\text{sur}}(i) \) and \( H_{E}(i) \) are true. Note that we did not prove \( H_{E}(3) \), which is not even defined, but we don’t need it. Actually, in step \( k \) of the recursion, \( H_{E}(k-1) \) is used only in the proof of \( H_{N}(k) \). For proving \( H_{N}(4) \), the use of \( H_{E}(3) \) is replaced by the use of the definition of \( V_3^{*} \) (Definition 5).

**Proof of \( H_{N}(k) \).** Let \( x \in V_k \). We denote by \( a_1, \ldots, a_l \) the elements of the set \( N_{k-1}(x) \). Let \( i_1, i_2 \in [1, l] \). If \( k \geq 5 \), by hypothesis \( H_{E}(k-1) \), we have that \( N_p(a_{i_1}) = N_p(a_{i_2}) = N_p(x) \) for all \( p \in [0, k-3] \) \( \setminus \{1\} \). If \( k = 4 \), by Definition 5, we have that \( N_0(a_{i_1}) = N_0(a_{i_2}) = N_0(x) \). Thus, independently of the value of \( k \geq 4 \) we have \( N_p(a_{i_1}) = N_p(a_{i_2}) = N_p(x) \) for all \( p \in [0, k-3] \setminus \{1\} \). Then, using the Definition 8 of the characterising sequence, it follows that, for \( p = 0 \), we obtain \( O_1(a_{i_1}) = O_1(a_{i_2}) \) and for \( p \in [2, k-3] \), we obtain \( O_p(a_{i_1}) = O_p(a_{i_2}) = O_p(x) \).

Let \( j \in [2, k-3] \) and \( i \in [1, l] \). From recursion hypothesis \( H_{N}(k-1) \) applied to \( a_i \), we have \( N_j(a_i) = W_j(a_i) \). Since for all \( q \in [1, k-3] \) we have \( O_q(x) = O_q(a_i) \), then, from the definition of \( W_j \), it follows that \( W_j(x) = W_j(a_i) \). Finally, since \( N_j(x) = N_j(a_i) \), we obtain \( N_j(x) = W_j(x) \), for all \( 2 \leq j \leq k-3 \). Then we just have to prove that \( N_{k-2}(x) = W_{k-2}(x) \) and \( N_{k-1}(x) = W_{k-1}(x) \).

We start with \( N_{k-1}(x) \). We will first show that for any \( i \in [1, l] \) we have \( a_i \in W_{k-1}(x) \), that is \( N_{k-1}(x) \subseteq W_{k-1}(x) \). As explained above, we already know that, for all \( q \in [1, k-3] \), \( O_q(x) = O_q(a_i) \). Then, we only need to show that \( O_{k-2}(x) \subseteq O_{k-2}(a_i) \subseteq O_{k-1}(x) \).

Let us show the first inclusion: \( O_{k-2}(x) \subseteq O_{k-2}(a_i) \). By definition, \( N_{k-2}(x) = \bigcap_{i \in [1, l]} N_{k-2}(a_i) \subseteq N_{k-2}(a_i) \). Then, we have \( \bigcap_{b \in N_{k-2}(a_i)} N_1(b) \subseteq \bigcap_{b \in N_{k-2}(x)} N_1(b) \). And since, by definition, \( \bigcap_{b \in N_{k-2}(a_i)} N_1(b) = K(O_{k-2}(a_i)) \) and \( \bigcap_{b \in N_{k-2}(x)} N_1(b) = K(O_{k-2}(x)) \), then we obtain \( K(O_{k-2}(a_i)) \subseteq K(O_{k-2}(x)) \). Thus, from Remark 1, we have \( O_{k-2}(x) \subseteq O_{k-2}(a_i) \).

Let us now show the second inclusion: \( O_{k-2}(a_i) \subseteq O_{k-1}(x) \). Since \( N_{k-2}(a_i) \subseteq \bigcup_{i \in [1, l]} N_{k-2}(a_i) \) then \( \bigcap_{b \in N_{k-2}(a_i)} N_1(b) \subseteq \bigcap_{b \in \bigcup_{i \in [1, l]} N_{k-2}(a_i)} N_1(b) \), and by definition \( \bigcap_{b \in N_{k-2}(a_i)} N_1(b) = K(O_{k-2}(a_i)) \).

In order to show the inclusion we aim at, we will show that \( \bigcap_{b \in \bigcup_{i \in [1, l]} N_{k-2}(a_i)} N_1(b) = \bigcap_{i \in [1, l]} N_1(a_i) \). Since, by definition again, \( \bigcap_{i \in [1, l]} N_1(a_i) = K(O_1(x)) \), this will give \( K(O_{k-2}(a_i)) \subseteq K(O_{k-2}(x)) \), which implies \( O_{k-2}(a_i) \subseteq O_{k-2}(a_i) \), the inclusion we aim at.
Then, let us show the equality $\bigcap_{b \in \bigcup_{t \in [1,l]} N_{k-2}(a_t)} N_1(b) = \bigcap_{t \in [1,l]} N_1(a_t)$ by a double inclusion. Let $z \in \bigcap_{t \in [1,l]} N_1(a_t)$, then for all $t \in [1,l]$, $z \in N_1(a_t)$. It follows that for all $b \in N_{k-2}(a_t)$, $z \in N_1(b)$. Since this holds for all $t \in [1,l]$ and for all $b \in N_{k-2}(a_t)$, then we obtain $z \in \bigcap_{b \in \bigcup_{t \in [1,l]} N_{k-2}(a_t)} N_1(b)$. Conversely, let $z \in \bigcap_{t \in [1,l]} N_1(a_t)$. We then have $z \in N_1(a_t)$, for any $t \in [1,l]$. For all $b \in \bigcup_{t \in [1,l]} N_{k-2}(a_t)$ we have $z \in N_1(b)$. In particular, for any $t \in [1,l]$ and for any $b \in N_{k-2}(a_t)$, we have $z \in N_1(b)$, and so $z \in N_1(a_t)$. As this holds for any $t \in [1,l]$, then $z \in \bigcap_{t \in [1,l]} N_1(a_t)$, that is $\bigcap_{b \in \bigcup_{t \in [1,l]} N_{k-2}(a_t)} N_1(b) \subseteq \bigcap_{t \in [1,l]} N_1(a_t)$. Finally, as we already showed the converse inclusion, we obtain $\bigcap_{b \in \bigcup_{t \in [1,l]} N_{k-2}(a_t)} N_1(b) = \bigcap_{t \in [1,l]} N_1(a_t)$.

