

SYMBOLIC METHOD AND DIRECTED GRAPH ENUMERATION

É. DE PANAFIEU AND S. DOVGAL

ABSTRACT. We introduce the arrow product, a systematic generating function technique for directed graph enumeration. It provides short proofs for previous results of Gessel on the number of directed acyclic graphs and of Liskovets, Robinson and Wright on the number of strongly connected directed graphs. We also recover Robinson's enumerative results on directed graphs where all strongly connected components belong to a given family.

1. INTRODUCTION

The enumeration of two important digraph families, the *Directed Acyclic Graphs* (DAGs) and the strongly connected digraphs, has been successfully approached at least since 1969. Apparently, it was Liskovets [9, 10] who first deduced a recurrence for the number of strongly connected digraphs and also introduced and studied the concept of initially connected digraph, a helpful tool for their enumeration. Subsequently, Wright [18] derived a simpler recurrence for strongly connected digraphs and Liskovets [11] extended his techniques to the unlabeled case. Stanley counted labeled DAGs in [17], and Robinson, in his paper [14], counted unlabeled DAGs with a given number of sources, which was the culmination of a series of publications he started in 1970 independently of Stanley. In the unlabeled case, his approach is very much related to the Species Theory [1] which systematises the usage of cycle index series. Robinson also announced [15] a simple combinatorial explanation for the generating function of strongly connected digraphs in terms of the cycle index function. Publications on the exact enumeration of digraphs slowed down, until Gessel [3], in 1995, returned to the problem with a new approach, based on *graphic generating functions*, and Robinson, independently, with a more general approach [16]¹. It allowed them to enumerate DAGs by marking sources and sinks [4] and digraphs by marking source-like and sink-like components. The

Received June 7, 2019.

2010 *Mathematics Subject Classification.* Primary 05C30, 05C20, 05A15.

Élie de Panafieu is a member of Lincs www.lincs.fr.

Sergey Dovgal is supported by the French ANR project MetACOnC, ANR-15-CE40-0014.

The order of the authors is alphabetical.

¹The authors have discovered Robinson's paper [16] after the body of this work was finished and accepted for a publication. We are solving here the same problem with a similar method. We chose to maintain the publication, as we felt those results are of interest for the scientific community, and did not yet received the diffusion they deserve. Our further goal is to integrate these exact enumeration methods into an asymptotic framework in the future.

first English version paper we found containing the elegant expression for the generating function of strongly connected digraphs recalled in Corollary 3.5 is [12]. It points to an earlier publication [11] in Russian, which contains the proof, see also [16].

The symbolic method [1, 2] is a dictionary that translates combinatorial operations into generating function relations. In particular, it allows to manipulate the generating functions directly, avoiding working at the coefficient level. Our contribution is twofold. Firstly, we describe a new operation, the *arrow product* (Definition 2.2), which enriches the symbolic method. Secondly, we propose simple proofs, similar to those of [16], for the generating functions of directed acyclic digraphs (DAGs), strongly connected graphs (SCCs), and digraphs where all SCCs belong to a given family. Some variants are presented as well.

Similar techniques enabled precise description of simple graphs phase transition (see e.g. [7]), so the techniques developed here might enable the study of digraphs phase transition [13].

In this paper, we consider directed graphs (digraphs) with labeled vertices, without loops or multiple edges. Two vertices u, v can be simultaneously linked by both edges $u \rightarrow v$ and $v \rightarrow u$. We also consider simple graphs which are undirected graphs with neither multiple edges nor loops.

2. THE SYMBOLIC APPROACH

2.1. Definitions

Consider a sequence $(a_n(w))_{n=0}^\infty$. Define the *exponential generating function* (EGF) and the *graphic generating function* (GGF) (introduced in [3]) of the sequence $(a_n(w))_{n=0}^\infty$ as

$$A(z, w) := \sum_{n \geq 0} a_n(w) \frac{z^n}{n!} \quad \text{and} \quad \mathbf{A}(z, w) := \sum_{n \geq 0} \frac{a_n(w)}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!}.$$

To distinguish EGF from GGF, the latter are written in bold characters. The *special generating functions* of [16] correspond to GGFs with $w = 1$. The n th coefficient of a series $A(z)$ with respect to the variable z is denoted by $[z^n]A(z)$, so $A(z) = \sum_{n \geq 0} ([x^n]A(x))z^n$.

