

Numerical results for generalized attachment models for the genesis of graphs

Jens Gustedt, Pedro Schimit

► **To cite this version:**

Jens Gustedt, Pedro Schimit. Numerical results for generalized attachment models for the genesis of graphs. [Technical Report] RT-0361, INRIA. 2008, 74 p. inria-00349461

HAL Id: inria-00349461

<https://hal.inria.fr/inria-00349461>

Submitted on 31 Dec 2008

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

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



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

*Numerical results for generalized attachment models
for the genesis of graphs*

Jens Gustedt — Pedro Schimit

N° 0361

December 2008

Thème NUM

 *rapport
technique*

Numerical results for generalized attachment models for the genesis of graphs

Jens Gustedt, Pedro Schimit

Thème NUM — Systèmes numériques
Équipe-Projet AlGorille

Rapport technique n° 0361 — December 2008 — 71 pages

Abstract: Using the network generation model from Gustedt [2008], we simulate and analyze the clustering coefficient of the networks as some parameters changes

Key-words: clustering coefficient graph network random generation

Résultats numériques pour la généralisation de modèles d'attachement pour la génèse de graphes

Résumé : En utilisant le modèle de génération de réseaux de Gustedt [2008], nous simulons et analysons le coefficient de clustering de ces réseaux en variant certains de leurs paramètres

Mots-clés : coefficient regroupement graphique réseaux génération aléatoire

1 Introduction and Overview

A model proposed in Gustedt [2008] is used for the random generation of large sparse graphs that allows to produce families of graphs with non-vanishing clustering coefficient. The main objective of this paper is to observe the dependency of the clustering coefficient as some parameters to generate the graphs are changed.

In Section 2, we present the model, some of its variations, and relate it to some well-known random models of networks (*e.g.* Barabási and Albert [1999], Erdős and Rényi [1960]). We compare the models and presentation of the different simulation that were realized and the justify our choices for the genesis of the graphs and related computations, in particular the computation of the clustering coefficient.

In Section 3, all the results of our simulations are presented, analyzed and discussed. Furthermore, a graphical view is used to have another idea of what happens with the graph for the different used parameters. In Section 4, our conclusions are presented.

All figures are given in the Appendix, see page 15 for a table of contents.

2 Model

The model of reference for our genesis of generation of graphs can be described basically with two types of entities: *contexts* and *objects* which are considered a level up of organization for nodes and edges in a graph.

In our graph a observer is a node and a group of observers compounds a context. If a context co have the objects ob_1 , ob_2 and ob_3 , all of them is, by definition, interconnected between them. In the graph, it means that the nodes are $n_1 = ob_1$, $n_2 = ob_2$ and $n_3 = ob_3$, and the edges are: $e_1 = \{ob_1, ob_2\}$, $e_2 = \{ob_1, ob_3\}$, $e_3 = \{ob_2, ob_3\}$. We add and edge $e = \{ob_x, ob_y\} \in G$, where G is the graph, if the objects $ob_x \in Ob$ and $ob_y \in Ob$ are in a common context.

As a basic example cited throughout this paper we will use an interconnection network of which we, as scientists, are all concerned: the graph of co-authorship. In that graph, the authors of a paper are the objects in our model and a paper, which is a set of authors, is a context.

As we already see in this basic example, the implicit structure that we investigate is richer than the just the graph (Ob, G) . In particular we have an important family of implicit objects or co-objects Co which are the scientific papers. Each such paper $co \in Co$ describes the context of a collaboration between a set of colleagues and the relational structure G is derived from them, see Gustedt [2008].

2.1 Objects and their Contexts, Gustedt [2008]

So formally we will investigate pairs (Ob, Co) where Ob is a set (usually finite) and $Co \subseteq 2^{Ob}$ is some family of subsets over Ob . We will say that $ob_1, ob_2 \in Ob$ are *linked* if there is some $co \in Co$ such that $\{ob_1, ob_2\} \subseteq co$. The set of edges, relations or links E is then defined by

$$E = E_{Ob, Co} := \{\{ob_1, ob_2\} | \exists co \in Co, ob_1 \neq ob_2 \subseteq co\}.$$

First, observe that from that definition (Ob, E) has no loops. Second, observe that in this formal definition Ob and Co play “opposite” sites in a bipartite relation that is defined by the containment relation \subseteq . In view of the combinatorial structure the emphasis of Ob being the first set of the pair and Co the second is arbitrary (“just” given by the application). For the example, we could equally well be interested in the relationship among the papers, linking two papers if they share a common author.

These pairs (Ob, Co) are considered as parts of a process the “genesis” of a growing structure, namely we look at sequences $((Ob_\tau, Co_\tau))_{\tau=\dots,0,1,\dots}$ where the parameter τ can be thought of as discretized time, and we have that

$$\begin{aligned} \dots &\subseteq Ob_0 \subseteq Ob_1 \subseteq \dots \\ \dots &\subseteq Co_0 \subseteq Co_1 \subseteq \dots \end{aligned}$$

In terms of the link graph this defines a growing sequence of graphs

$$\dots \subseteq (Ob_0, E_0) \subseteq (Ob_1, E_1) \subseteq \dots$$

What is usually observed in applications is only part of the genesis, *e.g* some or just one of the graphs. The amount of vertices (resp. edges) at time τ are denoted with n_τ and m_τ respectively, *i.e.*

$$\begin{aligned} n_\tau &= |Ob_\tau| \\ m_\tau &= |E_\tau|. \end{aligned}$$

To describe such a genesis we will assume that one step from (Ob_τ, Co_τ) to $(Ob_{\tau+1}, Co_{\tau+1})$ is given by exactly one new context. That is, there is an enumeration of the contexts $\dots, co_0, co_1, co_2, \dots$ such that

$$Ob_\tau = \bigcup_{r \leq \tau} co_r \tag{1}$$

$$Co_\tau = \bigcup_{r \leq \tau} \{co_r\}. \tag{2}$$

The potentially infinite base set of objects and contexts are denoted as

$$Ob_\infty = \bigcup_{r \geq 0} co_r \tag{3}$$

$$Co_\infty = \bigcup_{r \geq 0} \{co_r\}. \tag{4}$$

Generally, we will also suppose that the sequence has no redundancy, *i.e.* that for all τ there are $ob, ob' \in Ob_\tau$ such that $ob, ob' \notin E_{\tau-1}$. For all τ we will denote this set of non-redundant edges $\overline{E}_\tau = E_\tau \setminus E_{\tau-1}$ for which we thus have $\overline{E}_\tau \neq \emptyset$.

Another property that we assume for the sequence is that it respects inclusion in the following sense. $\tau < \kappa$, the new elements that appear in Ob_τ fulfill

$$Ob_\kappa \setminus Ob_{\tau-1} \not\subset Ob_\tau \setminus Ob_{\tau-1}, \quad (5)$$

i.e. no context appearing later than co_τ in the sequence will add less elements to $Ob_{\tau-1}$ than co_τ .

Even with this property the exact ordering of the contexts will be arbitrary. In fact, if $Ob_\tau = Ob_{\tau+1}$ the contexts co_τ and $co_{\tau+1}$ can be considered *interchangeable*. A subsequence $co_\tau, \dots, co_{\tau+l}$ in $(co_i)_{i=0, \dots}$ is *stable* if all adjacent elements are interchangeable, or, in other words $Ob_\tau = \dots = Ob_{\tau+l}$. It is *maximally stable* if it is stable and may not be extended to the left or right without loosing that property. By (5) we then also have that $Ob_\tau \cap Ob_{\tau-1} = \dots = Ob_{\tau+l} \cap Ob_{\tau-1}$.

With that definition we may subdivide our sequence uniquely into maximally stable subsequences. For each τ , $start_\tau$ denotes the start index of the maximal stable subsequence of co_τ and t_τ denotes the number of contexts in that subsequence. Both values are independent of the particular ordering of the subsequence. Also we associate to each such maximal stable subsequence the set of newly introduced objects, $create_\tau = Ob_\tau \setminus Ob_{start_\tau-1}$.

2.2 The generation of the graph

In a genesis as we attempt to describe here, new objects and contexts will emerge from ones that previously exist. Clearly this is only possible if we assume the initial existence of some of them, such that those that are then created may refer to. In a sequence of contexts we will thus assume that there is a finite number \aleph of predefined contexts $co_{(\aleph-1)}, \dots, co_0$. The parameters $n_{<0} = n_1$ and $m_{<0} = m_1$ are thus the amount of vertices and edges that we assume present before the genesis starts, and which we assume to be finite numbers.

In our genesis, the dependence of the process from previous choices is an important detail that we have to handle. We propose a relatively simple model, in which each new $co \in Co$ depends on one previously known other element. In our example of the graph of authorship, a new work often emerges from a previous one by slightly modifying the list of authors, some people cease contributing for the new one, others, such as experts of a particular subdomain or new PhD students join in.

For constructing co_τ , a pre-existing context co_ρ is randomly chosen. Then, the new context is formed by two sets of vertices. The first one, the intersection set, is formed by $co_\tau \cap co_\rho$. This vertices are randomly chosen from co_ρ and added to the new context co_τ ; this set will be referred by S_{i_ρ} .

The second one is formed by the new added vertices (generally these vertices are added inside a stable sequence), and could be represented by $co_\rho \setminus co_\tau$; this set will be referred by S_{n_ρ} .

Now, the type of transformations that are permitted when going from co_τ to co_ρ will be much dependent on the particular domain; different sets of rules will lead to specific families of graphs.

In the following we will introduce some parameters on the sizes of these sets that could describe the evolution in different application domains, either by following some deterministic rule, or just by some statistical correlation. These parameters may then be used to describe an observed sequence or to randomly sample a “typical” member of a specific family.

1. k_τ : the size of $co_\tau = S_{i_\tau} + S_{n_\tau}$;
2. t_τ : the length of the maximal stable sequence containing co_τ as defined above;
3. l_τ : the size (amount of nodes) of the maximal stable sequence.

2.3 Variations and proposed simulations

In this subsection, the different cases, their simulations and the justification for some choices are presented. The main objective is to have a general view of all the experiments realized.

2.3.1 Barabási-Albert, k -trees and its variations

Two well-known models of networks are applied to our model: the attachment model made popular by Barabási and Albert [1999] and k -tree networks.

It is easy to see that if the size of the context is constant $k = 2$, the intersection between an old context and a new one is $S_i = 1$, and the length of the stable sequence is $t = m$ (parameter m in the Barabási-Albert model), our generation process emulates the Barabási-Albert network: In this case, the contexts can be considered as edges because of their size. All contexts have the same probability to be chosen as a starting point for the next new context in the genesis. Thus, the probability that a new vertex w connects with a vertex v that already exists in the graph is proportional to the degree of v .

If we consider the size of the stable sequence with value $t = 1$, the generated graphs are K -trees, with $l = K$ (to not mix up with the size of the context k).

A variation of the K -tree model is when K is not constant, but it varies according to some predefined percentages. The generated graphs then belong to a wide class of graphs, namely *chordal graphs*. This case will be treated below.

2.3.2 Erdős-Rényi

An emulation of the model of Erdős and Rényi [1960] is also done in this work. In this case, we just raffle two different vertices to receive a connection, being an edge.

The total amount of edges is stipulated to be equal to the amount of edges of the Barabási-Albert simulations, *i.e.*, the amount of edges is set in the beginning of the simulation, and at the end of the simulation, the graph contains the same number of edges as one that we would have generated in the Barabási-Albert model.

In this case, the concept of a length of the stable sequence is not applicable. However, we separate the simulation as done in Barabási-Albert simulations, putting labels $l = 1, 2, 3, 4, 5$. But, what we really have is only a difference in the amount of edges.

2.3.3 Networks with random size of the contexts

If we want the sizes of the contexts to vary (as probably present in most applications), we have to choose a distribution for these sizes. We distinguish two different rules to choose new context sizes at random, one that only ensures that the average size of the contexts tends to a prescribed value, and one that prescribes the relative occurrence of each individual size.

In the first case, the size of the new context depends on the average size of all the contexts in the graph. Suppose that the average size is prescribed to be x , a real number. We write this number as an integer $n = \lfloor x \rfloor$ and fractional part $r = x - n$. Then, the probability for the new context to have size $k = n$ is set to $1 - r$, and the probability for size $k = n + 1$ is set to r .

With this kind of distribution is possible to have a well mixed proportion of the two sizes n and $n + 1$. Note that the initial contexts for the genesis of the graph is very important and will determine the family of the possible context sizes (because of the average size of them).

The second case is related to the chordal graphs as mentioned above. In this case, the size of each new context depends on some predefined values of probability for each size, *i.e.* we could have the following values: 50% for $k = 2$, 25% for $k = 3$ and 25% for size $k = 4$. In the end of the genesis, it is expected that 50% of the context have size $k = 2$ and so on for the other values.

For our simulations, we used values from an application graph, namely the co-author graph. The prescribed relative occurrences of context sizes were fixed according to the proportion of papers written by one, two, until ten authors as found in the NCSTRL (Networked Computer Science Technical Reference Library) database analyzed by Newman [2001].

A succession of limitations was discovered at implementation of the algorithm due to the application of the random sizes with length of the stable sequence greater than 1. To really assure random size to the new context at each iteration, sometimes it is necessary change the size of the stable sequence. A pseudo-code for this implementation is presented below (Algorithm 1).

2.3.4 About the simulations

The procedure to a simulation is described in the rest of this subsection. In the case of a fixed size k of the context, we also have to choose

- the size of the stable sequence l , and consequently the size of the intersection between a old and a new context, S_i ,
- the length t of the stable sequence are set,
- the range of the total size of the graph,
- an initial set of contexts.

Algorithm 1: Resolving the size of the new context for random cases

Input: size of the new context: integer K_n , size of the old context: integer K_o , size of the stable sequence: integer t

Output: size of the intersection between the old and the new context: integer S_i , and, if necessary, the amount of nodes to be added in the new context outside the stable sequence: integer t_{add}

Data: old context, size of the new context, stable sequence

```

1 begin
2   if  $K_n > t$  then
3     if  $K_n \geq K_o$  then
4       if  $K_o + t \geq K_n$  then
5          $S_i = K_n - K_o$ 
6       else
7          $S_i = K_o$ 
8          $t_{add} = K_n - K_o - t$ 
9     else
10       $S_i = 1$ 
11   else
12      $S_i = 1$ 
13      $t_{add} = K_n - 1$  (Not all the nodes of stable sequence are possible to add)
14 end

```

In the case of two random context sizes, we have to choose the length and size of the stable sequence. In this case, the initial set of context is very important because it determines the set of two possible sizes.

Concerning the other case of random sizes, in addition to the parameters provided in the previously case, it is necessary to supply the program with the percentage for each size which will have in the graph.

To be able to observe a real evolution of the clustering coefficient for growing sizes of graphs, we want to generate graphs randomly over several orders of magnitudes. It would not be appropriate to draw the size of the graphs uniformly in a given range, since then graphs from the smaller end would occur too rarely. Therefore, in all cases the total number of nodes in the graph is determined as follows: we choose a minimum I_{min} and the maximum I_{max} number and then choose a value α uniformly at random as $\ln(I_{min}) \leq \alpha \leq \ln(I_{max})$. The total amount of nodes is then set to $N = \lfloor e^\alpha \rfloor$.

In the case of fixed sizes for the contexts, it is possible to foresee the number of iterations to achieve the given number of nodes required by $N/(k - S_i)$. In the case of random sizes, when the given number of nodes required is achieved, the simulation stops.

2.3.5 The estimation of the clustering coefficient

Unfortunately, an exact computation of the clustering coefficients of as many graphs as we handle in our experiments would not be feasible. Therefore we

use the approach of Schank and Wagner [2005] to estimate this coefficient. The main idea is here to choose a node with degree greater or equal to 2 at random, then 2 of its neighbors and to verify if these neighbors are connected. This is iterating this n times, and if c connected neighbors have been found, n/c is taken as an estimate of the clustering coefficient. Clearly if $n \rightarrow \infty$ this value is expected to tend to the exact value of the clustering coefficient.

From here, the clustering coefficient will be denoted by CC .

The pseudo-code of the algorithm, adapted from the reference for our case (non-weighted graph), is presented below (Algorithm 2).

Algorithm 2: approximation for CC of Schank and Wagner [2005]

Input: integer k , vector $A_{1,\dots,|V'|}$ of nodes $V' = \{v \in V : d(v) \geq 2\}$
Output: Approximation of CC
Data: Nodes variables: u, v, j , integer variables: l

```

1 begin
2   for  $i \in (1, \dots, k)$  do
3      $j \leftarrow \text{RandomNode}(A)$ 
4      $u \leftarrow \text{RandomAdjacentNode}(A_j)$ 
5     repeat
6        $w \leftarrow \text{RandomAdjacentNode}(A_j)$ 
7     until  $u \neq w$  ;
8     if  $\text{ExistEdge}(u, w)$  then
9        $l \leftarrow l + 1$ 
10  return  $l/k$ 
11 end

```

A problem with this approach was discovered during the simulation: the initial number n_0 of iterations for the clustering coefficient is an important parameter. We had set this number as function of the total number of nodes of the graph.

Basically, the problem arises when a small number of edges contributes to the clustering coefficient. The estimated values then only fall into a discrete set of values and give a false impression of the distribution. This problem happens mainly for the Barabási-Albert ($k = 2$) and Erdős-Rényi models, where the CC tends to zero when the graphs are growing.

Our idea to resolve this problem is to do a function which multiplies the number of edges by a specific value to find the initial number of iterations for CC . The function is shown in Figure 74, and has a form of $f(x) = a.e^{(b.x)} + c.e^{(d.x)}$, although with a linear modification which for number of edges between 10^6 and 10^7 the amount of iteration is 0.5 times the number of edges, and higher than 10^7 , 0.1.

This function was used for the case $k = 2$, $l = 1$ for 10^3 to 10^7 vertices (larger graphs). It could also be used for the another case with $k = 2$, but due to the high time of simulation, we simulated them as a general case (with initial number of iterations equal to 5000).

Another idea could be increase the number of digits for relative precision between the value of CC of the last and the second last of each iteration to refine the final result.

3 Simulations results

In this section, all the results, graphs and analysis are presented. Firstly, a graphic view of the cases simulated are shown, to have an idea about the general behavior of the graph as some parameters changes. Secondly, the clustering coefficient of the many cases are presented. After all, a quick analysis of the time for computation the graph and the clustering coefficient and some general discussions about the results are done.