We can now finish the proof of the inclusion $N_{k-1}(x) \subseteq W_{k-1}(x)$. Remember that we have $\bigcap_{b \in N_{k-2}(a_t)} N_1(b) \supseteq \bigcap_{b \in \bigcup_{t \in [1,l]} N_{k-2}(a_t)} N_1(b)$ and that by definition $\bigcap_{b \in N_{k-2}(a_t)} N_1(b) = K(O_{k-2}(a_t))$. In addition, we just proved that $\bigcap_{b \in \bigcup_{t \in [1,l]} N_{k-2}(a_t)} N_1(b) = \bigcap_{t \in [1,l]} N_1(a_t)$, and, by definition again, we have $\bigcap_{t \in [1,l]} N_1(a_t) = K(O_{k-1}(x))$. We then obtain $K(O_{k-2}(a_t)) \supseteq K(O_{k-1}(x))$, and Remark 1 concludes that $O_{k-2}(a_t) \subseteq O_{k-1}(x)$, for all $t \in [1,l]$. So finally, putting everything together (we proved above that $O_{k-2}(x) \subseteq O_{k-2}(a_t)$ we get $O_{k-2}(x) \subseteq O_{k-2}(a_t) \subseteq O_{k-1}(x)$, for all $t \in [1,l]$, which completes our proof of $a_i \in W_{k-1}(x)$, for all $t \in [1,l]$, that is $N_{k-1}(x) \subseteq W_{k-1}(x)$.

Let us now prove the converse inclusion: $W_{k-1}(x) \subseteq N_{k-1}(x)$. Let $y \in W_{k-1}(x)$, we have $O_q(y) = O_q(x)$ for all $q \in [1,k-3]$. Moreover, as we showed at the beginning of the proof of $H_N(k)$, for all $i \in [1,l]$ and all $q \in [1,k-3]$ we have $O_q(a_i) = O_q(x)$ and so $O_q(a_i) = O_q(y)$. For $q = 1$, from the definition of $O_1$, this gives $O_0(a_i) = O_0(y)$. For $q \geq 2$ (which occurs only for $k \geq 5$), using recursion hypothesis $H_N(k-1)$ that characterises, for any $z \in V_{k-1}$ and any $q \geq 2$, $N_q(z)$ as a function of only $O_1(z), \ldots, O_q(z)$, we then obtain that since $y$ and $a_i$ have the same sequence $O_1(y), \ldots, O_{q-3}(y) = O_1(a_i), \ldots, O_{q-3}(a_i)$, they necessarily have the same neighbourhood $N_q(y) = N_q(a_i)$, for all $q \in [2,k-3]$. Since we showed that $y$ and $a_i$ also have the same neighbourhood on $V_0$, and since we showed at the beginning of the proof of $H_N(k)$ that $N_p(a_i) = N_p(x)$ for all $p \in [0,k-3] \setminus \{1\}$, it follows that for all $p \in [0,k-3] \setminus \{1\}$, we have $N_p(y) = N_p(x)$. Then, in order to show that $y \in N_{k-1}(x)$, we only need to show that $N_1(x) \subseteq N_1(y)$ and $N_{k-1}(x) \subseteq N_{k-1}(y)$, which implies, by maximality of $x$ (see Definition 1), that $y \in N_{k-1}(x)$. First, let us show that $N_1(x) \subseteq N_1(y)$. Since $y \in W_{k-1}(x)$, we have $O_{k-2}(y) \subseteq O_{k-1}(x)$, which implies $K(O_{k-1}(x)) \subseteq K(O_{k-2}(y))$. And since, from Remark 3, $K(O_{k-1}(x)) = N_1(x)$ and $K(O_{k-2}(y)) = N_1(y)$, then we get the desired inclusion: $N_1(x) \subseteq N_1(y)$.

Let us now show that $N_{k-2}(x) \subseteq N_{k-2}(y)$. Let $z \in N_{k-2}(x)$. By recursion hypothesis $H_N(k-1)$, we know that $N_{k-2}(x) = W_{k-2}(y)$. Thus, our aim is to show that $z \in W_{k-2}(y)$, that is $O_q(z) = O_q(y)$ for all $q \in [1,k-4]$ and $O_{k-3}(y) \subseteq O_{k-3}(z) \subseteq O_{k-2}(y)$. Let $i \in [1,l]$, since $z \in N_{k-2}(x)$ then $z \in N_{k-2}(a_i)$. It follows by recursion hypothesis $H_N(k-1)$ that for all $q \in [1,k-4]$, $O_q(z) = O_q(a_i)$. And since we already showed that $O_q(a_i) = O_q(x)$, then we obtain $O_q(z) = O_q(x)$ for all $q \in [1,k-4]$. On the other hand, since $y \in W_{k-1}(x)$, then $O_q(y) = O_q(x)$ for all $q \in [1,k-4]$, which finally gives that $O_q(y) = O_q(z)$ for all $q \in [1,k-4]$.

Let us now show that $O_{k-3}(z) \subseteq O_{k-2}(y)$. Since $y \in W_{k-1}(x)$, we have $O_{k-2}(x) \subseteq O_{k-2}(y)$. Then, it is sufficient to show that $O_{k-3}(z) \subseteq O_{k-2}(x)$. We state this fact as a proposition as we use it further in the proof:

**Prop. A** for any vertex $z \in N_{k-2}(x)$, we have $O_{k-3}(z) \subseteq O_{k-2}(x)$.

Clearly, since $z \in N_{k-2}(x)$, we have $N_1(z) \supseteq \bigcap_{b \in N_{k-2}(x)} N_1(b)$. By definition, $N_1(z) = K(O_{k-3}(z))$ and $\bigcap_{b \in N_{k-2}(x)} N_1(b) = K(O_{k-2}(x))$. We then obtain $K(O_{k-3}(z)) \supseteq K(O_{k-2}(x))$, which gives, from Remark 1, $O_{k-3}(z) \subseteq O_{k-2}(x)$. And since $y \in W_{k-1}(x)$, we have $O_{k-2}(x) \subseteq W_{k-1}(x)$. Therefore, we have $O_{k-3}(z) \subseteq W_{k-1}(x)$, which completes the proof that $O_{k-3}(z) \subseteq O_{k-2}(x)$.
$O_{k-2}(y)$. Thus, $O_{k-3}(z) \subseteq O_{k-2}(y)$.