The *exponential Hadamard product* of two series $A(z) = \sum_{n \geq 0} a_n \frac{z^n}{n!}$ and $B(z) = \sum_{n \geq 0} b_n \frac{z^n}{n!}$ is denoted by and defined as

$$A(z) \odot B(z) = \left(\sum_{n \geq 0} a_n \frac{z^n}{n!} \right) \odot \left(\sum_{n \geq 0} b_n \frac{z^n}{n!} \right) := \sum_{n \geq 0} a_n b_n \frac{z^n}{n!}.$$

All Hadamard products are taken with respect to the variable z . The Hadamard product can be used to convert between EGF and GGF (see Corollary 3.2). The exponential Hadamard product should not be confused with the ordinary Hadamard product $\sum_n ([z^n]A(z))([z^n]B(z))z^n$.

If $\mathcal{A}$ is a certain family of digraphs or graphs, we can associate to it a sequence of series $(a_n(w))_{n=0}^\infty$, such that $[w^m]a_n(w)$ is equal to the number of elements in $\mathcal{A}$ with n vertices and m directed edges. Consequently, we can associate both EGF and GGF to the same family of digraphs or graphs.

An advantage of the symbolic method is its ability to keep track of a collection of *parameters* in combinatorial objects. The two default parameters are the numbers of vertices and edges, and the arguments z and w of a generating function $F(z, w)$ correspond to these parameters. As a generalization, we consider multivariate generating functions

$$A(z, w, \mathbf{u}) := \sum_{n, \mathbf{p}} a_{n, \mathbf{p}}(w) \mathbf{u}^{\mathbf{p}} \frac{z^n}{n!} \quad \text{and} \quad \mathbf{A}(z, w, \mathbf{u}) := \sum_{n, \mathbf{p}} \frac{a_{n, \mathbf{p}}(w)}{(1+w)^{\binom{n}{2}}} \frac{\mathbf{u}^{\mathbf{p}} z^n}{n!},$$

where $\mathbf{u} = (u_1, \dots, u_d)$ is the vector of variables, $\mathbf{p} = (p_1, \dots, p_d)$ denotes a vector of parameters, and the notation $\mathbf{u}^{\mathbf{p}} := \prod_k u_k^{p_k}$ is used. We say that the variable u_k *marks* its corresponding parameter p_k , see [2].

2.2. Combinatorial operations

The next proposition recalls classic operations on EGFs (see [2]), which extend naturally to GGFs.

Proposition 2.1. *Consider two digraph (or graph) families $\mathcal{A}$ and $\mathcal{B}$. The EGF and GGF of the disjoint union of $\mathcal{A}$ and $\mathcal{B}$ are $A(z, w) + B(z, w)$ and $\mathbf{A}(z, w) + \mathbf{B}(z, w)$. The EGF and GGF of the digraphs from $\mathcal{A}$ where one vertex is distinguished are $z\partial_z A(z)$ and $z\partial_z \mathbf{A}(z, w)$. The EGF of sets of digraphs from $\mathcal{A}$ is $e^{A(z, w)}$. The EGF of pairs of digraphs (a, b) with $a \in \mathcal{A}$ and $b \in \mathcal{B}$ (re-labeled so that the vertex labels of a and b are disjoint, see [2]) is $A(z, w)B(z, w)$. If a variable u marks the number of specific items in the EGF $A(z, w, u)$ or the GGF $\mathbf{A}(z, w, u)$ of the family $\mathcal{A}$, then the EGF and GGF for the objects $a \in \mathcal{A}$ which have a distinguished subset of these specific items are $A(z, w, u+1)$ and $\mathbf{A}(z, w, u+1)$. Replacing $u \mapsto u-1$ corresponds to an inclusion-exclusion process.*

The next definition and proposition translate the combinatorial interpretation of the product of GGFs, already mentioned by [16], into the symbolic method framework. Gessel also used it implicitly in several proofs (e.g. [4]) at coefficient level, but did not express it at the generating function level. However, a combinatorial interpretation of the exponential of GGFs can be found in [3, 5].