3.1 A graphic view of the network

Using the software Graphviz-win v2.18, see graphviz, some pictures of the graphs with the range of parameters simulated was done. There are two sets of graphs: The graphs with constant size of the context (and their pictures for $k = 2, 3, 4, 5$ and stable sequence $l = 1, 2, 3, 4, 5$ - for all, $t = 1$) and graphs with random size. For this last case, we have three subsets: the graph with random size between two values (that is shown for $k = 3 - 4$, $k = 4 - 5$ and $k = 5 - 6$; $l = 1, 2, 3, 4, 5$ - for all, $t = 1$), and the random size with predefined proportion of each size (for $l = 1, 2, 3, 4, 5$).

What is viewable in Figures 1 to 8 is basically that as l increases, the number of edges grows, and the graph takes a higher density. For the random cases (Figures 6 to 8) even with $l = 1$, it is possible to view a larger condensation than for the fixed size of the contexts (Figures 1 to 4). Part of this higher condensation is also observable in the value of the clustering coefficient, as we will see in the next section.

The Erdős-Rényi graph is presented only in one figure due to the length of the stable sequence not be applicable for this case. There is only a difference in the amount of edges in the comparison with Barabási-Albert simulations with different values of l .

3.2 The Clustering Coefficient

3.2.1 First step

The first analyzed case is for $k = 2, 3, 4, 5$, $l = 1, 2, 3, 4, 5$ and $t = 1$. Figures from 9 to 28 show the CC in function of the number of nodes of the cases.

For $k = 2$ (emulation of the Barabási-Albert case), as expected, the CC vanishes as the number of nodes increases, even if l grows (it takes low levels slowly). There still are some artefacts in the simulations which seems not to be a property of the graph, but just a side effect of the estimation of the CC. The same problem appears for the larger graphs (next paragraph), but are already less significant due to the use of the function of the Figure 74 which determines the initial number of iterations in function of the number of edges to calculate the CC.

The value $k = 3$ is the first for which the theory from Gustedt [2008] proves a lower bound on CC. When $l = 1$ it is possible to see the CC stabilizing around 0.1 as the number of vertices grows. But for $l = 2$, we have a curve with higher values (but also stabilizes for larger graphs) and for $l = 3$ to $l = 5$, the curve goes down (always stabilizing in a specific value).

The general fact of the CC decreases as l grows can be a result of the boost of new neighbors of the new vertices in the context without a proportional number of new edges to connect the new neighborhood.

For $k = 4$ and $k = 5$, we have the same behavior, except that we do not have the higher curve for $l = 2$.

As described in section 2, we have used the amount of edges found in Barabási-Albert simulations for $l = 1, 2, 3, 4, 5$ for the emulation of Erdős-Rényi model. In fact, we have five cases, which are different only in the number of edges of the graph.

The probability of connection between two vertices is equal to $2.l/(n + 1)$, where n is the number of vertices in the graph, which has a range from 1000 (probability approximately 2^{-3}) to 1000000 (probability approximately 2^{-6}). We do not have a length of stable subsequences, but we have used all the different values of l ($l = 1, 2, 3, 4, 5$) of the Barabási-Albert simulations, which give us five different values for l in the expression of the probability of connection between two vertices.

3.2.2 Larger graphs

Simulations for larger graphs was also done to verify if the results found in the previous are still valid.

There was a limitation for the simulations due to the machine where the program was running. Therefore, the unique case possible to run is those one with length of stable sequence $l = 1$.

It is valid to note the number of iterations used in this case, which is according the function presented at the end of the section 2.

The simulated cases was for the sizes $k = 2, 3, 4, 5$, $l = 1$ with the intersection between a new generated context and the old one which it is connected is $S_i = k - 1$.

In Figures 34 to 37 we have the CC in function of the number of nodes of the graph. As expected, for the case with $k = 2$ (emulation of the Barabási-Albert case) the CC vanishes as the number of nodes grows.

For the other sizes, the CC maintains “almost” constant even for graph larger than 10^6 nodes and it increases as k grows.

3.2.3 Graphs with random size of contexts

In this subsection, it is presented the graphs with random sizes of the contexts, as described in subsection 2.3.3.

Figures 38 to 52 shows the CC in function of the number of vertices in the graph for the cases of random sizes with two possible values ($3 - 4$, $4 - 5$ and $5 - 6$), for $l = 1, 2, 3, 4, 5$. An initial analysis demonstrates a decrease behavior of the CC as the number of vertices in the graph grows.

Figures 53 to 57 mixes the three cases and plots in the same figure the CC in function of the average size of the contexts in the graph, for $l = 1, 2, 3, 4, 5$. There is a viewable increase in the CC as the average size grows, with higher values for larger l .

The simulations of the other case of random sizes (predefined probability of happen a specific size) is illustrated in Figures 58 to 62. What is observable from these graphs is that even for larger graphs, the CC tends to maintain a

constant value. Moreover, the decrease from the initial value (with 1000 nodes) is higher for bigger l , and the curve of CC has larger values as big is l .

3.2.4 Comparing the results

In this section, regressions of the presented cases are done to compare the results for the different simulated parameters in the same figure. For each case, a fitted function for the points is used to plot some cases in just one picture to make easier the visualization of the results as their analysis.

The first case is the constant size of contexts simulated from 10^3 to 10^6 vertices. In Figures 63 to 66 we have the CC for the sizes $k = 2, 3, 4, 5$ and each case has the curves of $l = 1, 2, 3, 4, 5$. It is easily viewable that except for the cases of Barabási-Albert ($k = 2$) and $k = 3, l = 2$ the curve of CC tends to be lower as l increases. Moreover, except for Barabási-Albert, the CC seems to stabilize even for larger graphs.

The comparison between the sizes can be viewable in Figure 68, which has the results of the simulations with graphs from 10^3 to 10^7 , $k = 2, 3, 4, 5$ and $l = 1$. The graphic emphasizes the vanish of CC for Barabási-Albert, otherwise, the cases $k = 3, 4, 5$ maintains the CC even for larger graphs.

The first random size of contexts case analyzed is that one with two possible cases in a graph. Figures 69 to 71 (3 – 4, 4 – 5 and 5 – 6 cases) shows the CC curves for $l = 1, 2, 3, 4, 5$. In all the cases, the CC is relatively high, but decreases as the number of vertices in the graph grows. The curve of CC is also higher for bigger sizes of the context and decreases as the length of the stable sequence decreases. Moreover, for all cases, it is viewable that the curves for higher l tends to a saturated curve which seems that even if we increase l , the curve will not change so much.

Figure 72 mixes the three cases and plots CC in function of the average size of the contexts for $l = 1, 2, 3, 4, 5$. For $l = 1$ and $l = 2$ the CC maintains almost constant as the number of nodes increases. But for $l = 3, 4, 5$, there is a little decrease in the value of CC as the number of nodes grows, stabilizing around 10^5 nodes, which seems to have a little increase until 10^6 .

In the constant size of the context, for $k = 2$, due to the curve formed by the points liken a line in a log-log scale, the function $f(x) = 10^{(b \cdot \log_{10}(x) + c)}$ was used to fit the points. For $k = 3, 4, 5$, due to the line be similar to a exponential decrease, the function $f(x) = 10^{(e^{(c \cdot \log_{10}(x))} + d \cdot \log_{10}(x) + f)}$ (where e is the Euler constant) was used.

For the random case with two possible values for the context, the function which best fitted was $f(x) = 10^{(e^{(c \cdot \log_{10}(x))} + d)}$. For the random case with predefined probability of each size, $f(x) = 10^{(e^{(c \cdot \log_{10}(x))} + d \cdot \log_{10}(x) + f)}$.

For the graph of the clustering coefficient in function of the average size of the context, it was difficult to find a good function, therefore, that one which best fitted the points was

$$f(x) = 10^{(b \cdot \log_{10}(\log_{10}(x)) + c \cdot \log_{10}(x) + d \cdot (1 - e^{(l \cdot \log_{10}(x))}) + f)}.$$

3.2.5 Execution time

To have an idea about the execution time to prepare the graph, and afterward to calculate the CC, some plots are presented.

Figures 75 to 94 show some cases of plots to understand how the parameters of the graph affects the computation of the graph and clustering coefficient. To prepare graph, *i.e.* to generate the graph with required number of nodes, the time grows quickly in function of the number of vertices in the graph, and the curve of the time ascends if l grows (there is more edges for larger l). The size of the context also dislocates the curve up.

Concerning the CC, all the printed cases have a similar behavior in the time to calculate the CC, except the $k = 2$ and Erdős-Rényi cases. In general, for more edges in the graph, more time to calculate the CC is required.

4 Conclusions

In an effort to study a new network generation model, many simulations have been done. The model presented in Section 2 is more complex than some well knows network models in your structure, but is also quite simple to understand to generate larger graphs.

One of the observed point of this model is its capacity to maintain the clustering coefficient near a specific range of values even for larger graphs, approximating natural networks.

Basically, two parameters of the model were analyzed in this report: the size of the context (basic unity of our model defined in Section 2) and the length of the stable sequence to add vertices in the graph (also defined in Section 2). For the size of the context, as larger it is, higher is the clustering coefficient of the network (it vanishes for $k = 2$, as expected, due to be the Barabási-Albert model or to be the Erdős-Rényi model), and for length of the stable sequence, we have a decrease in the clustering coefficient as it grows. As discussed in the text, it can happen because the bigger neighborhood of a vertex when it links with more contexts diffused in the graph.

Another fact reported was the higher values of clustering coefficient with random sizes of the contexts instead of that one with constant sizes. For small graphs, the random size model with two possible values in the graph has a higher clustering coefficient, but for larger graphs, this model can not maintain this value, which decreases, and the model of random size with predefined probability for each size (total: 10 values of sizes) is more successful and has a higher clustering coefficient.

5 Acknowledgments

In this document, we report the work which has been done during the internship financed via the international internship program INRIA, under the supervision of Jens Gustedt.

Pedro Schimit was supported by EGIDE during this research and is grateful for the help and support of Jens Gustedt.

References

- Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tnd. Akad. Mat. Kut. Int. Köz.*, 6:17–61, 1960.
- graphviz. Graphviz - graph visualization software, 2008. URL <http://www.graphviz.org>.
- Jens Gustedt. Generalized attachment models for the genesis of graphs with high clustering coefficient. Rapport de recherche, INRIA, 2008. URL <http://hal.inria.fr/inria-00312059/en/>.
- M. E. J. Newman. Scientific collaboration networks. i. network construction and fundamental results. *Physical Review E*, (64):016131.1–016131.8, 2001.
- Thomas Schank and Dorothea Wagner. Approximating clustering coefficient and transitivity. *J. Graph Algorithms Appl.*, 9(2):265–275, 2005.

Appendix

Contents

1	Introduction and Overview	3
2	Model	3
2.1	Objects and their Contexts, Gustedt [2008]	3
2.2	The generation of the graph	5
2.3	Variations and proposed simulations	6
2.3.1	Barabási-Albert, k -trees and its variations	6
2.3.2	Erdős-Rényi	6
2.3.3	Networks with random size of the contexts	7
2.3.4	About the simulations	7
2.3.5	The estimation of the clustering coefficient	8
3	Simulations results	10
3.1	A graphic view of the network	10
3.2	The Clustering Coefficient	10
3.2.1	First step	10
3.2.2	Larger graphs	11
3.2.3	Graphs with random size of contexts	11
3.2.4	Comparing the results	12
3.2.5	Execution time	12
4	Conclusions	13
5	Acknowledgments	13
A	Generated Graphs	21
B	The Evolution of the Clustering Coefficient	25
B.1	Constant Context Sizes	25
B.2	Varying Context Sizes	40
B.3	Contexts with a Fixed Average Size	49
B.4	Context Sizes with Prescribed Probabilities	52
C	Regressions for the Clustering Coefficient	55
D	Computing the Estimation of the Clustering Coefficient	61
D.1	Number of Iterations	61
D.2	Running Times	62

List of Figures

1	Graphical view of the graphs for $k = 2$	21
2	Graphical view of the graphs for $k = 3$	21
3	Graphical view of the graphs for $k = 4$	22
4	Graphical view of the graphs for $k = 5$	22
5	Graphical view of the graph for the Erdős-Rényi model	23
6	Graphical view of the graphs for random size of the context with $k = 3$ or $k = 4$ and varying l	23
7	Graphical view of the graphs for random size of the context with $k = 4$ or $k = 5$	24
8	Graphical view of the graphs with prescribed context size distri- bution as for the co-author graph, see Newman [2001].	24
9	Clustering coefficient in function of the number of nodes of a graph for $k = 2$, $S_i = 1$ and $l = 1$	25
10	Clustering coefficient in function of the number of nodes of a graph for $k = 2$, $S_i = 1$ and $l = 2$	25
11	Clustering coefficient in function of the number of nodes of a graph for $k = 2$, $S_i = 1$ and $l = 3$	26
12	Clustering coefficient in function of the number of nodes of a graph for $k = 2$, $S_i = 1$ and $l = 4$	26
13	Clustering coefficient in function of the number of nodes of a graph for $k = 2$, $S_i = 1$ and $l = 5$	27
14	Clustering coefficient in function of the number of nodes of a graph for $k = 3$, $S_i = 2$ and $l = 1$	27
15	Clustering coefficient in function of the number of nodes of a graph for $k = 3$, $S_i = 2$ and $l = 2$	28
16	Clustering coefficient in function of the number of nodes of a graph for $k = 3$, $S_i = 2$ and $l = 3$	28
17	Clustering coefficient in function of the number of nodes of a graph for $k = 3$, $S_i = 2$ and $l = 4$	29
18	Clustering coefficient in function of the number of nodes of a graph for $k = 3$, $S_i = 2$ and $l = 5$	29
19	Clustering coefficient in function of the number of nodes of a graph for $k = 4$, $S_i = 3$ and $l = 1$	30
20	Clustering coefficient in function of the number of nodes of a graph for $k = 4$, $S_i = 3$ and $l = 2$	30
21	Clustering coefficient in function of the number of nodes of a graph for $k = 4$, $S_i = 3$ and $l = 3$	31
22	Clustering coefficient in function of the number of nodes of a graph for $k = 4$, $S_i = 3$ and $l = 4$	31
23	Clustering coefficient in function of the number of nodes of a graph for $k = 4$, $S_i = 3$ and $l = 5$	32
24	Clustering coefficient in function of the number of nodes of a graph for $k = 5$, $S_i = 4$ and $l = 1$	32
25	Clustering coefficient in function of the number of nodes of a graph for $k = 5$, $S_i = 4$ and $l = 2$	33
26	Clustering coefficient in function of the number of nodes of a graph for $k = 5$, $S_i = 4$ and $l = 3$	33

27	Clustering coefficient in function of the number of nodes of a graph for $k = 5$, $S_i = 4$ and $l = 4$	34
28	Clustering coefficient in function of the number of nodes of a graph for $k = 5$, $S_i = 4$ and $l = 5$	34
29	Clustering coefficient in function of the number of nodes of a graph for Erdős-Rényi model with the same number of edges of the case Barabási-Albert with $l = 1$	35
30	Clustering coefficient in function of the number of nodes of a graph for Erdős-Rényi model with the same number of edges of the case Barabási-Albert with $l = 2$	35
31	Clustering coefficient in function of the number of nodes of a graph for Erdős-Rényi model with the same number of edges of the case Barabási-Albert with $l = 3$	36
32	Clustering coefficient in function of the number of nodes of a graph for Erdős-Rényi model with the same number of edges of the case Barabási-Albert with $l = 4$	36
33	Clustering coefficient in function of the number of nodes of a graph for Erdős-Rényi model with the same number of edges of the case Barabási-Albert with $l = 5$	37
34	Clustering coefficient in function of the number of nodes of a graph for $k = 2$, $S_i = 1$ and $l = 1$	38
35	Clustering coefficient in function of the number of nodes of a graph for $k = 3$, $S_i = 2$ and $l = 1$	38
36	Clustering coefficient in function of the number of nodes of a graph for $k = 4$, $S_i = 3$ and $l = 1$	39
37	Clustering coefficient in function of the number of nodes of a graph for $k = 5$, $S_i = 4$ and $l = 1$	39
38	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 3 and 4 and $l = 1$	40
39	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 3 and 4 and $l = 2$	40
40	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 3 and 4 and $l = 3$	41
41	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 3 and 4 and $l = 4$	41
42	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 3 and 4 and $l = 5$	42
43	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 4 and 5 and $l = 1$	43
44	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 4 and 5 and $l = 2$	43
45	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 4 and 5 and $l = 3$	44
46	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 4 and 5 and $l = 4$	44
47	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 4 and 5 and $l = 5$	45
48	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 5 and 6 and $l = 1$	46

49	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 5 and 6 and $l = 2$	46
50	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 5 and 6 and $l = 3$	47
51	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 5 and 6 and $l = 4$	47
52	Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 5 and 6 and $l = 5$	48
53	Clustering coefficient in function of the average size of the contexts for random size (mixed the three cases) with $l = 1$ 49
54	Clustering coefficient in function of the average size of the contexts for random size (mixed the three cases) with $l = 2$ 49
55	Clustering coefficient in function of the average size of the contexts for random size (mixed the three cases) with $l = 3$ 50
56	Clustering coefficient in function of the average size of the contexts for random size (mixed the three cases) with $l = 4$ 50
57	Clustering coefficient in function of the average size of the contexts for random size (mixed the three cases) with $l = 5$ 51
58	Clustering coefficient in function of the number of nodes of a graph for random size of context according to prescribed probability of each size and $l = 1$ 52
59	Clustering coefficient in function of the number of nodes of a graph for random size of context according to prescribed probability of each size and $l = 2$ 52
60	Clustering coefficient in function of the number of nodes of a graph for random size of context according to prescribed probability of each size and $l = 3$ 53
61	Clustering coefficient in function of the number of nodes of a graph for random size of context according to prescribed probability of each size and $l = 4$ 53
62	Clustering coefficient in function of the number of nodes of a graph for random size of context according to prescribed probability of each size and $l = 5$ 54
63	Regression of the clustering coefficient in function of the number of nodes of a graph for the case $k = 2$, $S_i = 1$ and $l = 1, 2, 3, 4, 5$	55
64	Regression of the clustering coefficient in function of the number of nodes of a graph for the case $k = 3$, $S_i = 2$ and $l = 1, 2, 3, 4, 5$	55
65	Regression of the clustering coefficient in function of the number of nodes of a graph for the case $k = 4$, $S_i = 3$ and $l = 1, 2, 3, 4, 5$	56
66	Regression of the clustering coefficient in function of the number of nodes of a graph for the case $k = 5$, $S_i = 4$ and $l = 1, 2, 3, 4, 5$	56
67	Regression of the clustering coefficient in function of the number of nodes of a graph for Erdős-Rényi model with the number of edges equal to the number of edges of the Barabási-Albert simulations from $l = 1$ to $l = 5$ 57
68	Regression of the clustering coefficient in function of the number of nodes of a graph for the cases $k = 2/S_i = 1$, $k = 3/S_i = 2$, $k = 4/S_i = 3$, $k = 5/S_i = 4$ and $l = 1$ - larger graph 58