We now show that $O_{k-3}(y) \subseteq O_{k-3}(z)$. Since $y \in W_{k-1}(x)$, then $O_{k-3}(y) = O_{k-3}(x)$. As we already showed, for any $i \in [1,l]$, we have $O_{k-3}(x) = O_{k-3}(a_i)$, and so $O_{k-3}(y) = O_{k-3}(a_i)$. Moreover, since $z \in N_{k-2}(x)$ then $z \in N_{k-2}(a_i)$. And by recursion hypothesis $H_K(k-1)$, we have $W_{k-2}(a_i) = N_{k-2}(a_i)$. Thus, $z \in W_{k-2}(a_i)$ satisfies $O_{k-3}(a_i) \subseteq O_{k-3}(z)$. Finally, we obtain $O_{k-3}(y) \subseteq O_{k-3}(z)$, which achieves the proof of $z \in N_{k-2}(y)$. Thus, we have $N_{k-2}(x) \subseteq N_{k-2}(y)$.

In summary, we showed that for any $y \in W_{k-1}(x)$, we have $N_p(y) = N_p(x)$ for all $p \in [0,k-3] \setminus \{1\}$ and $N_1(x) \subseteq N_1(y)$ and $N_{k-2}(x) \subseteq N_{k-2}(y)$. Then, by maximality of $x$ (see Definition 1), $y$ belongs to $N_{k-1}(x)$. That is, $W_{k-1}(x) \subseteq N_{k-1}(x)$. As we also showed the converse inclusion, we obtain $N_{k-1}(x) = W_{k-1}(x)$.

In order to achieve the proof of $H_N(k)$, we still have to show that $N_{k-2}(x) = W_{k-2}(x)$. We start with $W_{k-2}(x) \subseteq N_{k-2}(x)$. Let $z \in W_{k-2}(x)$, let $i \in [1,l]$, we show that $z \in N_{k-2}(a_i)$. By recursion hypothesis $H_N(k-1)$ we know that $N_{k-2}(a_i) = W_{k-2}(a_i) = \{ w \in V_{k-2} | (O_1(w), \ldots, O_{k-4}(w)) = (O_1(a_i), \ldots, O_{k-4}(a_i)) \text{ and } O_{k-2}(a_i) \subseteq O_{k-3}(w) \subseteq O_{k-2}(a_i) \}$. As we already mentioned several times, we have $(O_1(a_i), \ldots, O_{k-4}(a_i)) = (O_1(x), \ldots, O_{k-4}(x))$. Since $z \in W_{k-2}(x)$, we also have $(O_1(z), \ldots, O_{k-4}(z)) = (O_1(x), \ldots, O_{k-4}(x))$, and then $(O_1(z), \ldots, O_{k-4}(z)) = (O_1(a_i), \ldots, O_{k-4}(a_i))$. Again, since $z \in W_{k-2}(x)$, we have $O_{k-3}(a_i) \subseteq O_{k-3}(z) \subseteq O_{k-2}(x)$. Since we already proved that $N_{k-1}(x) = W_{k-1}(x)$, we know that $O_{k-3}(a_i) = O_{k-3}(x)$ and $O_{k-2}(x) \subseteq O_{k-2}(a_i)$. It follows that $O_{k-3}(a_i) \subseteq O_{k-3}(z) \subseteq O_{k-2}(a_i)$, which shows that $z \in N_{k-2}(a_i)$. As this holds for any $i \in [1,l]$, then we conclude that $z \in N_{k-2}(x)$.

Conversely, let $z \in N_{k-2}(x)$. Then, for all $i \in [1,l]$, $z \in N_{k-2}(a_i)$. From recursion hypothesis $H_N(k-1)$ applied to $a_i$, we get $(O_1(z), \ldots, O_{k-4}(z)) = (O_1(a_i), \ldots, O_{k-4}(a_i))$. And for the same reason, we also have $O_{k-3}(a_i) \subseteq O_{k-3}(z)$. As we know that $(O_1(a_i), \ldots, O_{k-4}(a_i)) = (O_1(x), \ldots, O_{k-4}(x))$, we obtain $(O_1(z), \ldots, O_{k-4}(z)) = (O_1(x), \ldots, O_{k-4}(x))$ and $O_{k-3}(a_i) \subseteq O_{k-3}(z)$. Then, the only thing left we have to show in order to prove that $z \in W_{k-2}(x)$, which is our goal, is to prove that $O_{k-3}(z) \subseteq O_{k-2}(x)$. In fact, we already proved this proposition in the proof of $W_{k-1}(x) \subseteq N_{k-1}(x)$ above, refered as (Prop. A) in the text. So we finally obtain that $N_{k-2}(x) \subseteq W_{k-2}(x)$, and since we already proved the converse inclusion, we obtain the equality between the two sets, $N_{k-2}(x) = W_{k-2}(x)$, which completes our proof of $H_N(k)$.

Proof of $H_{\text{tar}}(k)$. From $H_N(k)$ we know that for any $i \in [1,l]$, $(O_1(a_i), \ldots, O_{k-3}(a_i)) = (O_1(x), \ldots, O_{k-3}(x))$. Then, from recursion hypothesis $H_{\text{tar}}(k-1)$, we have $O_1(x) \subseteq \ldots \subseteq O_{k-3}(x)$ and, in particular, we have $O_1(x) \in \mathcal{O}$ and so $|O_1(x)| \geq 2$. Since $N_{k-2}(x) = W_{k-2}(x)$ and $|N_{k-2}(x)| > 1$, necessarily $O_{k-3}(x) \subseteq O_{k-2}(x)$. Similarly, the fact that $N_{k-1}(x) = W_{k-1}(x)$ and $|N_{k-1}(x)| > 1$ implies that $O_{k-2}(x) \subseteq O_{k-1}(x)$. At last, from Remark 3, we have $K(O_{k-1}(x)) = N_1(x)$, and since $|N_1(x)| \geq 2$, it follows that $|K(O_{k-1}(x))| \geq 2$. Combined with the fact that $|O_1(x)| \geq 2$, this implies that for all $j \in [1,k-1]$, we have $O_j(x) \in \mathcal{O}$. Thus, $H_{\text{tar}}(k)$ is true.