Definition 2.2. We define the *arrow product* of $\mathcal{A}$ and $\mathcal{B}$ as the family $\mathcal{C}$ of pairs (a, b) , with $a \in \mathcal{A}$, $b \in \mathcal{B}$ (re-labeled so that a and b have disjoint labels), where an arbitrary number of edges oriented from vertices of a to vertices of b are added (see Figure 1).

Proposition 2.3. *The GGF of the arrow product of the families $\mathcal{A}$ and $\mathcal{B}$ is equal to $\mathbf{A}(z, w)\mathbf{B}(z, w)$.*

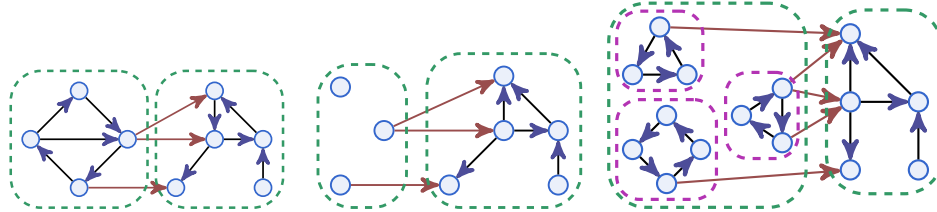


Figure 1. The arrow product. The vertex labels have been omitted.

Figure 2. Symbolic method for DAG.

Figure 3. Marking a subset of source-like SCC.

Proof. Consider two digraph families $\mathcal{A}$ and $\mathcal{B}$, with associated sequences $(a_n(w))$, $(b_n(w))$. Then the sequence associated to the GGF $\mathbf{A}(z, w)\mathbf{B}(z, w)$ is

$$\begin{aligned} c_n(w) &= (1+w)^{\binom{n}{2}} n! [z^n] \left(\sum_k \frac{a_k(w)}{(1+w)^{\binom{k}{2}}} \frac{z^k}{k!} \right) \left(\sum_\ell \frac{b_\ell(w)}{(1+w)^{\binom{\ell}{2}}} \frac{z^\ell}{\ell!} \right) \\ &= \binom{n}{k} \sum_{k+\ell=n} (1+w)^{k\ell} a_k(w) b_\ell(w). \end{aligned}$$

This series has the following combinatorial interpretation: it is the generating function (the variable w marks the edges) of digraphs with n vertices, obtained by

- choosing digraphs a of size k in $\mathcal{A}$, b of size ℓ in $\mathcal{B}$, such that $k + \ell = n$,
- choosing a subset of $\{1, \dots, n\}$ for the labels of a (and b receives the complementary set for its labels),
- for any vertices u in a , v in b , the oriented edge (u, v) is or not added.

Hence, $(c_n(w))$ is the sequence associated to the arrow product of $\mathcal{A}$ and $\mathcal{B}$. $\square$

3. GENERATING FUNCTIONS FROM THE SYMBOLIC METHOD

We start by defining the building bricks for the symbolic method of directed graphs.

Proposition 3.1. *The EGF of all graphs $\mathbf{G}(z, w)$, GGF of all digraphs $\mathbf{D}(z, w)$, and GGF of sets $\mathbf{Set}(z, w)$ (labeled graphs that contain no edge) are*

$$\mathbf{G}(z, w) = \mathbf{D}(z, w) = \sum_{n \geq 0} (1+w)^{\binom{n}{2}} \frac{z^n}{n!} \quad \text{and} \quad \mathbf{Set}(z, w) = \sum_{n \geq 0} \frac{1}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!}.$$

Proof. Consider a graph with n vertices. Each unordered pair of distinct vertices is either linked by an edge, or not. Thus, the sequence of series associated to the family of graphs and its EGF are

$$g_n(w) = (1+w)^{\binom{n}{2}}, \quad \mathbf{G}(z, w) = \sum_{n \geq 0} g_n(w) \frac{z^n}{n!} = \sum_{n \geq 0} (1+w)^{\binom{n}{2}} \frac{z^n}{n!}.$$

In a digraph with n vertices, each ordered pair of distinct vertices is either linked by an oriented edge, or not. So the sequence of series associated to the family of

digraphs and its GGF are

$$d_n(w) = (1+w)^{n(n-1)}, \quad \mathbf{D}(z, w) = \sum_{n \geq 0} \frac{d_n(w)}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!} = \sum_{n \geq 0} (1+w)^{\binom{n}{2}} \frac{z^n}{n!}.$$