69	Regression of the clustering coefficient in function of the number of nodes of a graph for random size of context varying between 3 and 4 and $l = 1, 2, 3, 4, 5$	58
70	Regression of the clustering coefficient in function of the number of nodes of a graph for random size of context varying between 4 and 5 and $l = 1, 2, 3, 4, 5$	59
71	Regression of the clustering coefficient in function of the number of nodes of a graph for random size of context varying between 5 and 6 and $l = 1, 2, 3, 4, 5$	59
72	Regressions of the clustering coefficient in function of the average size of the contexts for random size (mixed the three cases) and $l = 1, 2, 3, 4, 5$	60
73	Regression of the clustering coefficient in function of the number of nodes of a graph according to prescribed probability of each size and $l = 1, 2, 3, 4, 5$	60
74	Factor of multiplication to the number of edges of a graph to obtain the initial amount of iterations for the calculation of the clustering coefficient of a graph	61
75	Execution time to prepare the graph in the constant size of the context case for $k = 2, S_i = 1$ and $l = 1$	62
76	Execution time to prepare the graph in the constant size of the context case for $k = 2, S_i = 1$ and $l = 5$	62
77	Execution time to prepare the graph in the constant size of the context case for $k = 4, S_i = 3$ and $l = 1$	63
78	Execution time to prepare the graph in the constant size of the context case for $k = 4, S_i = 3$ and $l = 5$	63
79	Execution time to prepare the graph in the Erdős-Rényi case for the number of edges equal to the number of edges of the Barabási-Albert simulations with $l = 1$	64
80	Execution time to prepare the graph in the Erdős-Rényi case for the number of edges equal to the number of edges of the Barabási-Albert simulations with $l = 5$	64
81	Execution time to prepare the graph in the random size of the context case with two possible values (4 and 5) for $l = 1$	65
82	Execution time to prepare the graph in the random size of the context case with two possible values (4 and 5) for $l = 5$	65
83	Execution time to prepare the graph in the case with prescribed probability of each size of context for $l = 1$	66
84	Execution time to prepare the graph in the case with prescribed probability of each size of context for $l = 5$	66
85	Execution time to calculate the clustering coefficient in the constant size of the context case for $k = 2, S_i = 1$ and $l = 1$	67
86	Execution time to calculate the clustering coefficient in the constant size of the context case for $k = 2, S_i = 1$ and $l = 5$	67
87	Execution time to calculate the clustering coefficient in the constant size of the context case for $k = 4, S_i = 3$ and $l = 1$	68
88	Execution time to calculate the clustering coefficient in the constant size of the context case for $k = 4, S_i = 3$ and $l = 5$	68

89	Execution time to calculate the clustering coefficient in the Erdős-Rényi case for the number of edges equal to the number of edges of the Barabási-Albert simulations with $l = 1$	69
90	Execution time to calculate the clustering coefficient in the Erdős-Rényi case for the number of edges equal to the number of edges of the Barabási-Albert simulations with $l = 5$	69
91	Execution time to calculate the clustering coefficient in the random size of the context case with two possible values (4 and 5) for $l = 1$	70
92	Execution time to calculate the clustering coefficient in the random size of the context case with two possible values (4 and 5) for $l = 5$	70
93	Execution time to calculate the clustering coefficient in the case with prescribed probability of each size of context for $l = 1$. . .	71
94	Execution time to calculate the clustering coefficient in the case with prescribed probability of each size of context for $l = 5$. . .	71

A Generated Graphs

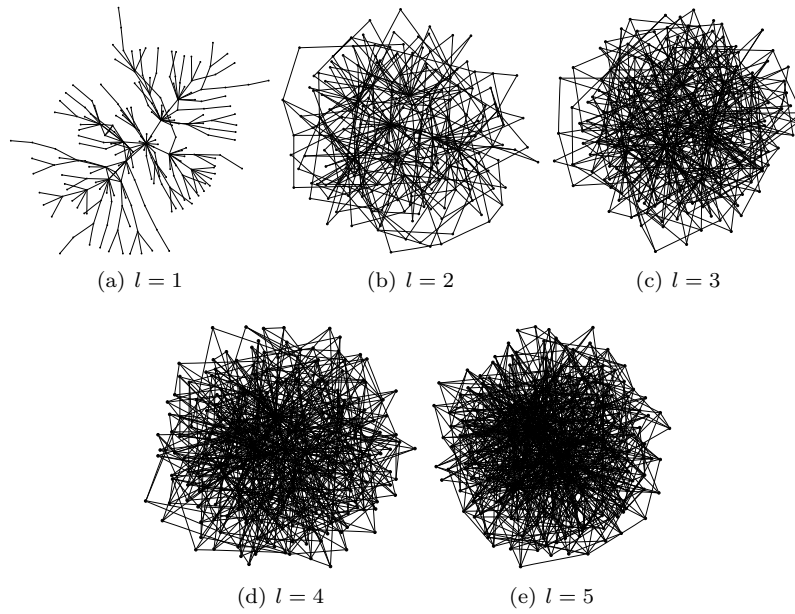


Figure 1: Graphical view of the graphs for $k = 2$

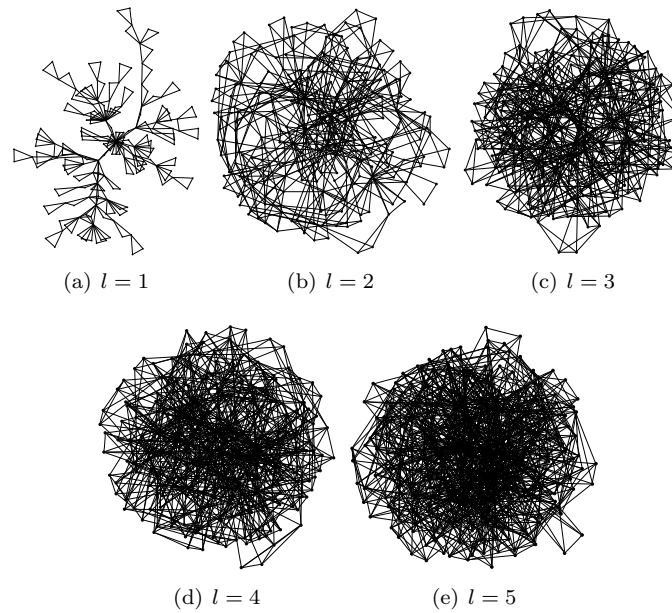
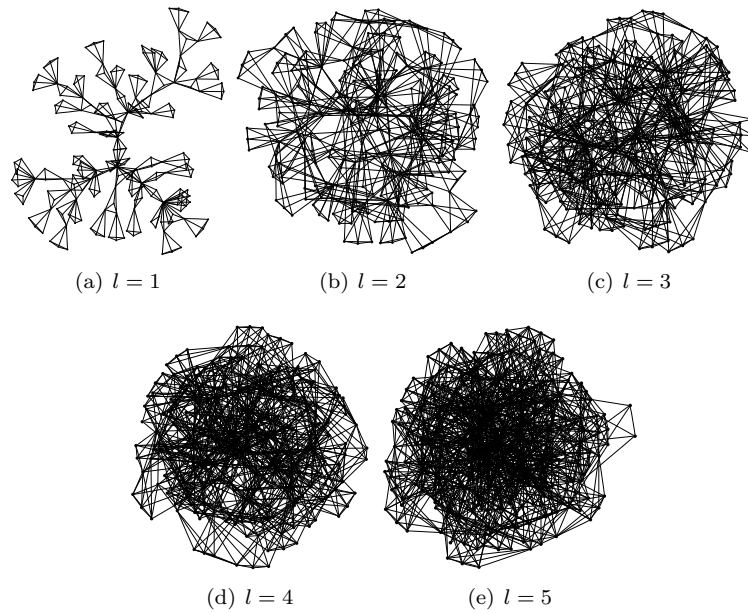
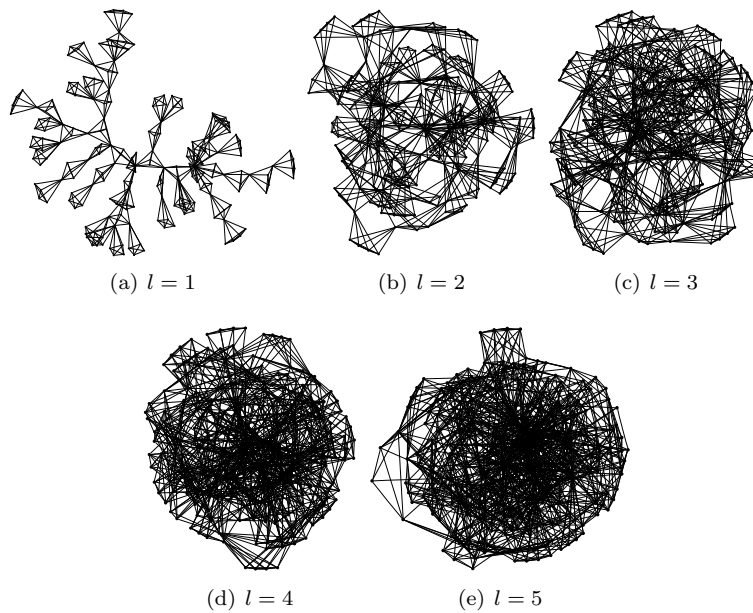


Figure 2: Graphical view of the graphs for $k = 3$

Figure 3: Graphical view of the graphs for $k = 4$ Figure 4: Graphical view of the graphs for $k = 5$

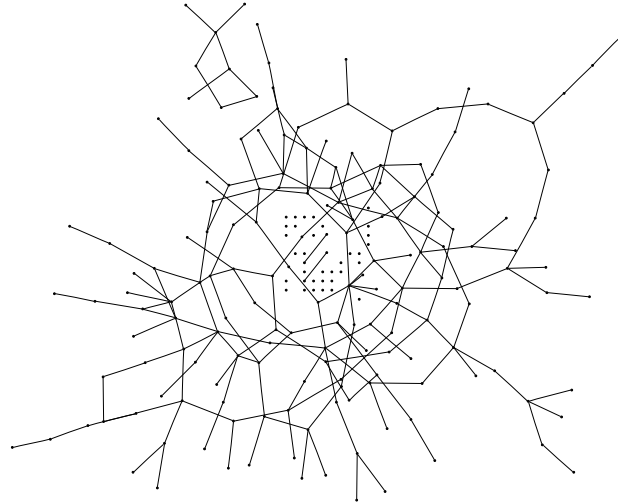


Figure 5: Graphical view of the graph for the Erdős-Rényi model

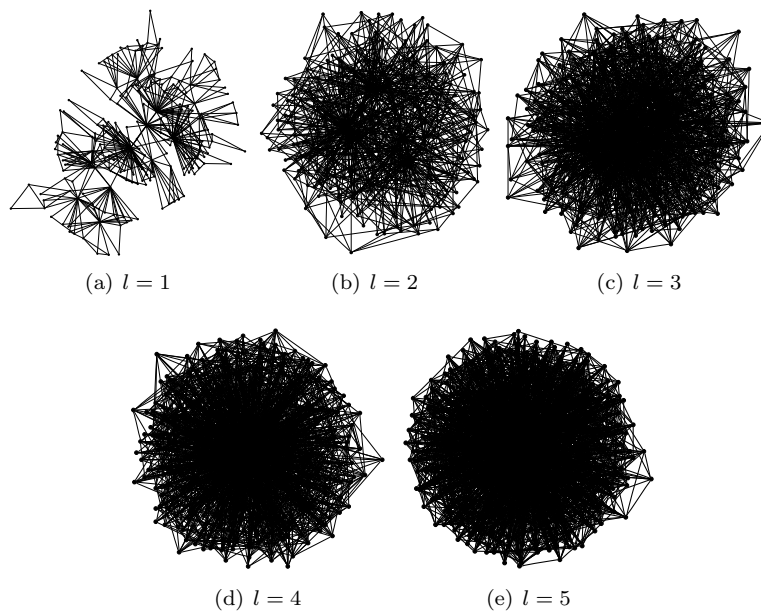


Figure 6: Graphical view of the graphs for random size of the context with $k = 3$ or $k = 4$ and varying l

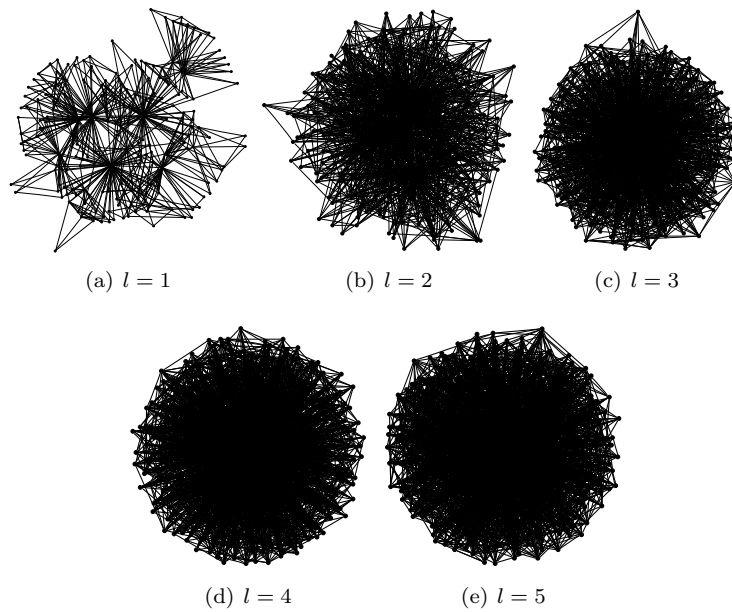


Figure 7: Graphical view of the graphs for random size of the context with $k = 4$ or $k = 5$

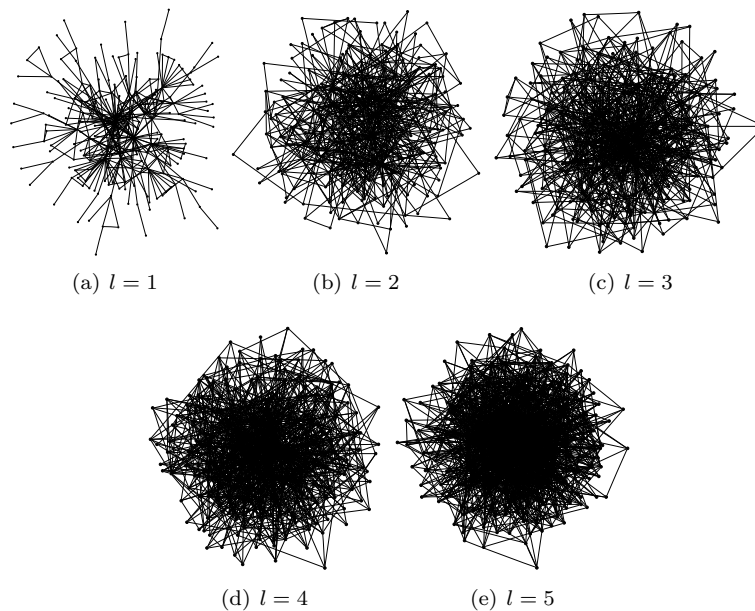


Figure 8: Graphical view of the graphs with prescribed context size distribution as for the co-author graph, see Newman [2001].

B The Evolution of the Clustering Coefficient

B.1 Constant Context Sizes

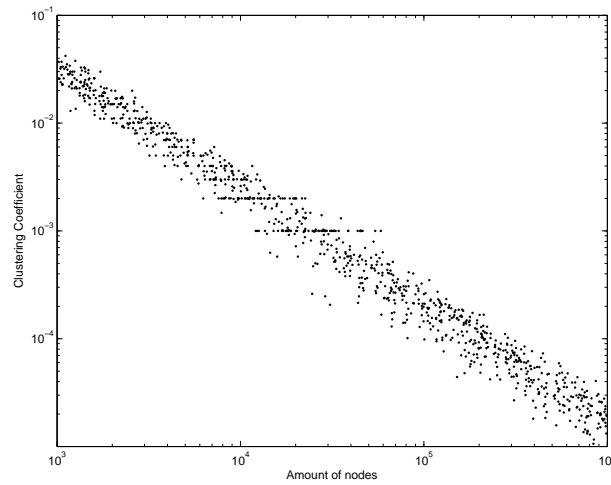


Figure 9: Clustering coefficient in function of the number of nodes of a graph for $k = 2$, $S_i = 1$ and $l = 1$

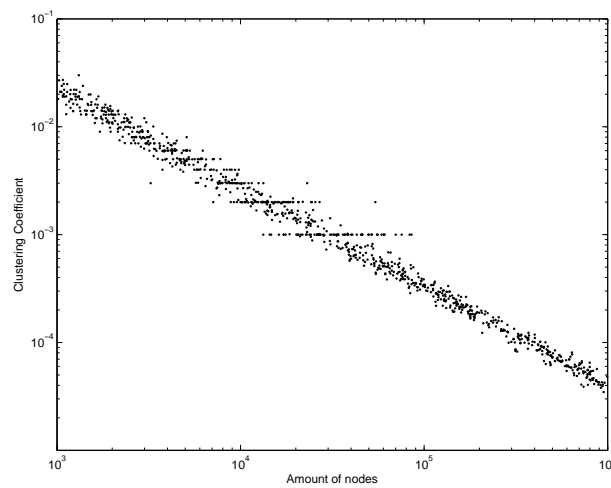


Figure 10: Clustering coefficient in function of the number of nodes of a graph for $k = 2$, $S_i = 1$ and $l = 2$

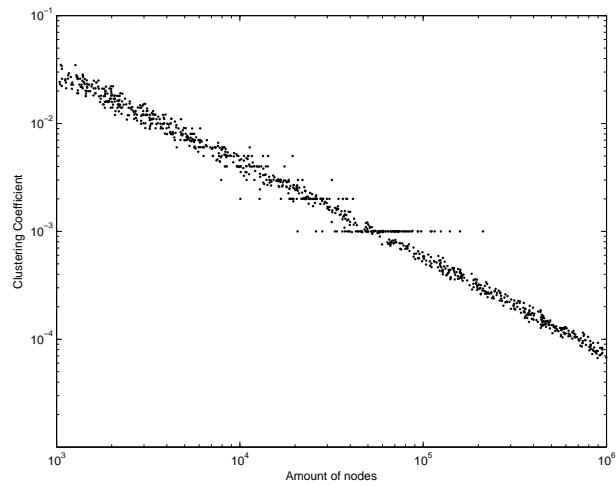


Figure 11: Clustering coefficient in function of the number of nodes of a graph for $k = 2$, $S_i = 1$ and $l = 3$