Proof of $H_{\text{inj}}(k)$. Let $x, x' \in V_k$ such that $S(x) = S(x')$. From $H_N(k)$, $N_{k-1}(x) = W_{k-1}(x)$ and $N_{k-1}(x') = W_{k-1}(x')$. And since $S(x) = S(x')$, we have $W_{k-1}(x) = W_{k-1}(x')$. As a consequence, $N_{k-1}(x) = N_{k-1}(x')$ and so $x = x'$. Therefore $H_{\text{inj}}(k)$ is true.

Proof of $H_{\text{sur}}(k)$. Let $(O_1, \ldots, O_{k-1}) \in \mathcal{O}^{k-1}$ such that $O_1 \subseteq \ldots \subseteq O_{k-1}$. From recursion hypothesis $H_{\text{sur}}(k-1)$, for any $P \in \mathcal{O}$ such that $O_{k-3} \subseteq P$, there exists $y_P \in V_{k-1}$ such that $S(y_P) = (O_1, \ldots, O_{k-3}, P)$. We denote by $Y$ the set $Y = \{ y \in V_{k-1} | (O_1(y), \ldots, O_{k-3}(y)) = (O_1, \ldots, O_{k-3}) \text{ and } O_{k-2} \subseteq O_{k-2}(y) \subseteq O_{k-1} \}$. Let $x = Y \cup \bigcap_{y \in Y} N(y)$. We will show that $x$ is maximal in $V_k$ and that the corresponding element of $V_k$ has the desired sequence $S(x) =$
(O_1, \ldots, O_{k-1}).

Let us start by showing that \( x \in V_k^* \). Since \( O_{k-2} \subseteq O_{k-1} \), we have \( |Y| \geq 2 \), that is \( |x \cap V_{k-1}| \geq 2 \). From recursion hypothesis \( H_2(k-1) \), for any \( y \in Y \), \( N_{k-2}(y) = \{ t \in V_{k-2} \mid (O_1(t), \ldots, O_{k-4}(t)) = (O_1, \ldots, O_{k-4}) \} \) and \( O_{k-3} \subseteq O_{k-3}(t) \subseteq O_{k-2}(y) \). And since, by definition, \( O_{k-2} \subseteq O_{k-2}(y) \), then \( N_{k-2}(y) \) contains at least the two elements of \( V_{k-2} \) having characterising sequences \( (O_1, \ldots, O_{k-4}, O_{k-3}) \) and \( (O_1, \ldots, O_{k-4}, O_{k-2}) \), which do exist from recursion hypothesis \( H_{sur} \). Since this is true for all \( y \in N_{k-2}(x) \), then \( x \) itself has these two elements as neighbours on level \( V_{k-2} \). Then, \( |x \cap V_{k-2}| \geq 2 \). Let us now show that \( |x \cap V_1| \geq 2 \). For any \( y \in Y \), from Remark 3, we have \( N_1(y) = K(O_{k-2}(y)) \). Since \( O_{k-2}(y) \subseteq O_{k-1} \) then \( K(O_{k-2}(y)) \supseteq K(O_{k-1}) \). Thus, we obtain \( x \cap V_1 = \bigcap_{y \in Y} N_1(y) = \bigcap_{y \in Y} K(O_{k-2}(y)) \supseteq K(O_{k-1}) \). And since \( O_{k-1} \subseteq O \), \( O_{k-1} \) contains at least two elements and so does \( x \cap V_1 \). In order to complete the proof of \( x \in V_k^* \), we need to show that for all \( y, y' \in Y \), we have \( N_{k-3}(y) = N_{k-3}(y') \). First, note that, by definition, \( (O_1(y), \ldots, O_{k-3}(y)) = (O_1(y'), \ldots, O_{k-3}(y')) = (O_1, \ldots, O_{k-3}) \). Moreover, recursion hypothesis \( H_N(k-1) \) gives that the neighbourhood at level \( V_{k-3} \) of any vertex \( z \in V_{k-1} \) only depends on the piece of sequence \( (O_1(z), \ldots, O_{k-3}(z)) \). And since \( y \) and \( y' \) have the same such pieces of sequence, it follows that \( N_{k-3}(y) = N_{k-3}(y') \). Thus \( x \in V_k^* \).

We will now show that \( x \) is maximal in \( V_k^* \). Let \( z \in V_k^* \setminus Y \), we show that for some \( j \in \{ 1, k-3 \} \), \( O_j(z) \neq O_j \) then there is no element \( B \in V_k^* \) containing \( Y \cup \{ z \} \). We denote \( y \in Y \) an arbitrary element of \( Y \) and we distinguish between the case where \( k = 4 \) and the case where \( k \geq 5 \).

Let us start with the general case where \( k \geq 5 \), we show that \( N_{k-3}(z) \neq N_{k-3}(y) \), which implies, from Definition 5, that there is no element of \( V_k^* \) containing both \( y \) and \( z \). So let \( j \in \{ 1, k-3 \} \), such that \( O_j(z) \neq O_j \). Since \( O_j(y) = O_j \), then we have \( O_j(z) \neq O_j(y) \). Moreover, from recursion hypothesis \( H_N(k-1) \), we have \( N_{k-3}(z) = W_{k-3}(z) \) and \( N_{k-3}(y) = W_{k-3}(y) \). We again distinguish several cases depending on the value of \( j \).