There is exactly one labeled graph without any edges, so the sequence of series associated to the set family and its GGF are

$$\text{set}_n(w) = 1, \quad \mathbf{Set}(z, w) = \sum_{n \geq 0} \frac{1}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!}.$$

□

Corollary 3.2. *The EGF and GGF of a family $\mathcal{A}$ are linked by the relations*

$$\mathbf{A}(z, w) = \mathbf{G}(z) \odot \mathbf{A}(z, w) \quad \text{and} \quad \mathbf{A}(z) = \mathbf{Set}(z, w) \odot \mathbf{A}(z, w).$$

Proof. Consider a family $\mathcal{A}$ with sequence of series $(a_n(w))$. By definition of the EGF, GGF and exponential Hadamard product, we have

$$\mathbf{G}(z) \odot \mathbf{A}(z) = \left(\sum_n (1+w)^{\binom{n}{2}} \frac{z^n}{n!} \right) \odot \sum_n \frac{a_n(w)}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!} = \sum_n a_n(w) \frac{z^n}{n!} = \mathbf{A}(z),$$

and similarly

$$\mathbf{Set}(z) \odot \mathbf{A}(z) = \left(\sum_n \frac{1}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!} \right) \odot \sum_n a_n(w) \frac{z^n}{n!} = \sum_n \frac{a_n(w)}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!} = \mathbf{A}(z).$$

□

3.1. Generating functions of various digraph families

The next proposition comes from [4, 14, 17]. We present a proof relying on the arrow product.

Proposition 3.3. *The GGF of directed acyclic graphs (DAGs) with an additional variable u marking the sources (i.e. there are no oriented edge pointing to those vertices) is*

$$\mathbf{DAG}(z, w, u) = \frac{\mathbf{Set}((u-1)z, w)}{\mathbf{Set}(-z, w)}.$$

Proof. The GGF of DAGs where each source is either marked, or left unmarked by the variable u , is $\mathbf{DAG}(z, w, u+1)$ (see Theorem 2.1). Such a DAG is decomposed as the arrow product of a set (the marked sources) with a digraph (Figure 2), so

$$\mathbf{DAG}(z, w, u+1) = \mathbf{Set}(zu, w) \mathbf{DAG}(z, w).$$

Observe that $\mathbf{DAG}(z, w, 0)$ is the GGF of DAGs without any source. The only DAG satisfying this property is the empty DAG, so $\mathbf{DAG}(z, w, 0) = 1$. Taking $u = -1$ gives $1 = \mathbf{Set}(-z, w) \mathbf{DAG}(z, w)$, so $\mathbf{DAG}(z, w) = 1/\mathbf{Set}(-z, w)$. Replacing u with $u-1$ gives $\mathbf{DAG}(z, w, u) = \mathbf{Set}((u-1)z, w)/\mathbf{Set}(-z, w)$. This second proof also illustrates the translation into the generating function world of the inclusion-exclusion principle. □

Let us recall that the *condensation* of a digraph is the directed acyclic graph (DAG) obtained from it by contracting each strongly connected component (SCC) to a vertex. The SCCs of the digraph corresponding to sources of the condensation are called *source-like SCCs*. The proof from Theorem 3.3 for expressing the generating function of DAGs with marked sources is now extended to digraphs with marked source-like components and SCCs belonging to a given family (similar proof published by [16]).

Theorem 3.4. *Consider a nonempty family $\mathcal{A}$ of SCCs (the empty digraph is not strongly connected by convention, so it cannot belong to $\mathcal{A}$). The GGF of digraphs where all SCCs belong to $\mathcal{A}$ is equal to*

$$\mathbf{D}_{\mathcal{A}}(z, w) = \frac{1}{\mathbf{Set}(w, z) \odot e^{-A(z, w)}}.$$

The GGF of the same digraph family where an additional variable u marks the source-like components is

$$\mathbf{D}_{\mathcal{A}}(z, w, u) = \frac{\mathbf{Set}(w, z) \odot e^{(u-1)A(z, w)}}{\mathbf{Set}(w, z) \odot e^{-A(z, w)}}.$$

