

The Lattice of Chemical Organisations

Pietro Speroni di Fenizio ¹

¹ Bio Systems Analysis Group, Jena Centre for Bioinformatics, Friedrich Schiller-University Jena, D-07737 Jena
ecal2015@piespe.net

Abstract

The paper describes how, in an Artificial Chemistry under flow conditions, the set of organisations form a lattice. The consequences of this are described, in particular how a series of theorems, valid for lattices, can be applied to more easily discover the complete set of organisations. An algorithm is then developed that uses such theorems to explore such lattice. The algorithm is applied first to the NTop Artificial Chemistry and then to an extension of it. Due to its complexity this system is also suggested as a benchmark case to test new Artificial Chemistries' algorithms.

Introduction

In 1994 Walter Fontana and Leo Buss introduced the concept of Organisation to represent the fixed points in Constructive Dynamical Systems (Fontana and Buss, 1994a, 1994b, 1996). Looking at Dynamical Systems they observed how they could predict only quantitative dynamics, and no qualitative development. True novelty could never appear through Ordinary Differential Equations. But novelty is inherent in this world. And it is unpredictability is what makes the world so interesting. So, to describe how a system could transform the qualitative space of possibilities, they presented the concept of Constructive Dynamical Systems. Those were a molecular based system, where new molecules could be generated through the interaction of existing ones. Such system was thought as a model for the macro-molecules inside a cell, but this was just one of the possible examples, of such a general theory.

Using standard dynamical systems as a metaphor, they then went on describing what would be the equivalent fixed points in their system; points where no novelty would appear, and called such cases Organisations. A set of molecules would be an organisation if, and only if, for each molecule inside the set, there was a reaction among the molecules in the set that would produce it; and, given any reaction among the molecules of the set, the result would always be a molecule of the set. The first property was called Self-Maintenance and the second Closure. A set satisfying both of those properties would be called an Organisation.

Later (Dittrich, Speroni di Fenizio, 2007) Organisations were then renamed Semi-Organisations when it became clear that those properties were not enough to permit to those sets to be dynamically stability. Organisations were so re-defined as special Semi-Organisations where it is possible for all the reactions among molecules inside them to be active, and have no molecule type diminish.

Studying Organisations and Semi-Organisations, it was soon found that under quite common assumptions, those structures would form a Lattice. It should be noted that this it is not

always true for every reaction system, but it is true if we assume that every molecule has a certain probability to disappear (i.e. every molecule would have an out-flux greater than zero). Those systems were every molecule has an out flux were called *Flow Systems* to distinguish them from the more general *General Reaction Systems* (Dittrich, Speroni, 2007). It should be noted how all of Fontana and Buss' models had an out-flux applied to each molecule that would destroy it at the same speed at which others were produced. So those models were Flow Systems, and the set of organisations would form a Lattice.

Other Flow Systems are also possible; for example if we consider cells, and thus cell growth, although it is not true that each molecule has a certain probability higher than zero of being excreted, it is true that as the cell grows the relative concentration of each molecule that is not generates diminishes, having the same effective result as if each molecule was subject to an out-flux. Then when the cell reproduces the average amount of that molecule halves, eventually reaching zero after enough reproductions. Thus the set of molecules inside a cell form a Flow System, and the set of possible Organisations in a living system form a lattice. Vice-versa, the reaction system in the atmosphere of a planet is not a Flow System, but a General Reaction System, and the Organisations do not form a lattice but just a Partially Ordered Set. Flow Systems and General Reaction Systems are not the only possible type of Artificial Chemistries: a more detailed analysis can distinguish also Catalytic Flow Systems as a specific type of Flow System where each molecule reacts in a catalytic way. In other words they are not used up in the

Catalytic Flow Systems	Flow Systems	General Reaction Systems
<i>All molecules in each reaction are catalytic; all molecules have an out flux</i>	<i>All molecules have an out-flux</i>	<i>No requirements on the out-flux nor on which molecules are catalytic</i>
each semi-organisation is an organisations	some semi-organisations are organisations	some semi-organisations are organisations
both organisations and semi-organisations form a lattice	both organisations and semi-organisations form a lattice	neither organisations nor semi-organisations form a lattice

Table 1: Different type of Artificial Chemistries produce sets of organisations with different properties.

reaction, while the prime material that generates the result is supposedly coming from a substrate of basic material floating around. Substrate to which each molecule eventually decays through the out-flux. Catalytic Flow Systems, Flow Systems, and Reaction Systems are quite different in terms of their relative properties (see Table 1).

This paper aims to investigate the consequences of the fact that under flow conditions, the set of Organisations form a lattice.

What does it mean to *understand* an Artificial Chemistry

In this context understanding an Artificial Chemistry means having a list of all the possible organisations; for each organisation know what are the organisations directly above and below it. With the organisations A being directly above (directly below) the organisation B we mean that A contains B (B contains A), $A > B$, ($A < B$) and there is no other organisation C that contain B and is contained by A (that contain A and is contained by B), $\nexists C$ such that $A > C > B$ ($\nexists C$ such that $A < C < B$). Thus being able to predict, given a starting condition, where the system will likely evolve, and if we mix two different systems what the result of it will be. I am speaking here of broad qualitative prediction, not precise quantitative ones. Knowing what molecules will be present, while ignoring the relative quantities, is a satisfactory result. Suppose you have two test tubes; each with a different experiment of the same Artificial Chemistry (AC); each having a small out-flux, thus that we can be sure that the set of organisations of this AC is a lattice.