If \( j \leq k-5 \) (which may occur only when \( k \geq 6 \)), from recursion hypothesis \( H_{sur}(k-3) \), there exists \( t_1 \in V_{k-3} \) such that \( S(t_1) = (O_1(z), \ldots, O_{k-4}(z)) \). Clearly, from the definition of \( W_j(z) \), we have \( t_1 \in W_{k-3}(z) = N_{k-3}(z) \). On the opposite, from the definition of \( W_j(y) \), and since \( O_j(z) \neq O_j(y) \) with \( j \leq k-5 \), we obtain \( t_1 \not\in W_{k-3}(y) = N_{k-3}(y) \) and it follows that \( N_{k-3}(z) \neq N_{k-3}(y) \).

If \( j = k-4 \). Since \( O_{k-4}(z) \neq O_{k-4}(y) \), then one of the two sets \( O_{k-4}(z), O_{k-4}(y) \) is not included in the other, say \( O_{k-4}(y) \not\subseteq O_{k-4}(z) \) without loss of generality. Consider again the element \( t_1 \in N_{k-3}(z) \) described above. Since \( O_{k-4}(t_1) = O_{k-4}(z) \not\subseteq O_{k-4}(y) \), then it follows, from the definition of \( W_{k-4}(y) \), that \( t_1 \not\in W_{k-3}(y) = N_{k-3}(y) \), and so \( N_{k-3}(z) \neq N_{k-3}(y) \).

If \( j = k-3 \), without loss of generality we can assume that \( O_{k-3}(z) \not\subseteq O_{k-3}(y) \). From recursion hypothesis \( H_{sur}(k-3) \), there exists \( t_2 \in V_{k-3} \) such that \( O_{k-4}(t_2) = O_{k-3}(z) \) and \( (O_1(z), \ldots, O_{k-5}(z)) = (O_1(z), \ldots, O_{k-5}(z)) \) (using, as usual, the convention \( (O_1(z), \ldots, O_{k-5}(z)) = () \) if \( k = 5 \)). From the definition of \( W_{k-3}(z) \) and \( W_{k-3}(y) \), we obtain that \( t_2 \in W_{k-3}(z) = N_{k-3}(z) \) but, since \( O_{k-3}(z) \not\subseteq O_{k-3}(y) \), we have \( t_2 \not\in W_{k-3}(y) = N_{k-3}(y) \). Thus, in all cases where \( k \geq 5 \), if there exists some \( j \in \{ 1, k-3 \} \) such that \( O_j(z) \neq O_j \), then \( N_{k-3}(z) \neq N_{k-3}(y) \).

Let us now deal with the particular case where \( k = 4 \). In this case, necessarily the index \( j \in \{ 1, k-3 \} \) such that \( O_j(z) \neq O_j \) is \( j = 1 \). We immediately obtain that \( N_0(z) = O_1(z) \neq O_1(y) = N_0(y) \). Then, from Definition 5 (case \( k = 4 \), there is no element \( B \in V_k^* \) containing both \( y \) and \( z \). Finally, we conclude that, regardless of the value of \( k \geq 4 \), if for some \( j \in \{ 1, k-3 \} \), \( O_j(z) \neq O_j \) then there is no element \( B \in V_k^* \) containing \( Y \cup \{ z \} \).

Thus, to show that \( x \) is maximal in \( V_k^* \) we only need to show that for any \( z \in V_{k-1} \) such that \( (O_1(z), \ldots, O_{k-3}(z)) = (O_1, \ldots, O_{k-3}) \), if \( O_{k-2} \not\subseteq O_{k-2}(z) \) or \( O_{k-2}(z) \not\subseteq O_{k-1} \), then we have \( \bigcap_{y \in Y} N(y) \not\subseteq N(z) \).
We first treat the case where $O_{k-2}(z) \not\subseteq O_{k-1}$. In this case, from Remark 1, $K(O_{k-1}) \not\subseteq K(O_{k-2}(z))$. From Remark 3, we have $K(O_{k-2}(z)) = N_1(z)$. We now show that $K(O_{k-1}) = \bigcap_{y \in \mathcal{V}_k} N_1(y)$, which will give us the desired result: $\bigcap_{y \in \mathcal{V}_k} N_1(y) \not\subseteq N_1(z)$. From Remark 3, for any $y \in \mathcal{V}_k$, $N_1(y) = K(O_{k-2}(y))$. It follows that $\bigcap_{y \in \mathcal{V}_k} N_1(y) = \bigcap_{y \in \mathcal{V}_k} K(O_{k-2}(y))$ and we also have $\bigcap_{y \in \mathcal{V}_k} K(O_{k-2}(y)) = \bigcap_{y \in \mathcal{P} \subseteq \mathcal{O}} K(P)$, from $H_{\text{sur}}(k-1)$ and the definition of $x$. From Lemma 2, we get $\bigcap_{y \in \mathcal{P} \subseteq \mathcal{O}} K(P) = K(\bigcup_{y \in \mathcal{P} \subseteq \mathcal{O}} K(P))$, which is clearly equal to $K(O_{k-1})$. And so we have $\bigcap_{y \in \mathcal{V}_k} N_1(y) = K(O_{k-1})$. As a consequence, we obtain $\bigcap_{y \in \mathcal{V}_k} N_1(y) \not\subseteq N_1(z)$. Then, adding $z$ to $x \cap \mathcal{V}_k$ would strictly decrease $\bigcap_{y \in \mathcal{V}_k} N_1(y)$.