Proof. The GGF of the digraph family considered, where each source-like component is either marked, or left unmarked by the variable u , is $\mathbf{D}_{\mathcal{A}}(z, w, u + 1)$ (see Theorem 2.1). Such a digraph is decomposed as the arrow product of a set of SCCs from $\mathcal{A}$ (the marked source-like components) with a digraph, so

$$\mathbf{D}_{\mathcal{A}}(z, w, u + 1) = \left(\mathbf{Set}(z, w) \odot e^{uA(z, w)} \right) \mathbf{D}_{\mathcal{A}}(z, w).$$

Taking $u = -1$ gives

$$1 = \left(\mathbf{Set}(z, w) \odot e^{-A(z, w)} \right) \mathbf{D}_{\mathcal{A}}(z, w), \text{ so } \mathbf{D}_{\mathcal{A}}(z, w) = \left(\mathbf{Set}(z, w) \odot e^{-A(z, w)} \right)^{-1}.$$

Replacing u with $u - 1$ gives $\mathbf{D}_{\mathcal{A}}(z, w, u) = \left(\mathbf{Set}(z, w) \odot e^{(u-1)A(z, w)} \right) \mathbf{D}_{\mathcal{A}}(z, w)$. $\square$

When the family $\mathcal{A}$ contains only the SCC with one vertex and no edges, so $A(z, w) = z$, then $\mathbf{D}_{\mathcal{A}}(z, w)$ becomes the GGF of DAGs. Thus, Theorem 3.4 generalizes Theorem 3.3. Several interesting corollaries follow. The first one is our new proof for the EGF of strongly connected digraphs (original result from [11, 12, 16]).

Corollary 3.5. *The exponential generating function of strongly connected digraphs is equal to*

$$\text{SCC}(z, w) = -\log \left(G(z, w) \odot \frac{1}{G(z, w)} \right).$$

Proof. When $\mathcal{A}$ is the family of all SCCs, the first result of Theorem 3.4 becomes

$$\mathbf{D}(z, w) = \frac{1}{\mathbf{Set}(w, z) \odot e^{-\text{SCC}(z, w)}}.$$

By inversion and Hadamard product with $G(z, w)$, we obtain

$$e^{-\text{SCC}(z, w)} = G(z, w) \odot \frac{1}{\mathbf{D}(z, w)}.$$

Replacing $\mathbf{D}(z, w)$ with $G(z, w)$ (see Theorem 3.1) and taking the logarithm gives the final result. $\square$

This formula enables fast computation of the numbers of strongly connected digraphs: $\mathcal{O}(nm \log(n + m))$ arithmetic operations to compute the array of SCCs with at most n vertices and at most m edges, $\mathcal{O}(n \log(n))$ for the SCCs with at most n vertices without edge constraint. The next corollary might prove useful to investigate the birth of the giant SCC in random digraph, following [7].

Corollary 3.6. *Consider a nonempty SCC family $\mathcal{B}$. The GGF of digraphs with a variable u marking the number of SCCs from $\mathcal{B}$ is*

$$\frac{1}{\mathbf{Set}(w, z) \odot e^{(1-u)\mathbf{B}(z, w) - \text{SCC}(z, w)}}.$$

Proof. When $\mathcal{A}$ is the family of all SCCs, with an additional variable u marking the SCCs from $\mathcal{B}$, then $A(z, w, u) = \text{SCC}(z, w) + (u - 1)\mathbf{B}(z, w)$, and the first result of Theorem 3.4 finishes the proof. $\square$

3.2. Initially connected digraphs

Initially connected digraphs are defined as digraphs where any vertex is reachable from the vertex with label 1 via an oriented path. Their analysis has been linked to the study of SCCs, so we provide or recall some results on them for completeness.