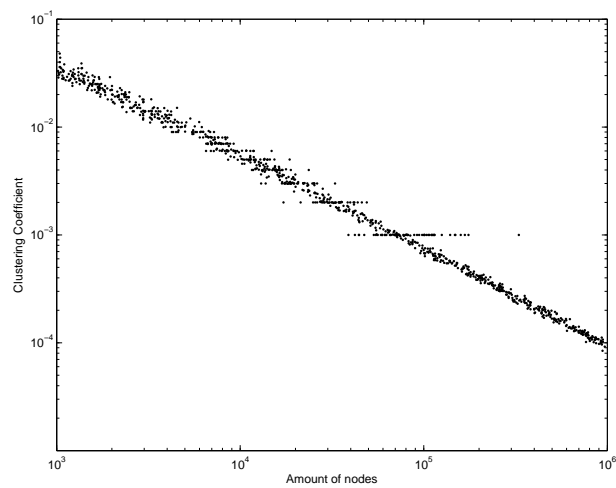


Figure 12: Clustering coefficient in function of the number of nodes of a graph for $k = 2$, $S_i = 1$ and $l = 4$

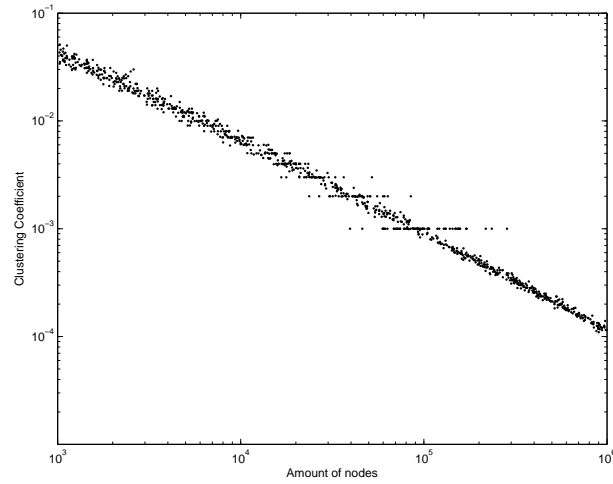


Figure 13: Clustering coefficient in function of the number of nodes of a graph for $k = 2$, $S_i = 1$ and $l = 5$

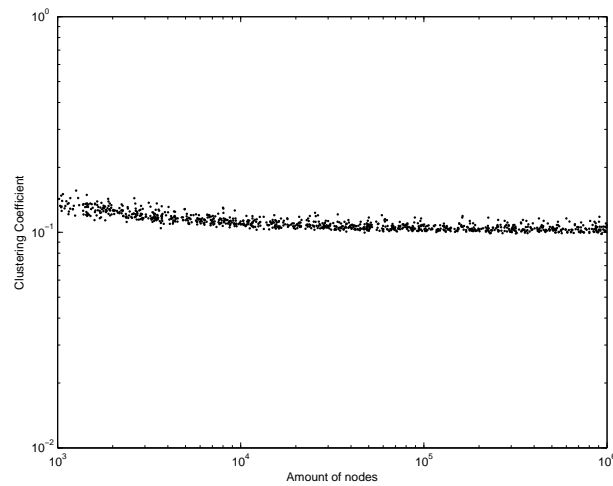


Figure 14: Clustering coefficient in function of the number of nodes of a graph for $k = 3$, $S_i = 2$ and $l = 1$

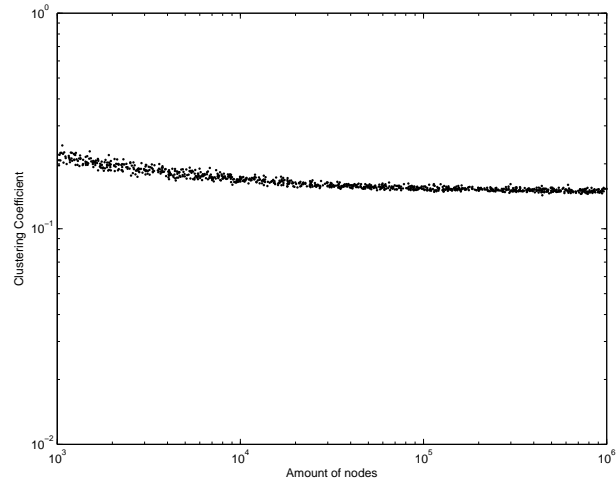


Figure 15: Clustering coefficient in function of the number of nodes of a graph for $k = 3$, $S_i = 2$ and $l = 2$

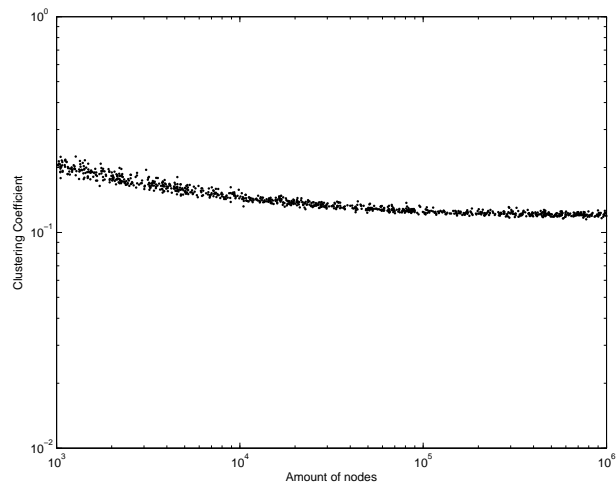


Figure 16: Clustering coefficient in function of the number of nodes of a graph for $k = 3$, $S_i = 2$ and $l = 3$

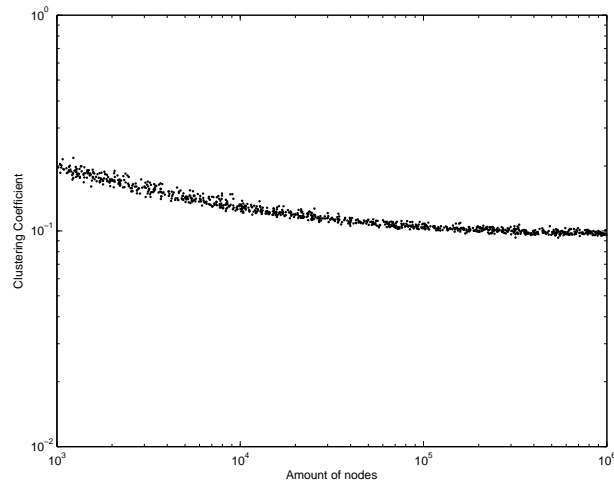


Figure 17: Clustering coefficient in function of the number of nodes of a graph for $k = 3$, $S_i = 2$ and $l = 4$

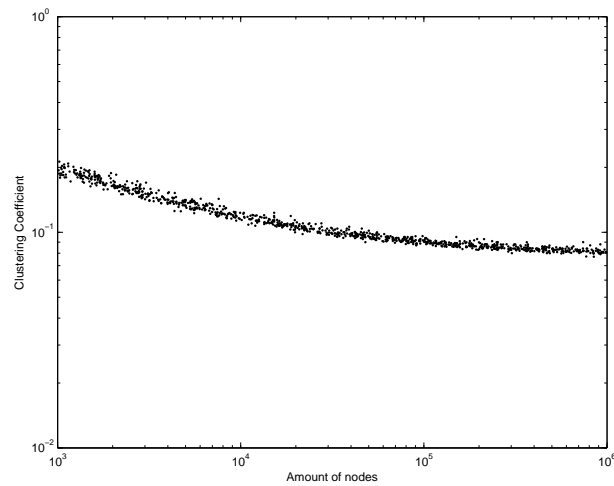


Figure 18: Clustering coefficient in function of the number of nodes of a graph for $k = 3$, $S_i = 2$ and $l = 5$

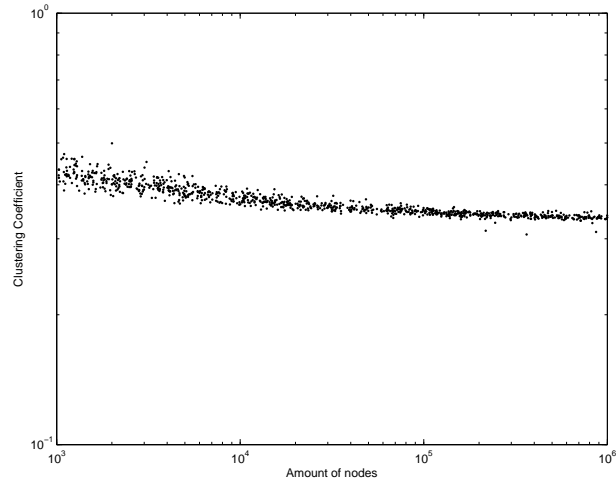


Figure 19: Clustering coefficient in function of the number of nodes of a graph for $k = 4$, $S_i = 3$ and $l = 1$

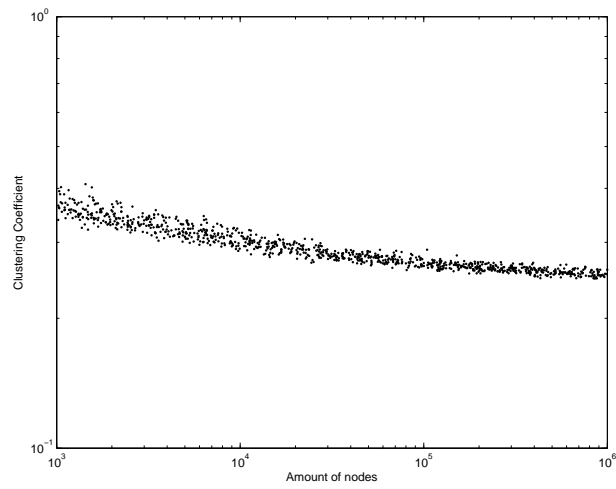


Figure 20: Clustering coefficient in function of the number of nodes of a graph for $k = 4$, $S_i = 3$ and $l = 2$

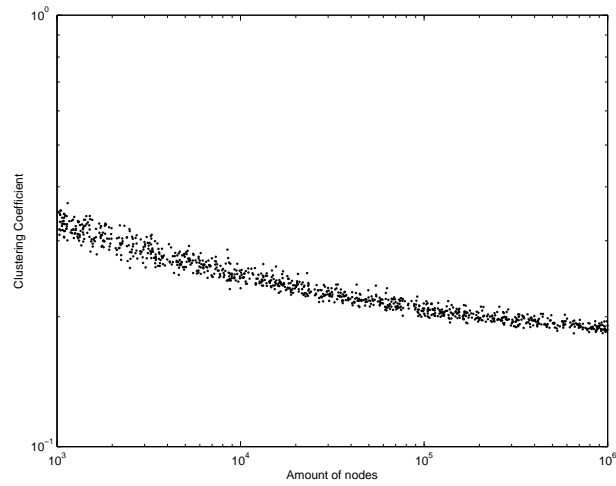


Figure 21: Clustering coefficient in function of the number of nodes of a graph for $k = 4$, $S_i = 3$ and $l = 3$

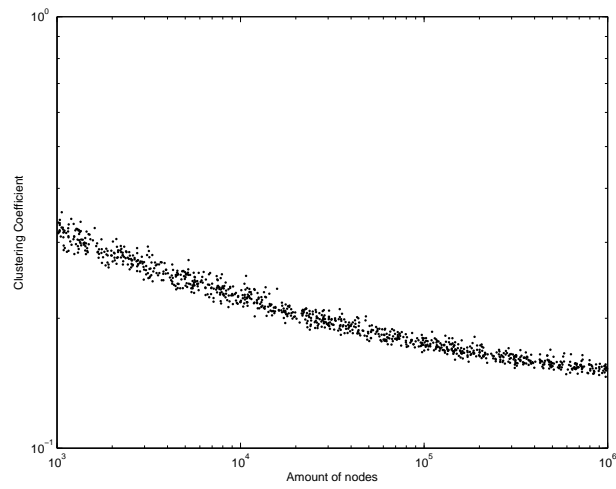


Figure 22: Clustering coefficient in function of the number of nodes of a graph for $k = 4$, $S_i = 3$ and $l = 4$

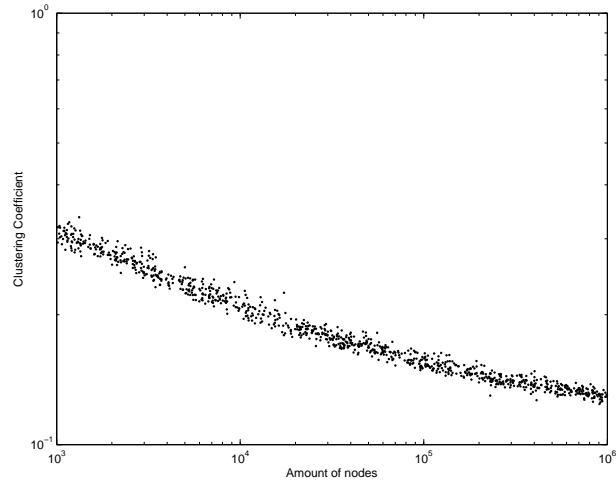


Figure 23: Clustering coefficient in function of the number of nodes of a graph for $k = 4$, $S_i = 3$ and $l = 5$

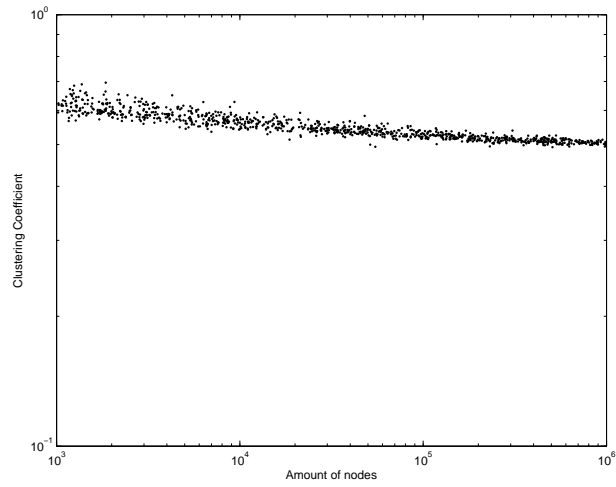


Figure 24: Clustering coefficient in function of the number of nodes of a graph for $k = 5$, $S_i = 4$ and $l = 1$

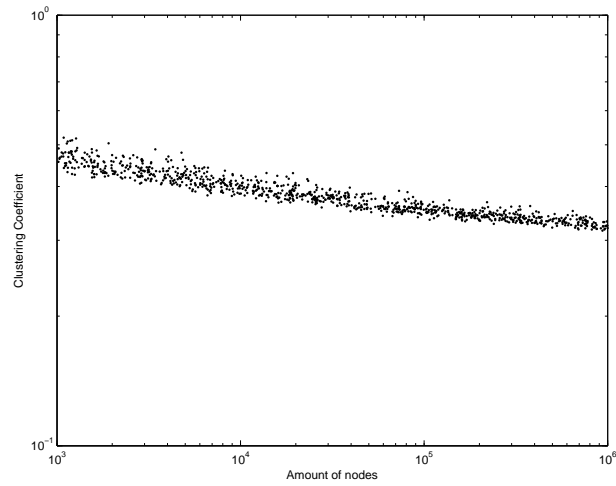


Figure 25: Clustering coefficient in function of the number of nodes of a graph for $k = 5$, $S_i = 4$ and $l = 2$

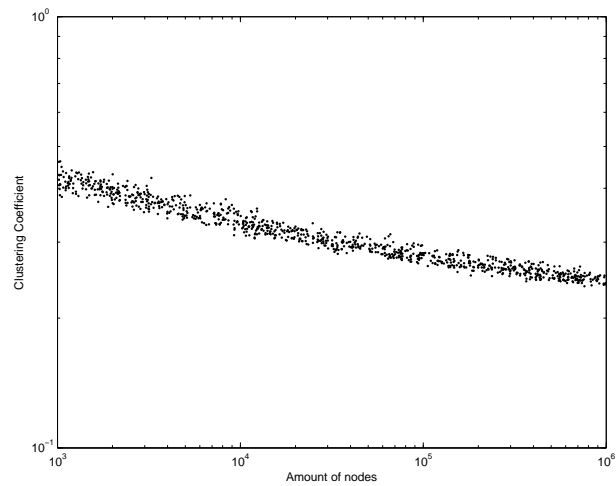


Figure 26: Clustering coefficient in function of the number of nodes of a graph for $k = 5$, $S_i = 4$ and $l = 3$

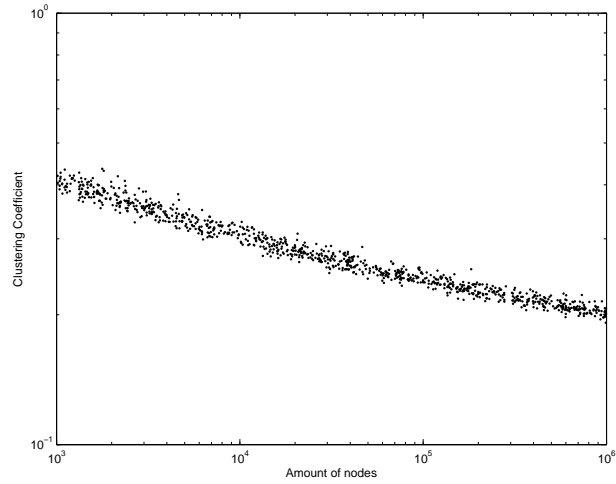


Figure 27: Clustering coefficient in function of the number of nodes of a graph for $k = 5$, $S_i = 4$ and $l = 4$

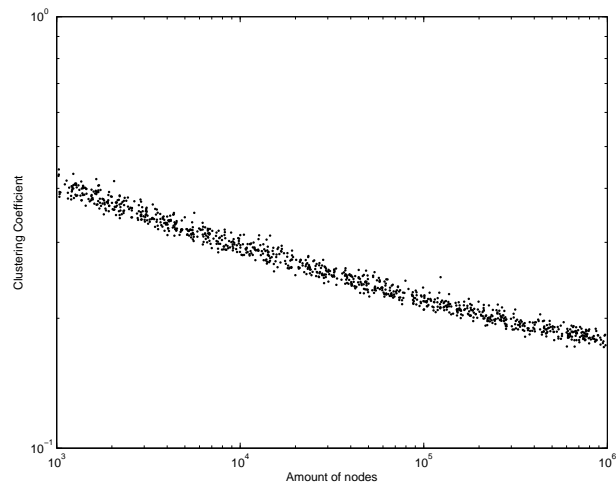


Figure 28: Clustering coefficient in function of the number of nodes of a graph for $k = 5$, $S_i = 4$ and $l = 5$