Let us now consider the case where $O_{k-2} \not\subseteq O_{k-2}(z)$. Using recursion hypothesis $H_N(k-1)$, for all $y \in x \cap \mathcal{V}_k$ we have $N_{k-2}(y) = \{t \in \mathcal{V}_k \mid O_{k-3}(y) \subseteq O_{k-3}(t) \subseteq O_{k-2}(y) \text{ and } (O_1(t), \ldots, O_{k-4}(t)) = (O_1, \ldots, O_{k-4})\}$. Let us denote $Z = \{t \in \mathcal{V}_k \mid \bigcap_{y \in \mathcal{V}_k} O_{k-3}(y) \subseteq O_{k-3}(t) \subseteq \bigcap_{y \in \mathcal{V}_k} O_{k-2}(y) \text{ and } (O_1(t), O_{k-4}(t)) = (O_1, \ldots, O_{k-4})\}$. We show that $\bigcap_{y \in \mathcal{V}_k} N_{k-2}(y) = Z$. Let $t \in \bigcap_{y \in \mathcal{V}_k} N_{k-2}(y)$, then for all $y \in x \cap \mathcal{V}_k$, $O_{k-3}(y) \subseteq O_{k-3}(t) \subseteq O_{k-2}(y)$ and so $\bigcup_{y \in \mathcal{V}_k} O_{k-3}(y) \subseteq O_{k-3}(t) \subseteq \bigcup_{y \in \mathcal{V}_k} O_{k-2}(y)$, that is $t \in Z$. Conversely, if $t \in Z$ then we have $\bigcup_{y \in \mathcal{V}_k} O_{k-3}(y) \subseteq O_{k-3}(t) \subseteq \bigcup_{y \in \mathcal{V}_k} O_{k-2}(y)$ and so $\bigcap_{y \in \mathcal{V}_k} N_{k-2}(y) \subseteq O_{k-2}(y)$ for all $y \in x \cap \mathcal{V}_k$, that is $t \in \bigcap_{y \in \mathcal{V}_k} N_{k-2}(y)$. Thus, $\bigcap_{y \in \mathcal{V}_k} N_{k-2}(y) = Z$. By definition, for all $y \in x \cap \mathcal{V}_k$, $O_{k-3}(y) = O_{k-3}$ and $O_{k-2} \subseteq O_{k-2}(y)$. It follows that $\bigcap_{y \in \mathcal{V}_k} O_{k-3}(y) = O_{k-3}$ and $O_{k-2} \subseteq \bigcap_{y \in \mathcal{V}_k} O_{k-2}(y)$. Moreover, from recursion hypothesis $H_{\text{sur}}$, there exists $t' \in \mathcal{V}_k$ such that $O_{k-3}(t') = O_{k-2}$ and $(O_1(t'), \ldots, O_{k-4}(t')) = (O_1, \ldots, O_{k-4})$. From what precedes, since $O_{k-3} \subseteq O_{k-2}$, we have $N_{k-2}(z) = \{t \in \mathcal{V}_k \mid O_{k-3} \subseteq O_{k-3}(t) \subseteq O_{k-2}(z) \text{ and } (O_1(t), \ldots, O_{k-4}(t)) = (O_1, \ldots, O_{k-4})\}$. And since $O_{k-2} \not\subseteq O_{k-2}(z)$, it follows that $t' \not\in N_{k-2}(z)$, while $t' \in Z = \bigcap_{y \in \mathcal{V}_k} N_{k-2}(y)$ and adding $z$ to $x \cap \mathcal{V}_k$ would strictly decrease $\bigcap_{y \in \mathcal{V}_k} N_{k-2}(y)$. Finally, $x$ is maximal in $\mathcal{V}_k$ and is therefore an element of $\mathcal{V}_k$.

In order to conclude the proof of $H_{\text{sur}}(k)$, let us now show that the element $x$ of $\mathcal{V}_k$ has the desired characterising sequence $(O_1, \ldots, O_{k-1})$. First, from $H_N(k)$, which we already proved, we know that $(O_1(x), \ldots, O_{k-3}(x)) = (O_1(y), \ldots, O_{k-3}(y))$ for any $y \in N_{k-1}(x)$, which gives $(O_1(x), \ldots, O_{k-3}(x)) = (O_1, \ldots, O_{k-3})$, from the definition of $x = Y \bigcap_{y \in \mathcal{N}} N_1(y)$. Second, from Remark 3, we have $K(O_{k-1}(x)) = N_1(x)$ and we already know that $|N_1(z)| \geq 2$ (see beginning of the proof of $H_{\text{sur}}(k)$). This gives $|K(O_{k-1}(x))| \geq 2$ and as we have $|O_1(x)| \geq 2$ and $O_1(x) \subseteq \ldots \subseteq O_{k-2}(x) \subseteq O_{k-1}(x)$, we obtain that $O_{k-2}(x), O_{k-1}(x) \in \mathcal{O}$. Now, from $H_N(k)$, we know that the couple $(O_{k-2}(x), O_{k-1}(x))$ is such that $N_{k-1}(x) = \{y \in \mathcal{V}_k \mid O_{k-2}(x) \subseteq O_{k-2}(y) \subseteq O_{k-1}(x) \text{ and } (O_1(y), \ldots, O_{k-3}(y)) = (O_1, \ldots, O_{k-3})\}$. And by definition of $x$, the couple $(O_{k-2}, O_{k-1})$ also satisfies this condition. But since $O_{k-2}, O_{k-1}, O_{k-2}(x), O_{k-1}(x)$ all belong to $\mathcal{O}$, then, from recursion hypothesis $H_{\text{sur}}(k-1)$, for any $P \in \{O_{k-2}, O_{k-1}, O_{k-2}(x), O_{k-1}(x)\}$ there exists $y \in \mathcal{V}_k$ such that $S(y) = (O_1, \ldots, O_{k-3}, P)$. Then, for $P = O_{k-1}$, using the definition of $N_1(x)$ based on the couple $(O_{k-2}, O_{k-1})$, we obtain that $y \in N_{k-1}(x)$, and consequently, using the definition of $N_{k-1}(x)$ based on the couple $(O_{k-2}(x), O_{k-1}(x))$, we have $O_{k-1} \subseteq O_{k-1}(x)$. Similarly, for $P = O_{k-2}(x)$ we obtain $O_{k-1}(x) \subseteq O_{k-1}$, and it follows that $O_{k-1}(x) = O_{k-1}$. Analogously, choosing $P = O_{k-2}$ and then $P = O_{k-2}(x)$ shows that $O_{k-2}(x) = O_{k-2}$. Thus, $H_{\text{sur}}(k)$ is true.