Lemma 3.7. *For a given number of vertices and edges, initially connected digraphs with one distinguished vertex are in bijection with digraphs which have a unique source-like component, and where one vertex of that component is distinguished.*

Proof. Let $\mathcal{A}$ and $\mathcal{B}$ denote the two digraph families from the lemma. Consider a digraph $a \in \mathcal{A}$. Since a is initially connected, it contains exactly one source-like SCC. If the distinguished vertex belongs to the source-like SCC, then $a \in \mathcal{B}$. Otherwise, by switching the distinguished vertex with the vertex of label 1, we obtain a digraph from $\mathcal{B}$. Reciprocally, if the distinguished vertex of a digraph $b \in \mathcal{B}$ is in the same SCC as the vertex 1, then $b \in \mathcal{A}$. Otherwise, a digraph from $\mathcal{A}$ is obtained by switching those two vertices. $\square$

The following lemma provides a relation between initially connected digraphs and connected graphs ([8], proof also available in the conclusion of [7]).

Lemma 3.8. *The GGF of initially connected digraphs is equal to the EGF of connected graphs*

$$\mathbf{IC}(z, w) = \mathbf{C}(z, w) = \log(G(z, w)).$$

4. CONCLUSION

Many digraph families can be enumerated using the techniques presented, *e.g.* marking sinks in DAGs and sink-like SCCs in digraphs. The next challenge is SCCs and DAGs asymptotics and, following [7], digraphs phase transition.

Acknowledgment. We would like to thank Cyril Banderier for pointing out the reference [12], and Vlady Ravelomanana for introducing us to the topic and for many fruitful discussions.

REFERENCES

1. Bergeron F., Labelle G. and Leroux P., *Combinatorial Species and Tree-like Structures*, Cambridge University Press vol. 67, 1998.
2. Flajolet P. and Sedgewick R., *Analytic Combinatorics*, Cambridge University Press, 2009.
3. Gessel I. M., *Enumerative applications of a decomposition for graphs and digraphs*, Discrete Math. **139** (1995), 257–271.
4. Gessel I. M., *Counting acyclic digraphs by sources and sinks*, Discrete Math. **160** (1996), 253–258.
5. Gessel I. M. and Sagan B. E., *The Tutte polynomial of a graph, depth-first search, and simplicial complex partitions*, Electron. J. Comb. **3** (1996), #9.
6. Goldschmidt C. and Stephenson R., *The scaling limit of a critical random directed graph*, arXiv:1905.05397.
7. Janson S., Knuth D. E., Luczak T. and Pittel B., *The birth of the giant component*, Random Structures Algorithms **4** (1993), 233–358.
8. Liskovets V. A., *Подсчет корневых инициально связанных ориентированных графов*, Известия АН БССР **5** (1969), 23–32. [Enumeration of rooted initially connected oriented graphs, Izv. Akad. Nauk BSSR]
9. Liskovets V. A., *Об одном рекуррентном методе подсчета графов с отмеченными вершинами*, Доклады Академии наук **184(6)** (1969), 1284–1287 [On one recurrent method of counting graphs with marked vertices, Dokl. Akad. Nauk].
10. Liskovets V. A., *The number of strongly connected directed graphs*, Math. notes of the Academy of Sciences of the USSR **8** (1970), 877–882.
11. Liskovets V. A., *К перечислению сильно связанных ориентированных графов*, ДАН БССР **17** (1973), 1077–1080. [A contribution to the enumeration of strongly connected digraphs, Dokl. Akad. Nauk BSSR]
12. Liskovets V. A., *Some easily derivable integer sequences*, J. Integer Seq. **3** (2000).
13. Luczak T. and Seierstad T. G., *The critical behavior of random digraphs*, Random Structures Algorithms **35** (2009), 271–293.
14. Robinson R. W., *Counting unlabeled acyclic digraphs*, Combinatorial Math. V (1977), 28–43.
15. Robinson R. W., *Counting strong digraphs*, J. Graph Theory **1** (1977), 189–190.
16. Robinson R. W., *Counting digraphs with restrictions on the strong components*, Combin. Graph Theory, Proc. of the Summer School and Conf. **1** (1995), 343–354.
17. Stanley R. P., *Acyclic orientations of graphs*, Discrete Math. **5** (1973), 171–178.
18. Wright E. M., *The number of strong digraphs*, B. Lond. Math. Soc. **3** (1971), 348–350.

É. de Panafieu, Nokia Bell Labs France, Nozay, France,
e-mail: depanafieuelie@gmail.com

S. Dovgal, Laboratoire d'Informatique de Paris Nord, Institut Galilée Université Paris, France,
e-mail: dovgal@lipn.univ-paris13.fr