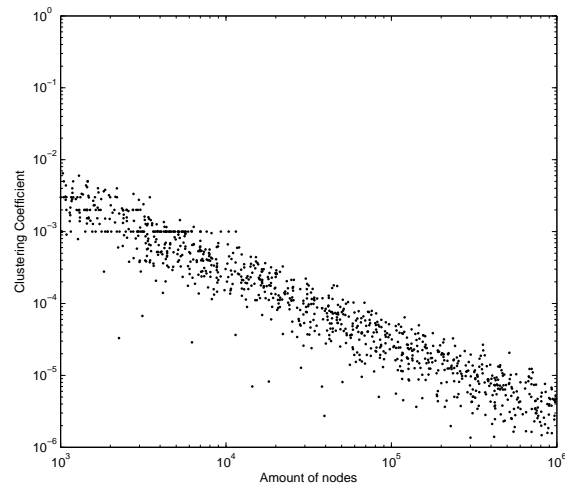


Figure 29: Clustering coefficient in function of the number of nodes of a graph for Erdős-Rényi model with the same number of edges of the case Barabási-Albert with $l = 1$

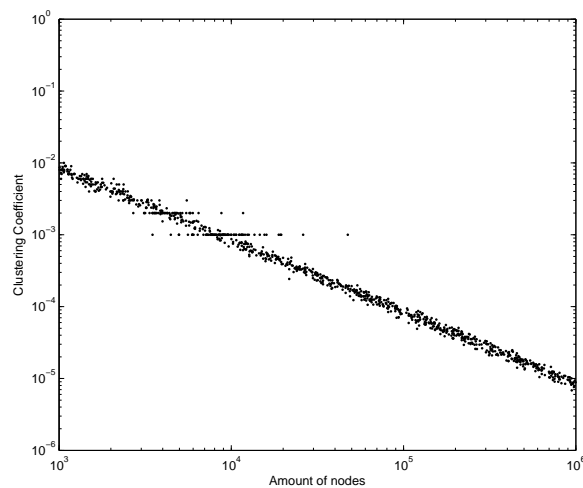


Figure 30: Clustering coefficient in function of the number of nodes of a graph for Erdős-Rényi model with the same number of edges of the case Barabási-Albert with $l = 2$

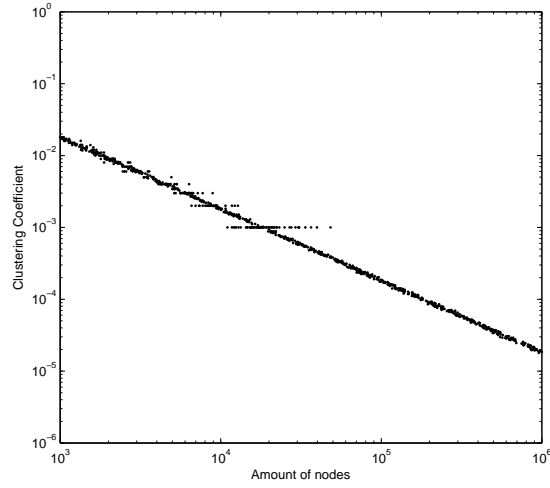


Figure 31: Clustering coefficient in function of the number of nodes of a graph for Erdős-Rényi model with the same number of edges of the case Barabási-Albert with $l = 3$

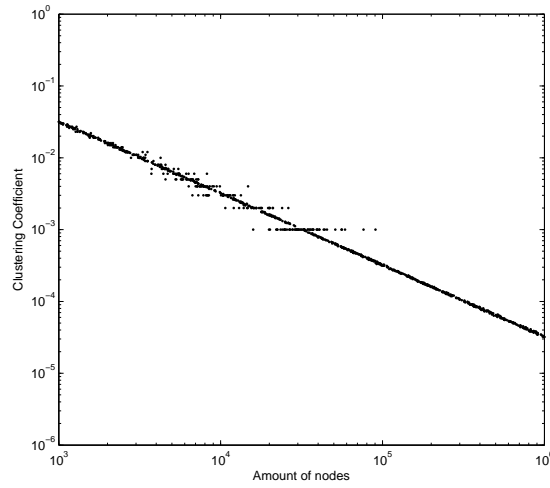


Figure 32: Clustering coefficient in function of the number of nodes of a graph for Erdős-Rényi model with the same number of edges of the case Barabási-Albert with $l = 4$

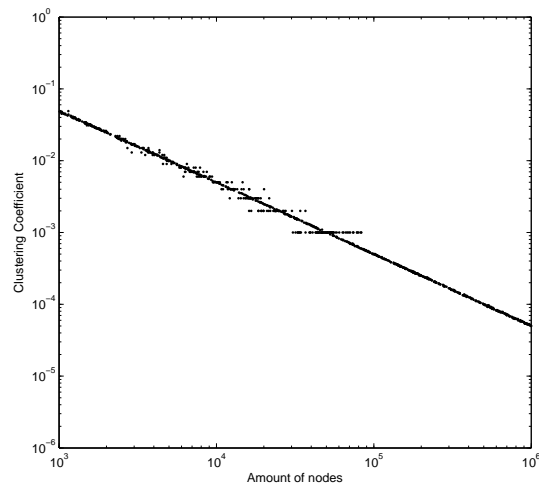


Figure 33: Clustering coefficient in function of the number of nodes of a graph for Erdős-Rényi model with the same number of edges of the case Barabási-Albert with $l = 5$

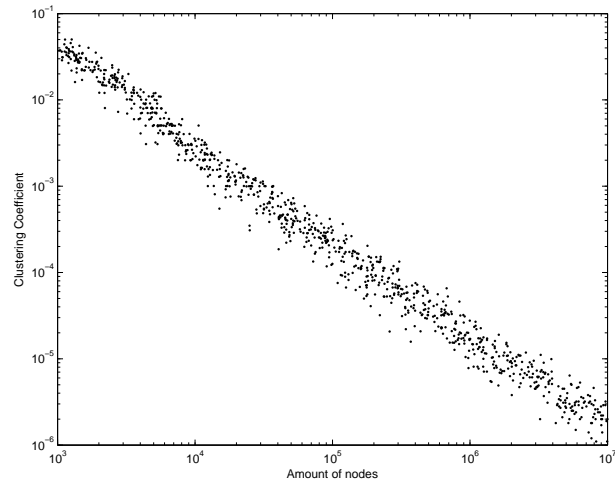


Figure 34: Clustering coefficient in function of the number of nodes of a graph for $k = 2$, $S_i = 1$ and $l = 1$

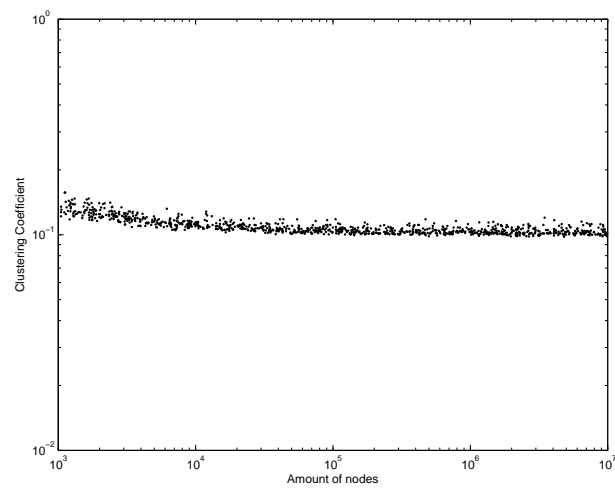


Figure 35: Clustering coefficient in function of the number of nodes of a graph for $k = 3$, $S_i = 2$ and $l = 1$

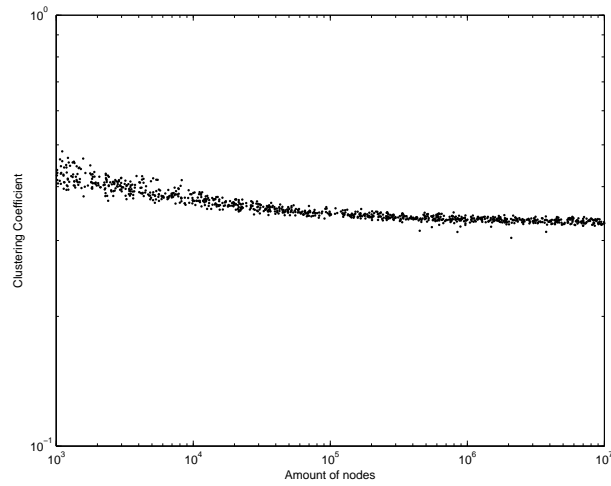


Figure 36: Clustering coefficient in function of the number of nodes of a graph for $k = 4$, $S_i = 3$ and $l = 1$

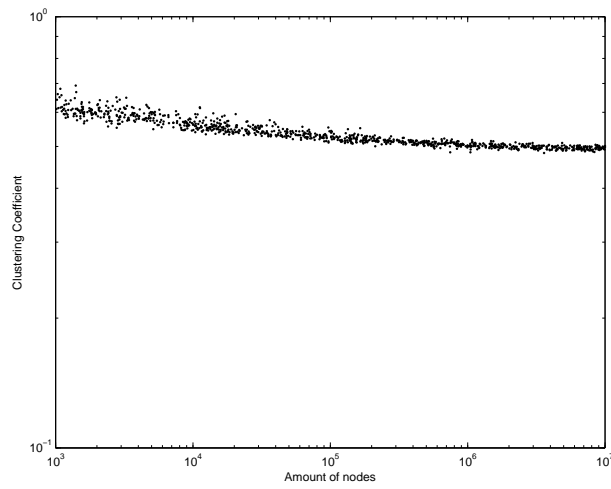


Figure 37: Clustering coefficient in function of the number of nodes of a graph for $k = 5$, $S_i = 4$ and $l = 1$

B.2 Varying Context Sizes

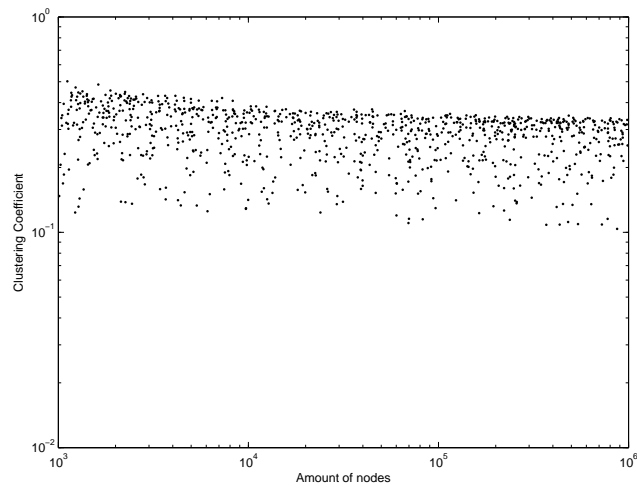


Figure 38: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 3 and 4 and $l = 1$

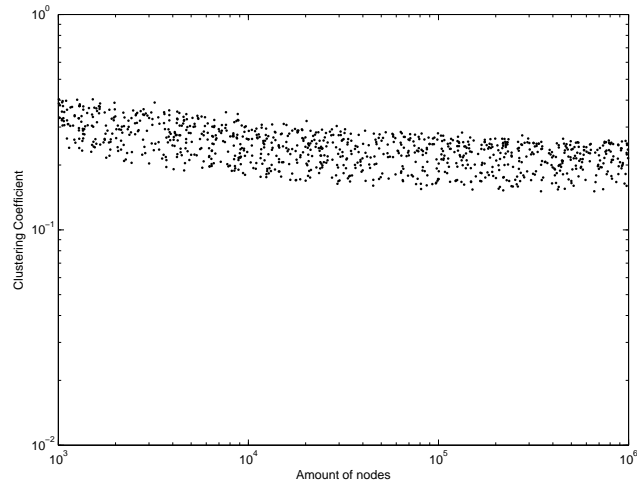


Figure 39: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 3 and 4 and $l = 2$

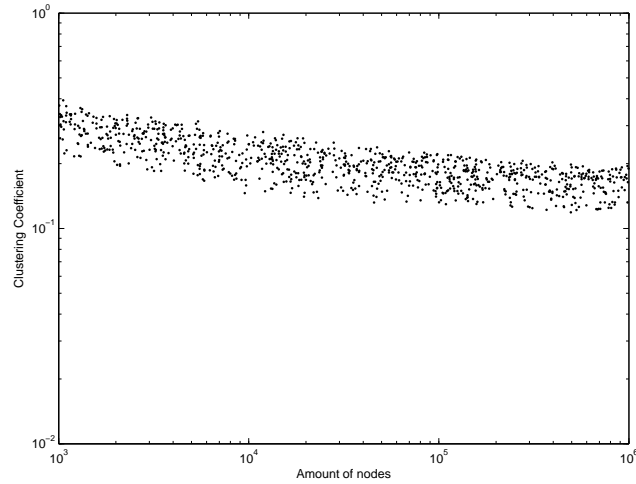


Figure 40: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 3 and 4 and $l = 3$

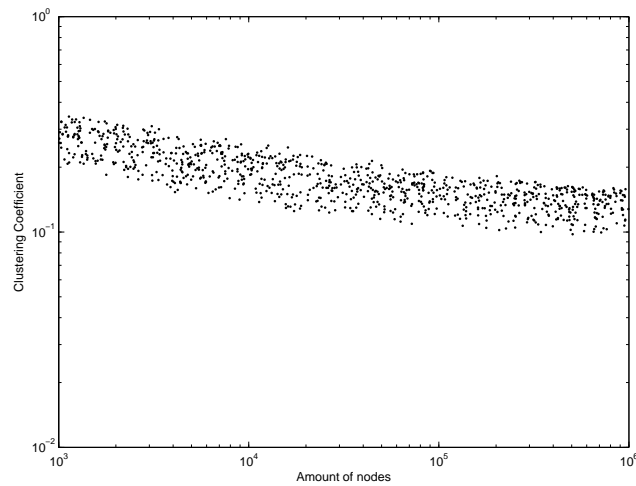


Figure 41: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 3 and 4 and $l = 4$

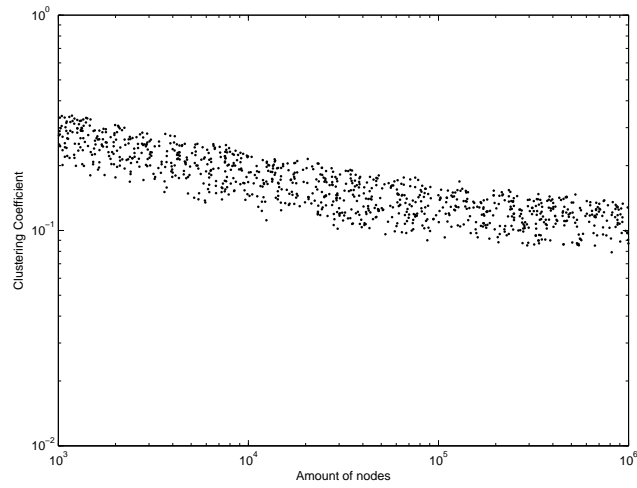


Figure 42: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 3 and 4 and $l = 5$

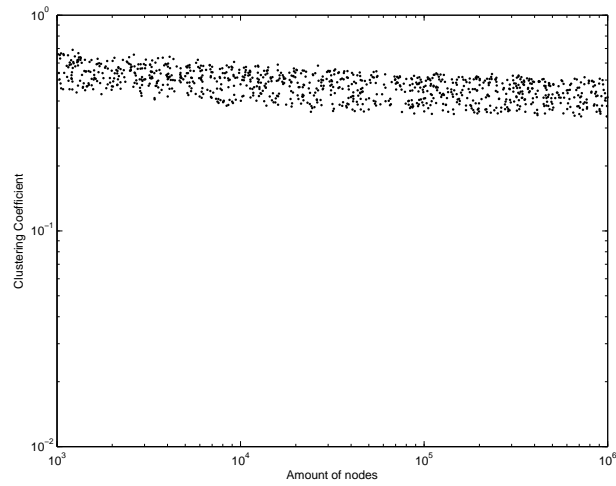


Figure 43: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 4 and 5 and $l = 1$

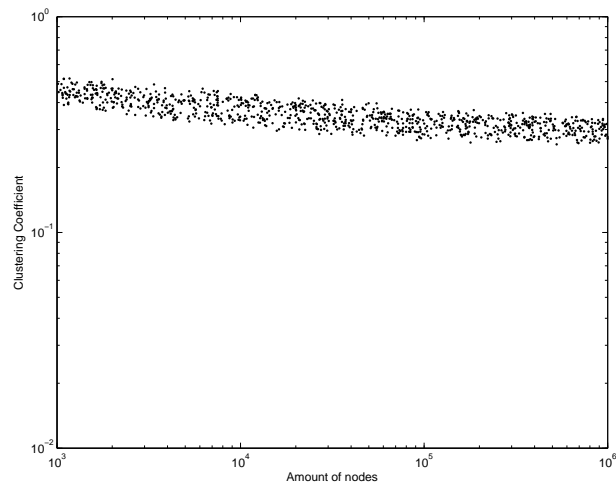


Figure 44: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 4 and 5 and $l = 2$

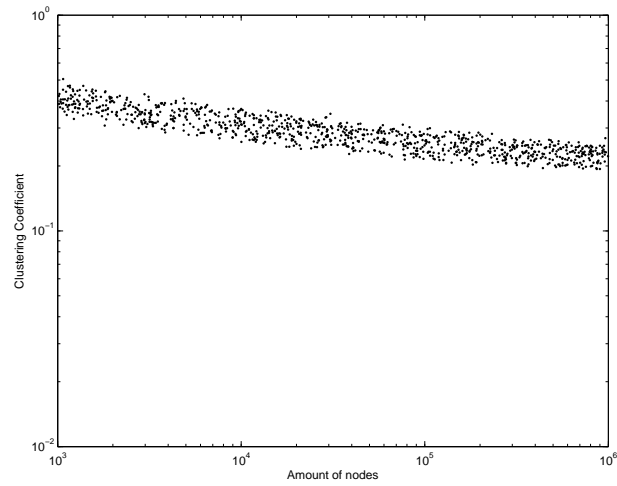


Figure 45: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 4 and 5 and $l = 3$