**Proof of $H_E(k)$**.

Let $k \geq 4$, and let $y_1, y_2 \in \mathcal{V}_k$ such that $N_{k-2}(y_1) = N_{k-2}(y_2)$. We will show that for all $p \in [0, k-2] \setminus \{1\}, N_p(y_1) = N_p(y_2)$. From $H_N(k)$ applied to $y_1$ and $y_2$, we get:
$N_{k-2}(y_1) = W_{k-2}(y_1) = \{ t \in V_{k-2} \mid (O_1(t), \ldots, O_{k-4}(t)) = (O_1(y_1), \ldots, O_{k-4}(y_1)) \text{ and } O_{k-3}(y_1) \subseteq O_{k-3}(t) \subseteq O_{k-2}(y_1) \}$, and $N_{k-2}(y_2) = W_{k-2}(y_2)$. Since $N_{k-2}(y_1) = N_{k-2}(y_2)$, then by considering a common element $t$ of these two sets (which are non empty since $O_{k-2}(y_1) \in \mathcal{O}$), we have $(O_1(y_1), \ldots, O_{k-4}(y_1)) = (O_1(y_2), \ldots, O_{k-4}(y_2))$ (using the usual convention on empty sequences). We now prove that $O_{k-3}(y_1) = O_{k-3}(y_2)$. From recursion hypothesis $H_{\text{sur}}(k)$, we know that there exists $t_1 \in V_{k-2}$ such that $S(t_1) = (O_1(y_1), \ldots, O_{k-3}(y_1))$. This element $t_1$ is clearly an element of $N_{k-2}(y_1)$, and then is an element of $N_{k-2}(y_2)$. Then, we have $O_{k-3}(y_2) \subseteq O_{k-3}(t_1) = O_{k-3}(y_1)$. Symmetrically, by considering an element $t_2 \in V_{k-2}$ such that $S(t_2) = (O_1(y_2), \ldots, O_{k-3}(y_2))$, we obtain $O_{k-3}(y_1) \subseteq O_{k-3}(y_2)$. And finally, we have $O_{k-3}(y_1) = O_{k-3}(y_2)$. Then, regardless of the value of $k \geq 4$, we have $N_0(y_1) = O_1(y_1) = O_1(y_2) = N_0(y_2)$, which is enough to prove $H_E(k)$ when $k = 4$. Let us complete the general case where $k \geq 5$ by considering some $p \in [2, k-3]$ and showing that $N_p(y_1) = N_p(y_2)$. From recursion hypothesis $H_N(k)$, we have $N_p(y_1) = \{ t \in V_p \mid (O_1(t), \ldots, O_{p-2}(t)) = (O_1(y_1), \ldots, O_{p-2}(y_1)) \text{ and } O_{p-1}(y_1) \subseteq O_{p-1}(t) \subseteq O_p(y_1) \}$. And since $p \leq k-3$, $(O_1(y_1), \ldots, O_{p}(y_1)) = (O_1(y_2), \ldots, O_{p}(y_2))$, which implies $N_p(y_1) = N_p(y_2)$. Thus, for all $p \in [0, k-2] \setminus \{1\}$, $N_p(y_1) = N_p(y_2)$.

This shows that $H_E(k)$ is true, which ends the recursion step and the proof of Theorem 1.

The convergence of the series directly follows from the bijection theorem (Theorem 1 above) between the vertices of the multipartite graph and the chains of $\mathcal{L}$.

**Theorem 2 (Convergence theorem)** For any graph $G$, the clean factor series $(G_i)_{i \geq 1}$ generated by $G$ converges.

**Proof:** Theorem 1 states that the characterising sequence $(O_1(x), \ldots, O_{k-1}(x))$ of any node $x$ at level $k$ is such that $O_1(x) \not\subseteq \ldots \not\subseteq O_{k-1}(x)$. The strict inclusions imply that the length of the characterising sequence, which is equal to $k-1$, cannot exceed $h+1$, where $h$ is the height of $\mathcal{L}$. Since $h \leq n-2$, necessarily $V_{n+1}$ is empty. It follows that the clean factor series converges and that the multipartite graph on which it stops has at most $n+1$ levels, that is the upper level has index at most $n$.

### 4 Practical utility of the model

In addition to the fundamental theoretic questions we addressed, our work was motivated by designing a model of complex networks that, while remaining very general, encompass both the local density and the heterogeneous degree distribution of those graphs encountered in practice. In this section, we emphasize on the fact that our model, the clean factor decomposition, is suitable for practical use, with regard to size and time of computation. This allowed us to compute the clean factor decomposition of very large graphs having hundreds of thousands of vertices and billions of edges. These practical results are not presented here (there will be the subject of another paper) since they are beyond the scope of this work. But, in the following, we give theoretic evidence of why the clean factor decomposition is a suitable model to manipulate large real-world instances of graphs, based on common properties of those graphs.

The size of the multipartite graph $M$ obtained at termination of the clean factor series can be exponential in theory, as the number of maximal cliques itself may be exponential. But in practice, its size is quite reasonable and it can be computed efficiently. Indeed, the size of $M$ mainly depends on the complexity of imbrication of maximal cliques, namely on the number of chains of $\mathcal{L}$ (Theorem 1). Theorem 3 below shows that under reasonable hypotheses, this complexity of imbrication is bounded by a constant and the size of $M$ only linearly depends on the number of vertices of $G$. 

[21]
Theorem 3 If every vertex of $G$ is involved in at most $k$ maximal cliques and if every maximal clique of $G$ contains at most $c$ vertices, then we have

$$|V(M)| \leq \min(k2^c c!, 2^k k! + 1) \times n$$

Proof: Thanks to Theorem 1, we obtain an upper bound on $|V(M)|$ by bounding the number of strictly increasing sequences of the form $(O_1, \ldots, O_i)$ such that $O_1, \ldots, O_{i-1} \in \mathcal{O}$.

First, we use the fact that all such sequences are sub-sequences of those obtained starting from a clique $O_i$ and recursively removing one vertex at each step until one obtains a pair $O_1$. The number of such sequences starting with a fixed clique is at most the number of orders on the $c$ vertices of the clique, that is $c!$. And the number of sub-sequences of a sequence of length $c$ is $2^c$. Finally, since each vertex is included in at most $k$ maximal cliques, the number of maximal cliques is at most $kn$. Then, there are at most $k2^c c! n$ increasing sequences made of elements of $\mathcal{O}$, which are in bijection with the vertices of $M$ of level at least 2. Moreover, note that our counting also includes the sequences made of one single maximal clique of $G$ and the sequences made of one single set $\mathcal{O}$ which is a singleton. Those particular sequences are in bijection with the vertices of $M$ at level 1 and 0 respectively. Then, we obtain $|V(M)| \leq k2^c c! n$.

Clearly, since $\mathcal{O} \subseteq \mathcal{O}'$, the number of strictly increasing sequences made of elements of $\mathcal{O}$ is at most the number of strictly increasing sequences made of elements of $\mathcal{O}'$. Another way to count those latter sequences is to count those starting with a fixed minimal set $O_{min} \in \mathcal{O}' \setminus \{\emptyset\}$. Since $\mathcal{O}'$ is closed under intersection, minimal elements of $\mathcal{O}' \setminus \{\emptyset\}$ are pairwise disjoint and therefore their number is at most $n$. The sequences having $O_{min}$ as first set can be formed by starting from a clique containing $O_{min}$ and iteratively intersecting it with another clique containing $O_{min}$. By hypothesis, there are at most $k$ cliques containing a given $O_{min}$, and therefore $k!$ orders on these $k$ cliques. Each order gives rise to a sequence of elements of $\mathcal{O}'$, which contains $2^k$ sub-sequences. Thus, there are at most $k!2^k$ strictly decreasing sequences of elements of $\mathcal{O}'$ having $O_{min}$ as first element. Note that the counting we made actually also comprises the sequences made of one single maximal clique of $G$. Consequently, the number of vertices in $M$ at level at least 1 is at most $k!2^k n$, and adding the $n$ vertices at level 0 to this count we obtain the bound $|V(M)| \leq (2^k k! + 1) \times n$, which completes the proof. \[\blacksquare\]

In practice, parameters $k$ and $c$ are quite small, as they are often constrained by the context where the graphs come from (e.g. social networks, computer networks, citation networks) independently from the size of the graph. Then, the size of $M$ is reasonable in practice, namely $O(n)$ for class of graphs where $k$ and $c$ are bounded. An important consequence is that for those graphs it is possible to compute $M$ in low polynomial time. For example, under those hypothesis, the algorithm of [25] enumerates all maximal cliques of graph $G$ in linear time with regard to the number of maximal cliques, that is $O(n)$ time in this case. Moreover, [26] shows that, for general bipartite graphs, it is possible to enumerate their maximal bicliques in $O(n^2)$ time per biclique (see also [27] for a survey on maximal bicliques enumeration). These facts explain that, in practice, using, as black boxes, the implementation [28] of [29]’s algorithm for enumeration of maximal cliques and the implementation [30] of [31]’s algorithm for enumeration of maximal bicliques, we could compute the clean factor series of graphs with thousands and even hundred of thousands of nodes. Indeed, we did so for a protein interaction network of 1458 vertices and 1948 edges, a movie actors network of 392 340 vertices and 15 038 083 edges, and a piece of the world-wide-web graph of 325 740 nodes and 1 090 108 edges (all can be found at [32]).

This shows that despite the NP-complete nature of the problem of computing the maximal cliques and bicliques, it is possible to compute the clean factor series of graphs encountered in practice, even with a large number of nodes and edges. This makes the clean factor model a very promising tool for modelling complex networks.
5 Conclusion

In this paper, we explored the termination of some multipartite-graph operators closely related to the clique-graph operator. Though the operators we defined may seem very natural, the problem of their termination is non trivial. We showed that the most immediate definition, the weak factor, do not always terminate and we introduced two variations: one of which we could prove termination for (clean-factor graph), the question remaining open for the other (factor graph). Moreover, we showed that the clean-factor series has very strong relationship with the inclusion order $\mathcal{L}$ of the non-trivial intersections of maximal cliques of $G$, which constitutes a striking property of this combinatorial object.

The originality of our work is to explore the possibility of modifying the definition of the operator in order to obtain convergence for all graphs, instead of determining on which graphs it converges. The fact that we could achieve this goal strongly relies on the multipartite nature of the operators we considered: the levels of the multipartite graphs of the series keep track of the history of the previous iterations of the operator. By requiring some additional conditions on the history of the vertices involved in a factorisation step, we showed that we can control the convergence of the series, and still get an object having a rich structure. Moreover, one important feature of our result is that the additional conditions we require involve only a bounded number of history steps. In other words, we need only to keep track of the recent history of factorisation in order to control the convergence of the series.

Many questions arise from our work. The first one is to find minimal restrictions of the weak-factor operator that guarantee termination. In particular we ask whether the factor series always terminates. Indeed, weakening the conditions imposed to force convergence of the series is crucial for modelling purposes, as all the constraints added to obtain the convergence must be respected during the random generation process of the structure, which makes it more complicated.

Another question which seems of great interest to us is the relationship between the classic clique-graph operators and the new operators we introduced here. In particular, one could consider the restriction of the clique-graph operator where one requires the two cliques of $G$ corresponding to the extremities of an edge of the clique graph of $G$ to share at least two vertices (instead of just one): is the convergence of this operator related to those we introduced here? more generally, is it possible to keep track of the factorisation history in the classic clique-graph series and to restrict the clique-graph operator with regard to this history so that we obtain convergence for all graphs?

Finally, the use of multipartite graphs as models of complex networks, in the spirit of the bipartite decomposition [8, 9], asks for several questions. In this context, the key issue is to generate a random multipartite graph while preserving the properties of the original graph. To do so, one has to express the properties to preserve as functions of basic multipartite properties (like degrees, for instance) and to generate random multipartite graphs with these properties. This is a promising direction for complex network modelling, but much remains to be done.

Acknowledgements. We warmly thank Thanh Qui Nguyen and The Hung Tran for helpful discussions on the subject, as well as Clémence Magnien and Stéphan Thomassé for their comments on the writing of the article.

References


URL http://dx.doi.org/10.1007/0-387-22444-0_5


URL http://dx.doi.org/10.1007/BF01793007


URL http://portal.acm.org/citation.cfm?id=1017373.1017385


URL http://dx.doi.org/10.1016/j.dam.2009.01.019


URL http://dx.doi.org/10.1007/BF00338741

24