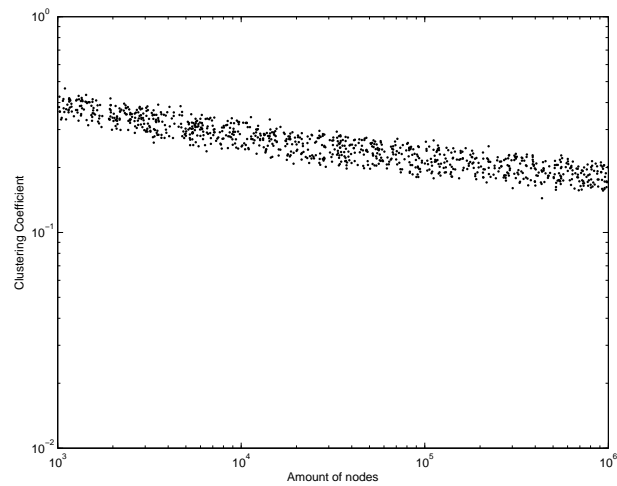


Figure 46: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 4 and 5 and $l = 4$

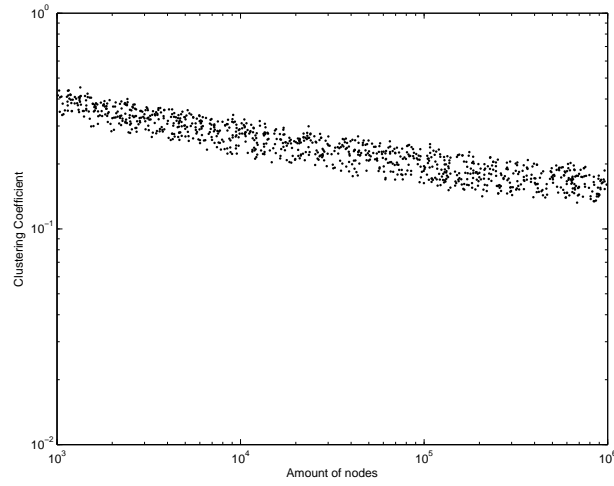


Figure 47: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 4 and 5 and $l = 5$

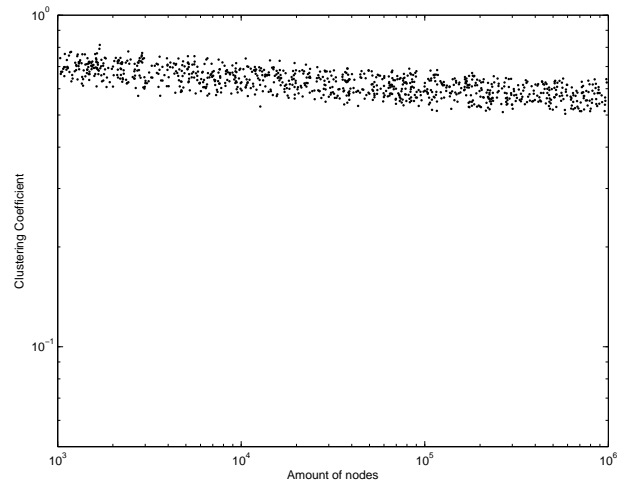


Figure 48: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 5 and 6 and $l = 1$

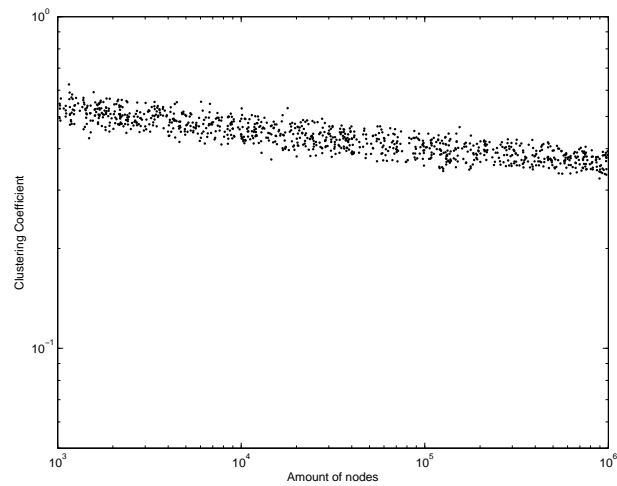


Figure 49: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 5 and 6 and $l = 2$

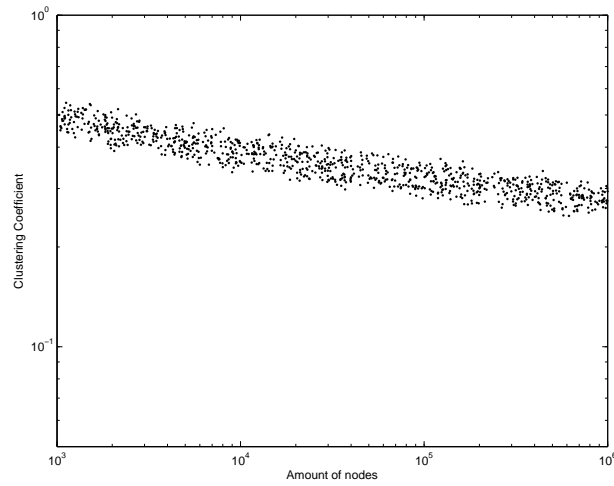


Figure 50: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 5 and 6 and $l = 3$

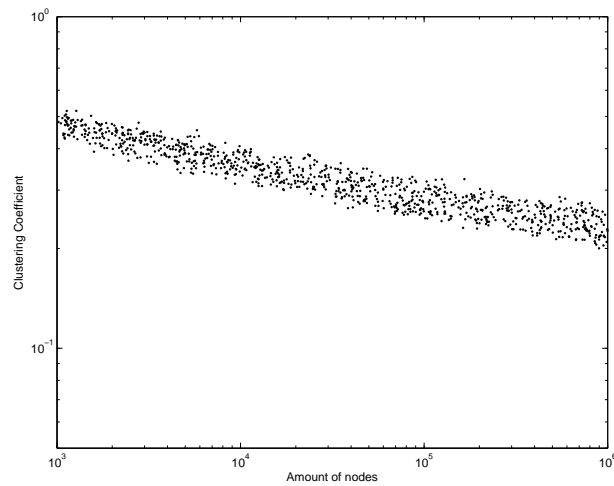


Figure 51: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 5 and 6 and $l = 4$

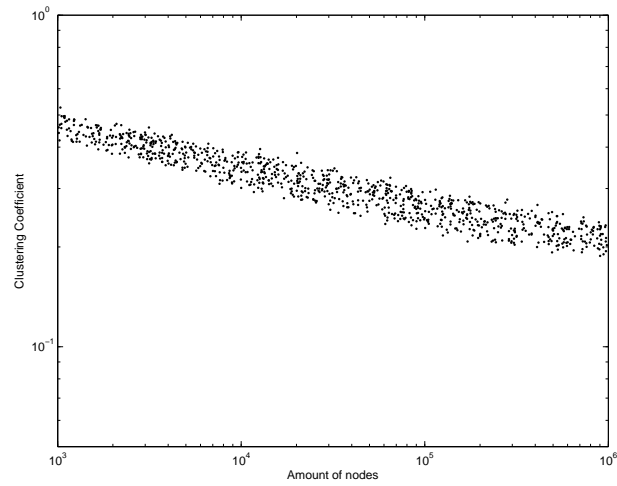


Figure 52: Clustering coefficient in function of the number of nodes of a graph for random size of context varying between 5 and 6 and $l = 5$

B.3 Contexts with a Fixed Average Size

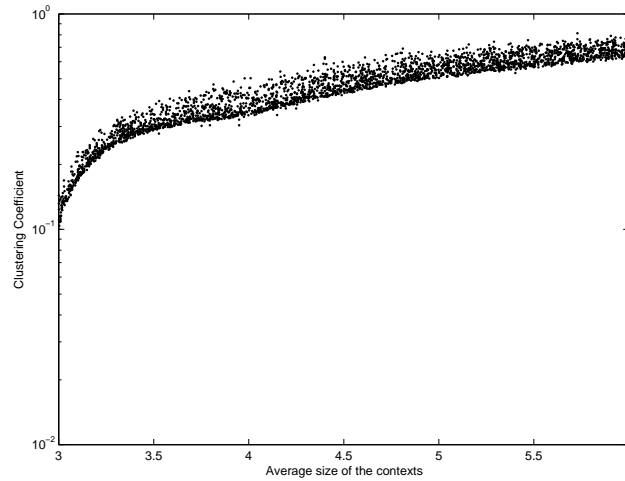


Figure 53: Clustering coefficient in function of the average size of the contexts for random size (mixed the three cases) with $l = 1$

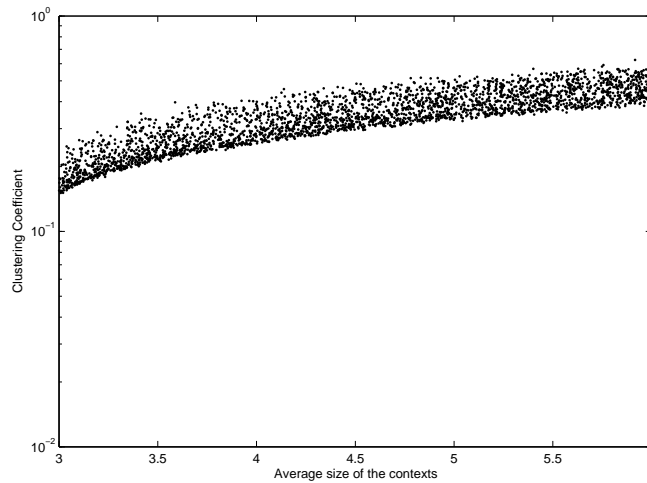


Figure 54: Clustering coefficient in function of the average size of the contexts for random size (mixed the three cases) with $l = 2$

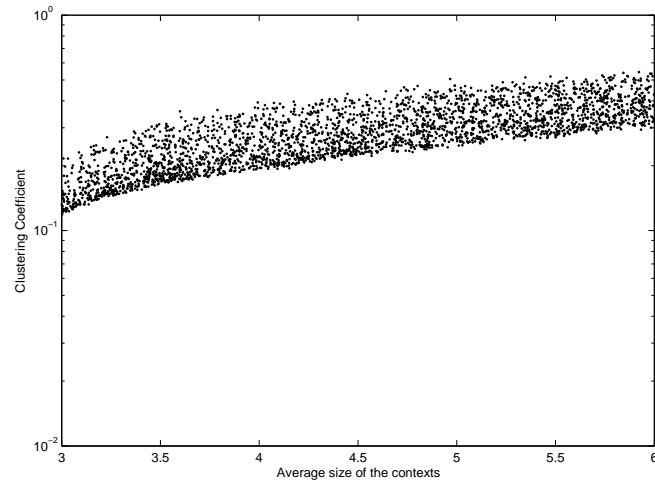


Figure 55: Clustering coefficient in function of the average size of the contexts for random size (mixed the three cases) with $l = 3$

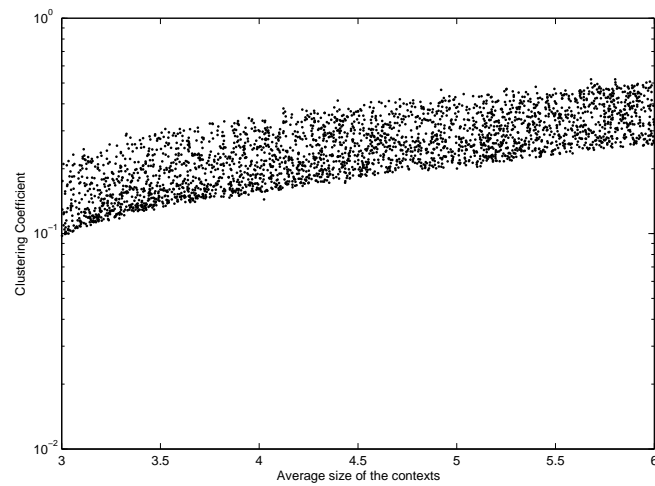


Figure 56: Clustering coefficient in function of the average size of the contexts for random size (mixed the three cases) with $l = 4$

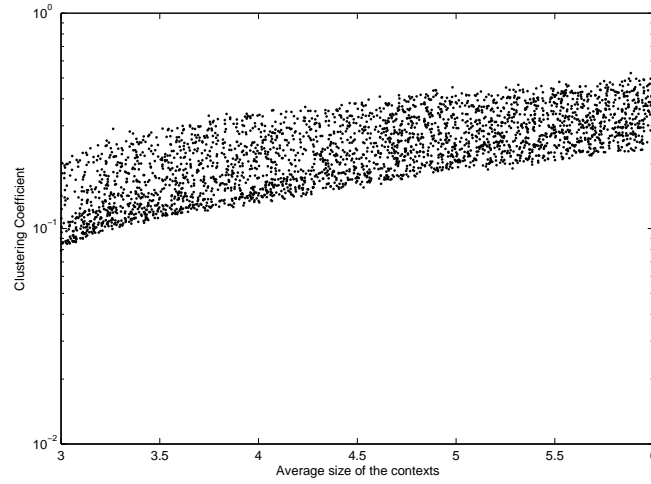


Figure 57: Clustering coefficient in function of the average size of the contexts for random size (mixed the three cases) with $l = 5$

B.4 Context Sizes with Prescribed Probabilities

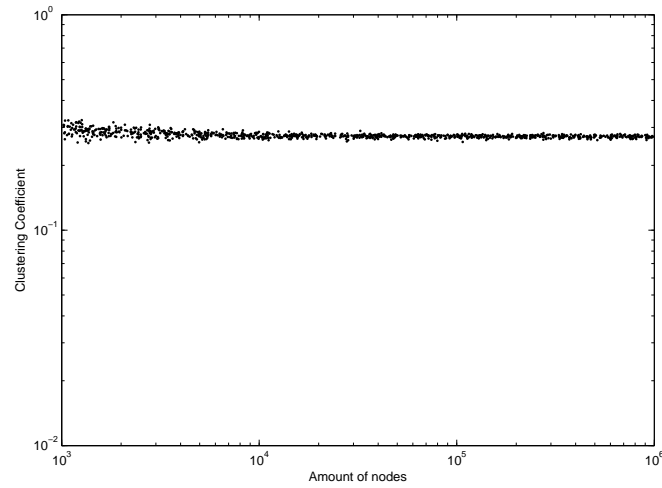


Figure 58: Clustering coefficient in function of the number of nodes of a graph for random size of context according to prescribed probability of each size and $l = 1$

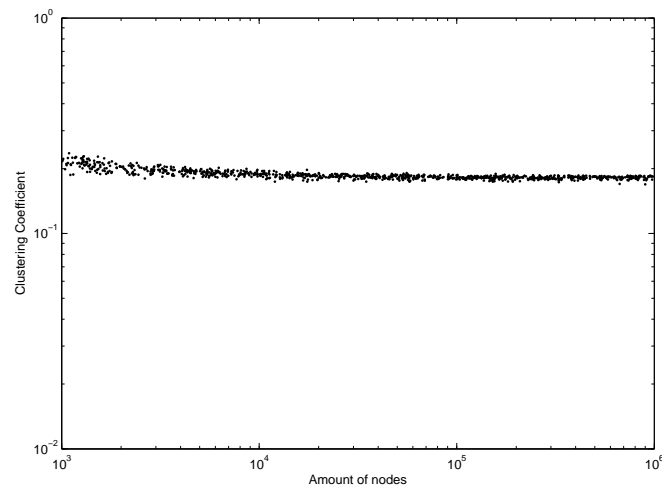


Figure 59: Clustering coefficient in function of the number of nodes of a graph for random size of context according to prescribed probability of each size and $l = 2$

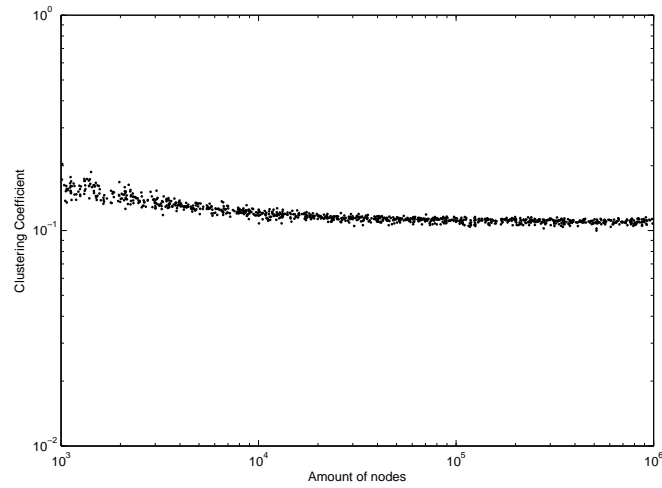


Figure 60: Clustering coefficient in function of the number of nodes of a graph for random size of context according to prescribed probability of each size and $l = 3$

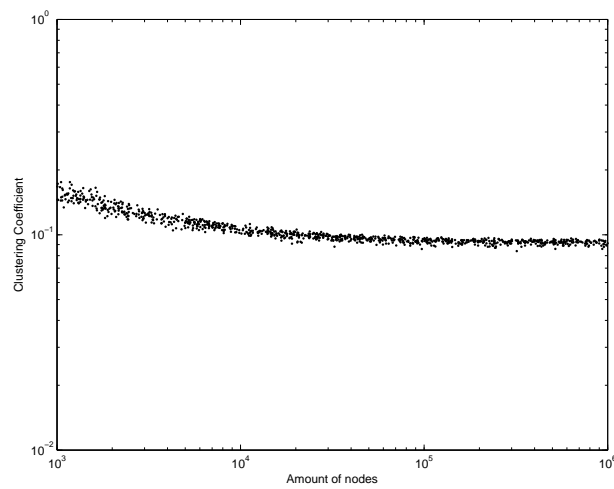


Figure 61: Clustering coefficient in function of the number of nodes of a graph for random size of context according to prescribed probability of each size and $l = 4$

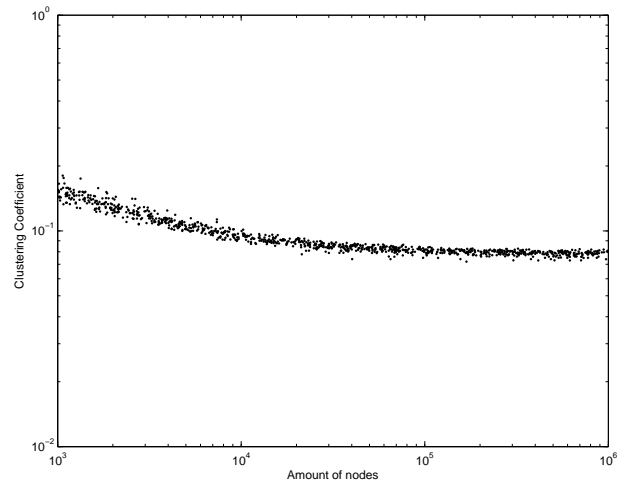


Figure 62: Clustering coefficient in function of the number of nodes of a graph for random size of context according to prescribed probability of each size and $l = 5$

C Regressions for the Clustering Coefficient

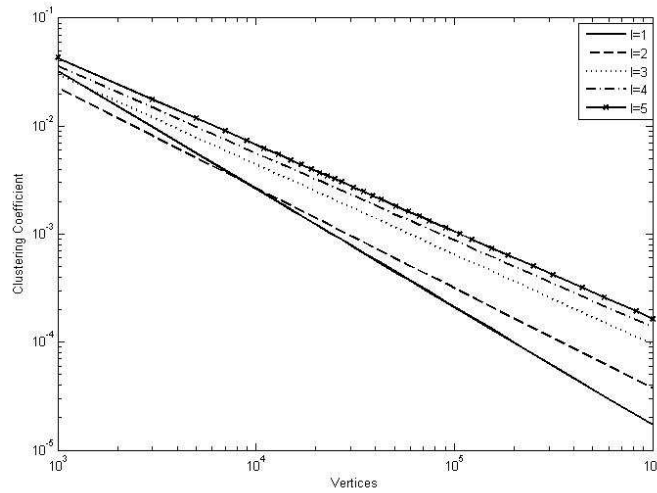


Figure 63: Regression of the clustering coefficient in function of the number of nodes of a graph for the case $k = 2$, $S_i = 1$ and $l = 1, 2, 3, 4, 5$

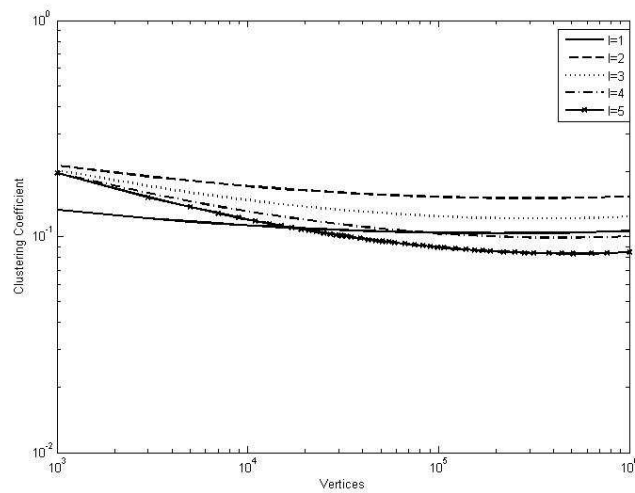


Figure 64: Regression of the clustering coefficient in function of the number of nodes of a graph for the case $k = 3$, $S_i = 2$ and $l = 1, 2, 3, 4, 5$

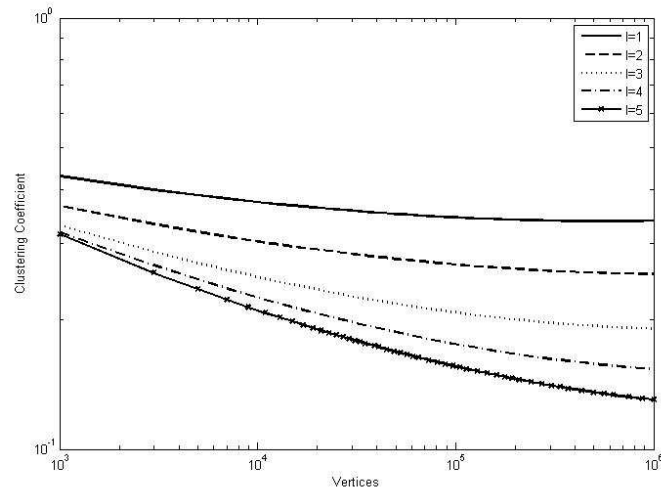


Figure 65: Regression of the clustering coefficient in function of the number of nodes of a graph for the case $k = 4$, $S_i = 3$ and $l = 1, 2, 3, 4, 5$

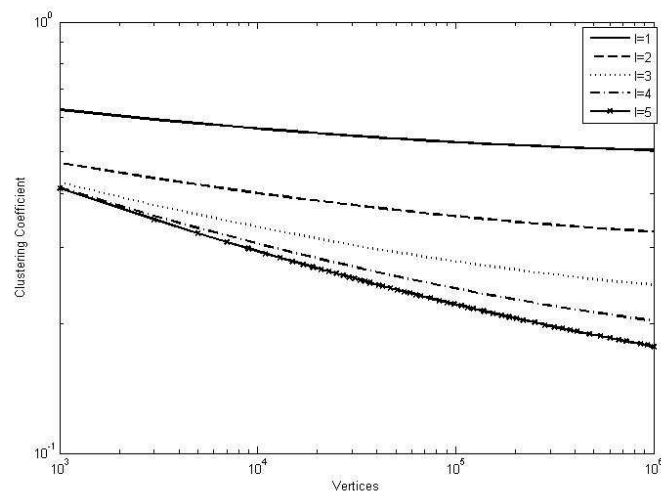


Figure 66: Regression of the clustering coefficient in function of the number of nodes of a graph for the case $k = 5$, $S_i = 4$ and $l = 1, 2, 3, 4, 5$

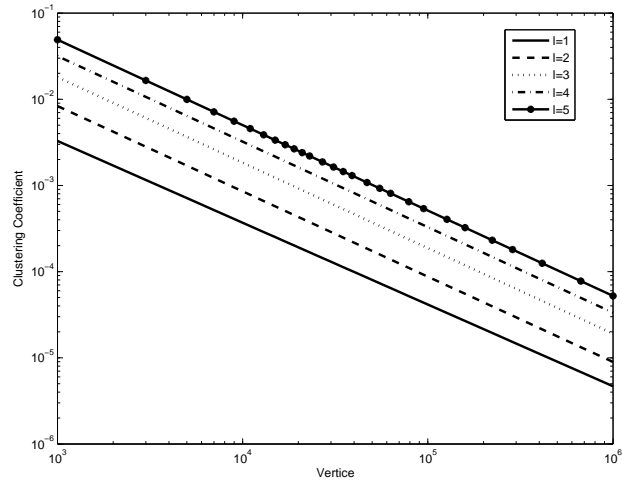


Figure 67: Regression of the clustering coefficient in function of the number of nodes of a graph for Erdős-Rényi model with the number of edges equal to the number of edges of the Barabási-Albert simulations from $l = 1$ to $l = 5$

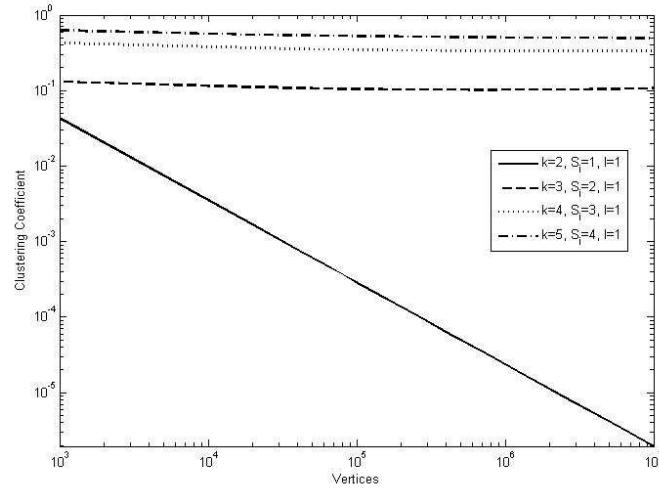


Figure 68: Regression of the clustering coefficient in function of the number of nodes of a graph for the cases $k = 2/S_i = 1$, $k = 3/S_i = 2$, $k = 4/S_i = 3$, $k = 5/S_i = 4$ and $l = 1$ - larger graph

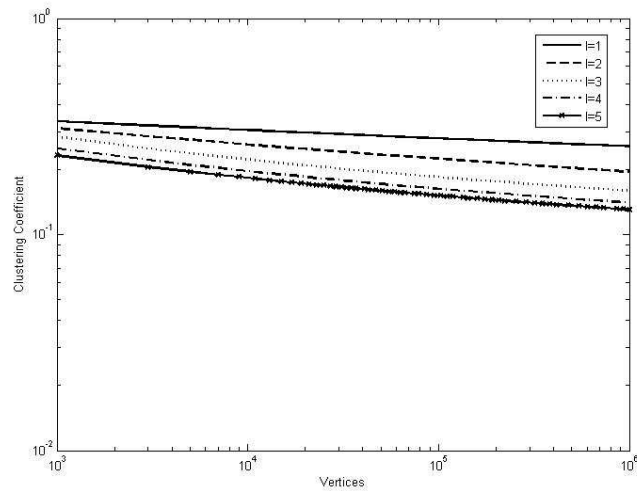


Figure 69: Regression of the clustering coefficient in function of the number of nodes of a graph for random size of context varying between 3 and 4 and $l = 1, 2, 3, 4, 5$

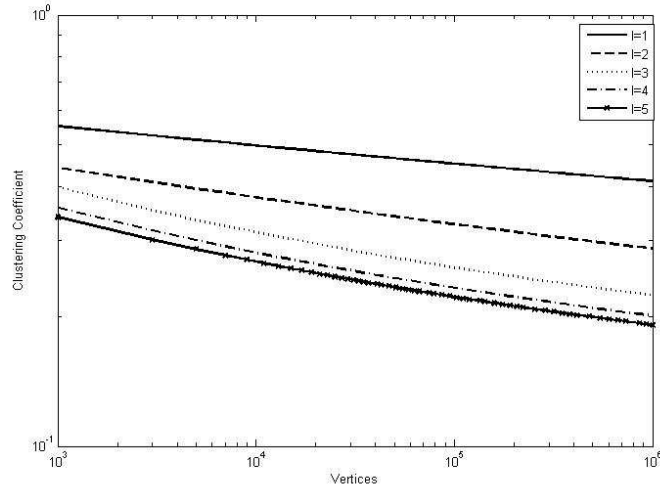


Figure 70: Regression of the clustering coefficient in function of the number of nodes of a graph for random size of context varying between 4 and 5 and $l = 1, 2, 3, 4, 5$

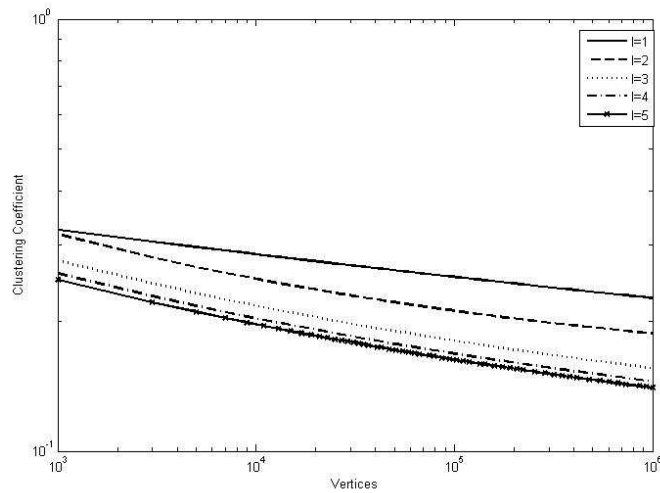


Figure 71: Regression of the clustering coefficient in function of the number of nodes of a graph for random size of context varying between 5 and 6 and $l = 1, 2, 3, 4, 5$

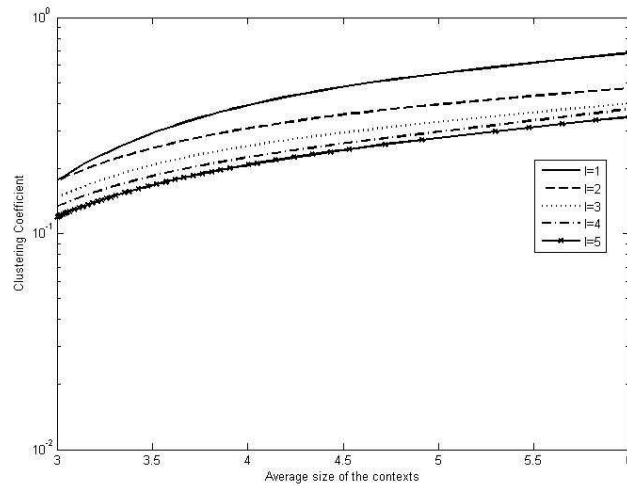


Figure 72: Regressions of the clustering coefficient in function of the average size of the contexts for random size (mixed the three cases) and $l = 1, 2, 3, 4, 5$

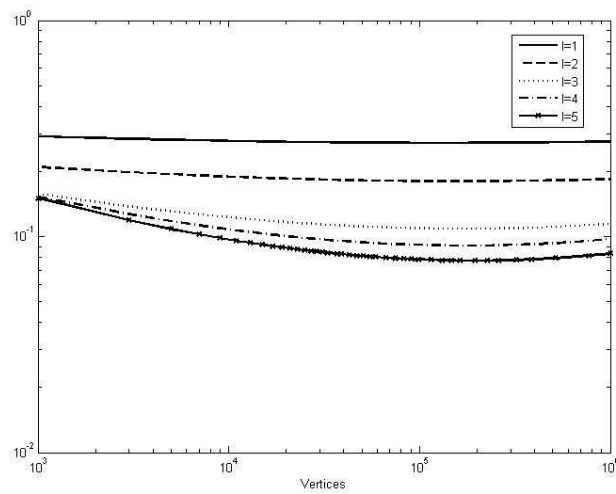


Figure 73: Regression of the clustering coefficient in function of the number of nodes of a graph according to prescribed probability of each size and $l = 1, 2, 3, 4, 5$

D Computing the Estimation of the Clustering Coefficient

D.1 Number of Iterations

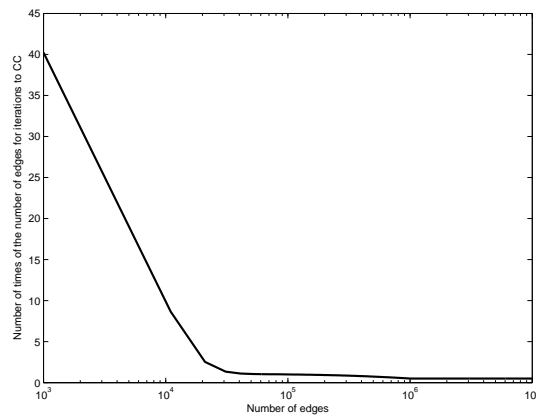


Figure 74: Factor of multiplication to the number of edges of a graph to obtain the initial amount of iterations for the calculation of the clustering coefficient of a graph

D.2 Running Times

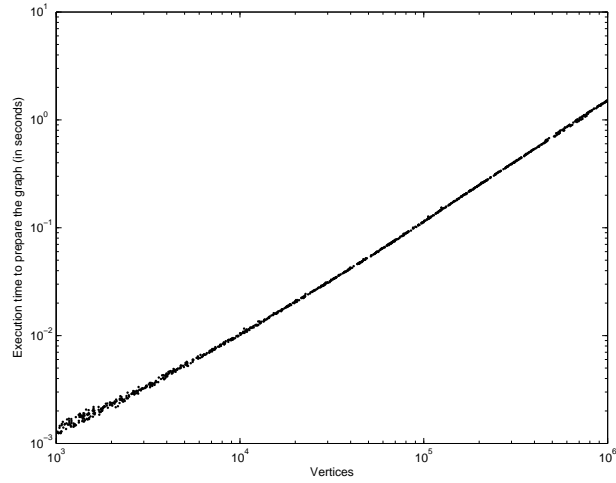


Figure 75: Execution time to prepare the graph in the constant size of the context case for $k = 2$, $S_i = 1$ and $l = 1$

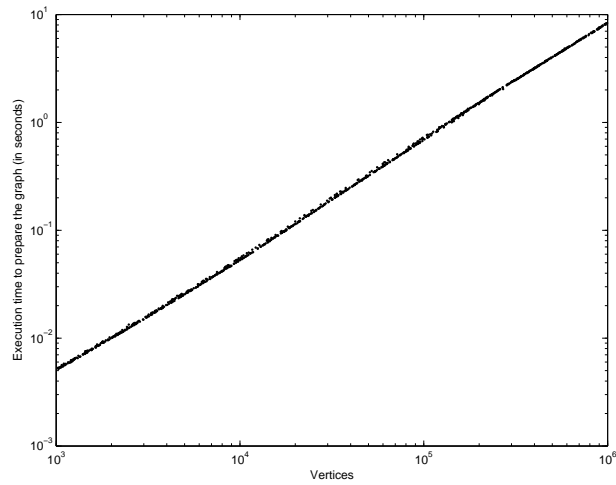


Figure 76: Execution time to prepare the graph in the constant size of the context case for $k = 2$, $S_i = 1$ and $l = 5$

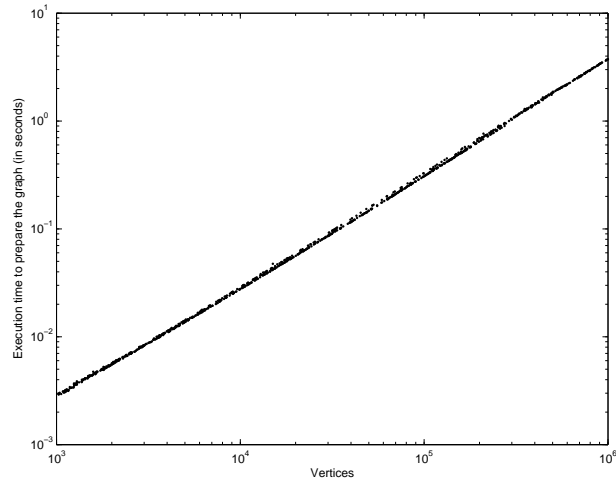


Figure 77: Execution time to prepare the graph in the constant size of the context case for $k = 4$, $S_i = 3$ and $l = 1$

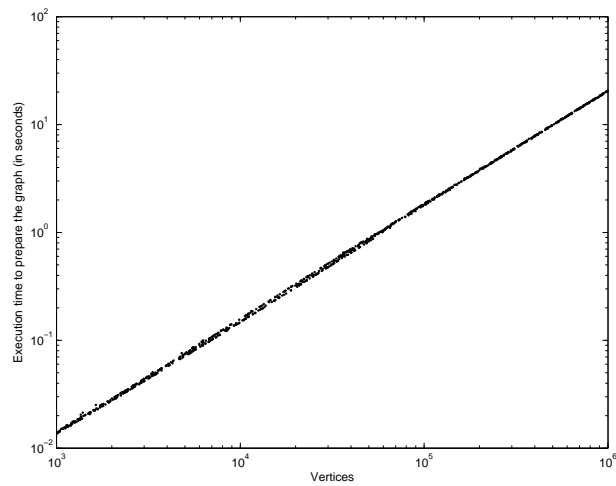


Figure 78: Execution time to prepare the graph in the constant size of the context case for $k = 4$, $S_i = 3$ and $l = 5$

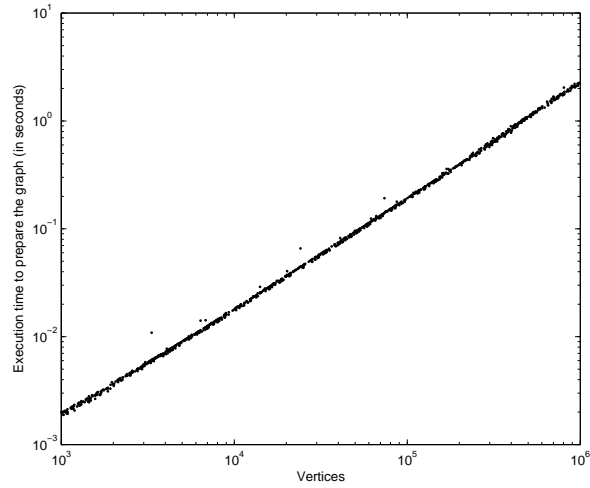


Figure 79: Execution time to prepare the graph in the Erdős-Rényi case for the number of edges equal to the number of edges of the Barabási-Albert simulations with $l = 1$

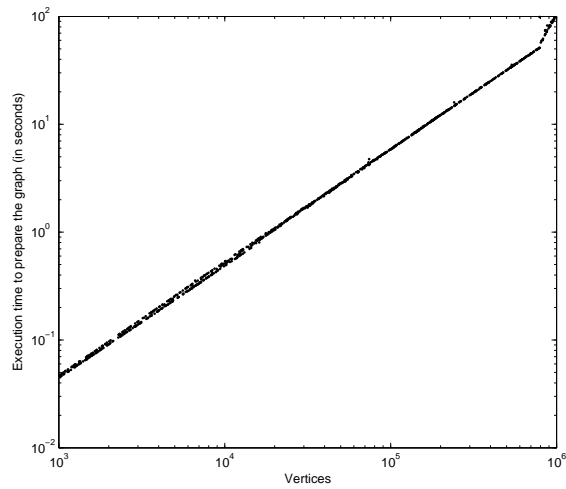


Figure 80: Execution time to prepare the graph in the Erdős-Rényi case for the number of edges equal to the number of edges of the Barabási-Albert simulations with $l = 5$

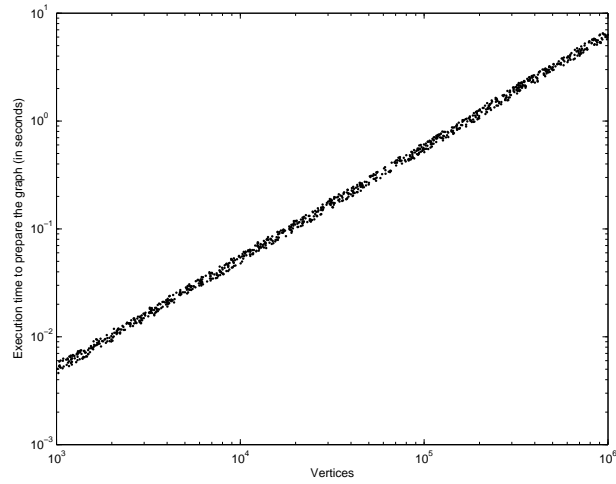


Figure 81: Execution time to prepare the graph in the random size of the context case with two possible values (4 and 5) for $l = 1$

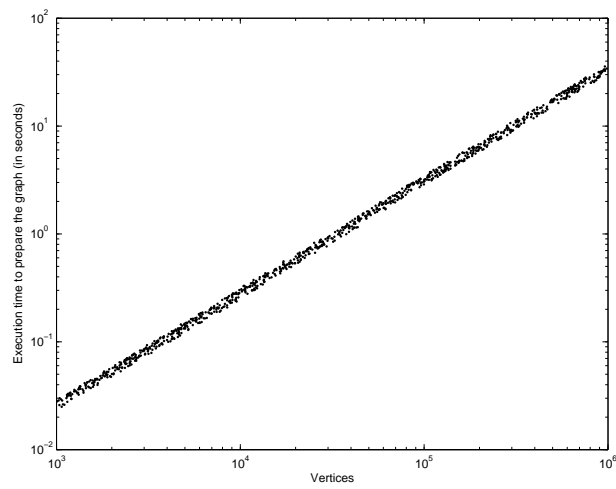


Figure 82: Execution time to prepare the graph in the random size of the context case with two possible values (4 and 5) for $l = 5$

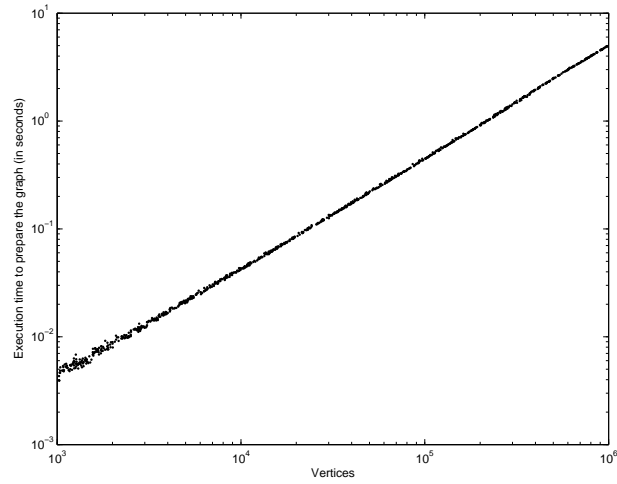


Figure 83: Execution time to prepare the graph in the case with prescribed probability of each size of context for $l = 1$

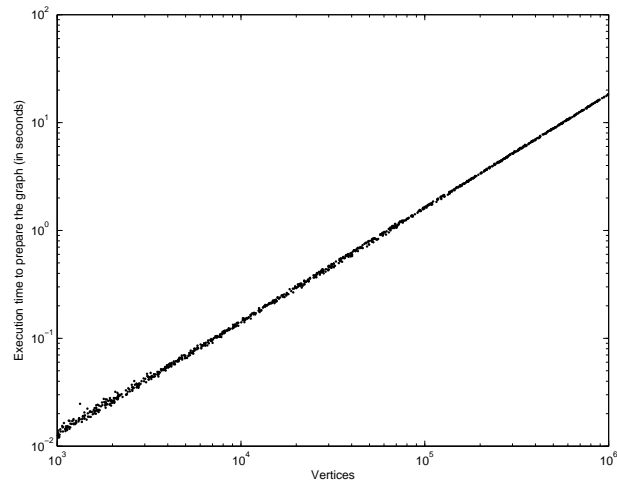


Figure 84: Execution time to prepare the graph in the case with prescribed probability of each size of context for $l = 5$

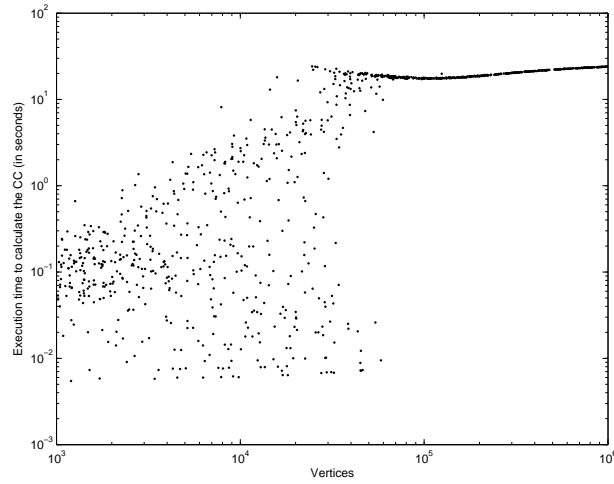


Figure 85: Execution time to calculate the clustering coefficient in the constant size of the context case for $k = 2$, $S_i = 1$ and $l = 1$

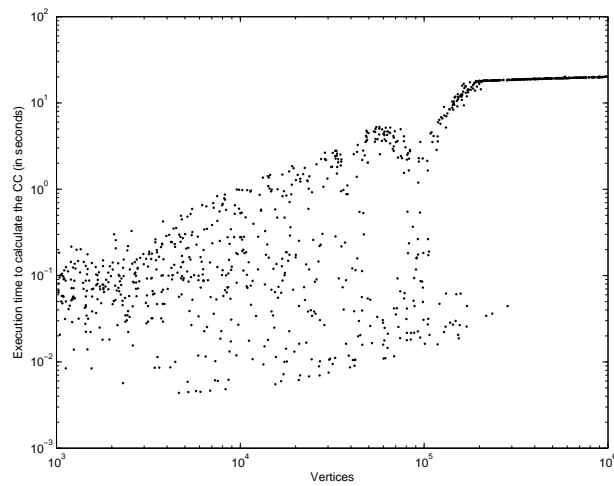


Figure 86: Execution time to calculate the clustering coefficient in the constant size of the context case for $k = 2$, $S_i = 1$ and $l = 5$

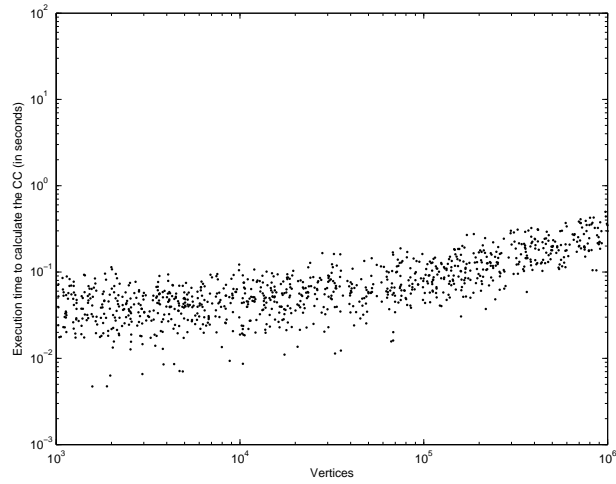


Figure 87: Execution time to calculate the clustering coefficient in the constant size of the context case for $k = 4$, $S_i = 3$ and $l = 1$

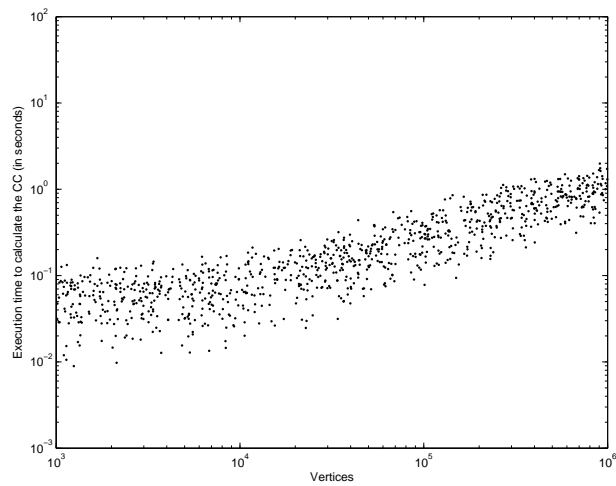


Figure 88: Execution time to calculate the clustering coefficient in the constant size of the context case for $k = 4$, $S_i = 3$ and $l = 5$

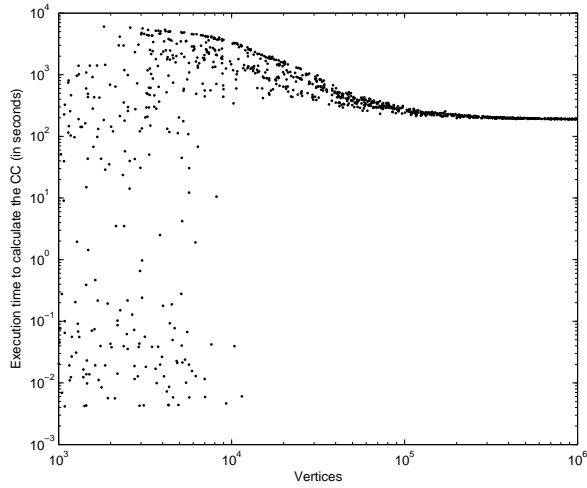


Figure 89: Execution time to calculate the clustering coefficient in the Erdős-Rényi case for the number of edges equal to the number of edges of the Barabási-Albert simulations with $l = 1$

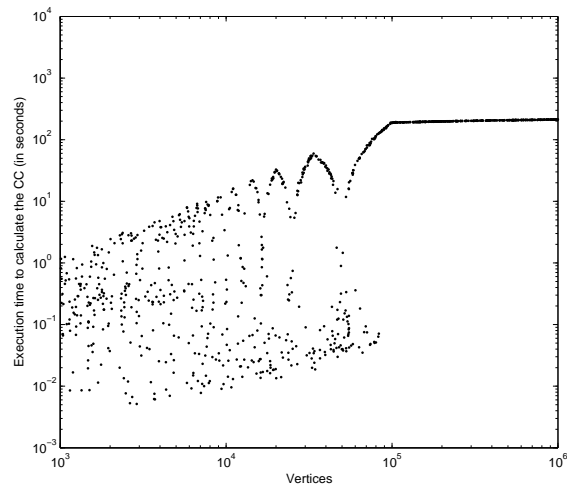


Figure 90: Execution time to calculate the clustering coefficient in the Erdős-Rényi case for the number of edges equal to the number of edges of the Barabási-Albert simulations with $l = 5$

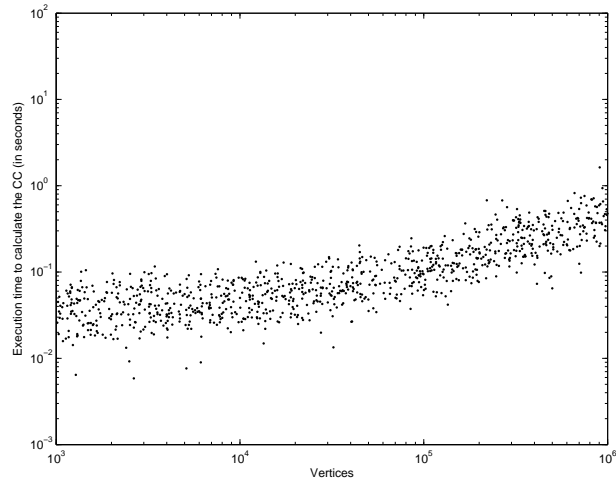


Figure 91: Execution time to calculate the clustering coefficient in the random size of the context case with two possible values (4 and 5) for $l = 1$

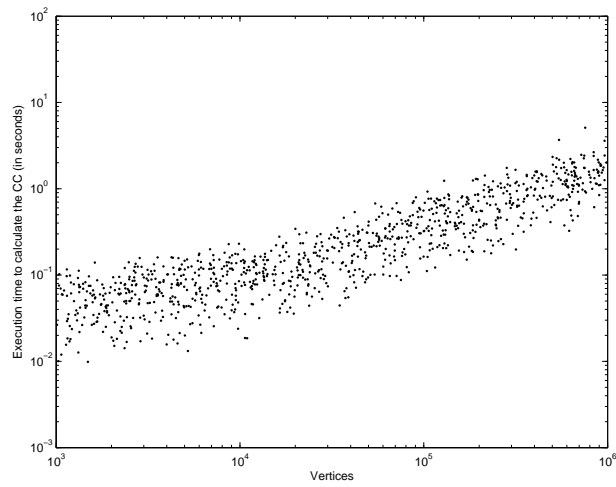


Figure 92: Execution time to calculate the clustering coefficient in the random size of the context case with two possible values (4 and 5) for $l = 5$

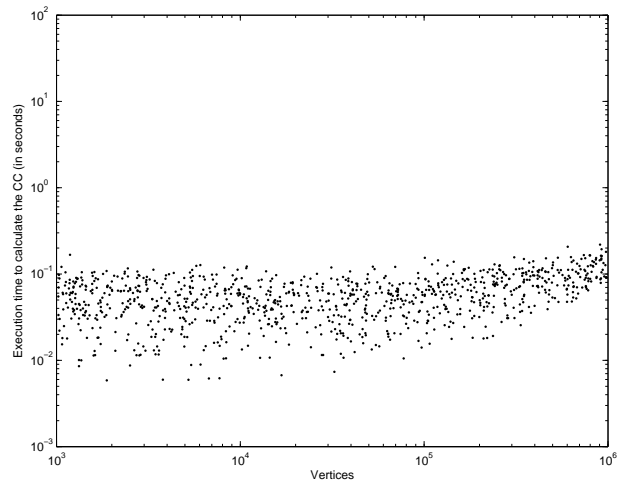


Figure 93: Execution time to calculate the clustering coefficient in the case with prescribed probability of each size of context for $l = 1$

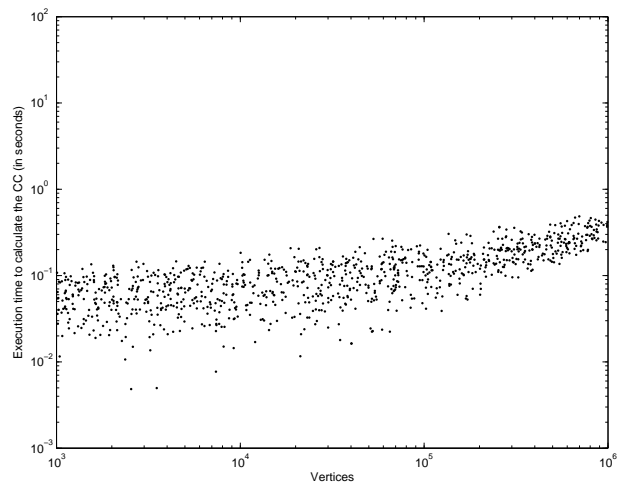


Figure 94: Execution time to calculate the clustering coefficient in the case with prescribed probability of each size of context for $l = 5$



Centre de recherche INRIA Nancy – Grand Est
LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
<http://www.inria.fr>
ISSN 0249-0803