If we consider the molecules inside, eventually they will express an organisation, as only the molecules which can be generated will be present and all the others will be lost through the out-flux. What will happen if we join those two test tubes? Because the organisations in a Flow System form a lattice, we can immediately say that such action will generate the organisation union of the two organisations. And this exist and is unique because in a lattice the union of two elements is always present and unique. Of course if the organisations present in the two test tubes are one inside the other, then the union is trivially the biggest one. More interestingly we can ask: suppose we have two test tubes, where two organisations are expressed, A and B; suppose we want to add some molecules to both A and B without changing the organisation inside; what can we add? And the answer is: we can add anything from the organisation C, where C is the organisation intersection between A and B. Again we know that C uniquely exists because the set of organisations form a lattice.

From this appears obvious that we do not only need to know the set of organisations, but also know the content of the two tables that given any two organisations A and B would tell us their union ($A \cup B$) and their intersection ($A \cap B$). Calculating those tables is also something that becomes more efficient by the fact that those organisations form a lattice and thus we can use theorems from lattice theory to calculate some of the results.

Thus understanding the set of organisations gives us some clear, practical abilities to understand what happens in an Artificial Chemistry. And when this is a lattice, it all becomes easier.

Basic Known Definitions from Previous Work

In a reaction system a closed set C is a set of molecules such that given any two molecules their reaction is still contained in C. A semi-self-maintaining set S is a set such that given any molecule m that is consumed in S, there exist a reaction inside S such that m is produced. Any set that is both semi-self-maintaining and closed is a semi-organisation. A self-maintaining set S is a semi-self-maintaining set such that there exist a reaction speed for all reactions among the molecules in S, with each reaction speed higher than zero, and such that each molecule has a production rate higher or equal to zero. In laymen's terms, a semi-self-maintaining set is a set where all molecules can be produced, and a self-maintaining one is one where this is globally possible. In Catalytic Flow systems the two coincide as every semi-self-maintaining set is self-maintaining [Speroni di Fenizio, 2007].

Given any set F, we can generate its closure as the smallest closed set containing F. This always exists (albeit it can be infinite in size, if the AC contains an infinite diversity of molecules) and it is unique; we will indicate this as $G_C(F)$.

Given two closed sets, we can define the closed union (\sqcup_C) as the closed set generate by the union of the two sets. Thus $\forall A, B$, closed sets; $A \sqcup_C B \equiv G_C(A \cup B)$. Similarly we can define, $A \sqcap_C B \equiv G_C(A \cap B)$. And then if C is the set of Closed Sets, $\langle C, \sqcup_C, \sqcap_C \rangle$ will be a Lattice.

Given any set F we can generate its semi-self-maintenance subset $G_{sSM}(F)$ as the biggest semi-self-maintaining subset contained in F. Again this exists and is unique. In a Flow System (but not in always in a reaction system), if we take a closed set C and we apply the operator $G_{sSM}(C)$ the result is still closed. So given any set F we can associate a semi-organisation to it as $G_{sO}(F) = G_{sSM}(G_C(F))$. This can be expanded by defining the operator Generate Self Maintaining set: $G_{SM}(T)$ as the operator that returns the biggest Self Maintaining Set contained in T.

Applying the equivalent definition of union and intersection among semi-self-maintaining sets, self maintaining sets, semi-organisations, and organisations, we obtain that in Flow Systems the set of semi-self maintaining sets, self maintaining sets, semi-organisations and organisations are all lattices with their respective union and intersection. Thus if O is the set of organisations in a Flow System $\langle O, \sqcup_O, \sqcap_O \rangle$ is a lattice. But not necessarily in a reaction system. We shall now explore the consequences of this.

Note: In the rest of the paper we shall use \sqcup to mean \sqcup_O and \sqcap to mean \sqcap_O .

Useful Theorems

There are several theorems from Lattice theory which can help in mapping the lattice of organisations, and thus understanding better an Artificial Chemistry. Generally speaking if we have two organisations A, B then $A \sqcup_O B = G_{SM}(G_{sSM}(G_C(A \cup B)))$. Those calculations can be long, so if we can shortcut the calculations just by working on the knowledge we already have this speeds up sensibly the work. In passing we note that in a Flow System for all sets A, although $G_{SM}(G_{sSM}(A)) = G_{SM}(A)$, it is usually faster to calculate it as $G_{SM}(G_{sSM}(A))$ as $|G_{sSM}(A)| < |A|$, and G_{sSM} being an algebraic operator is usually faster to calculate than G_{SM} .

Theorem 1: If A, B and C are elements in a Lattice $\langle L, \cup, \cap \rangle$, then we know that $(A \cup B) \cap C = A \cup (B \cap C)$. Now suppose that we have two organisations, S, C, and we are interested in calculating $T = S \sqcup C$. If we can express S as $S = A \sqcup B$ then $T = S \sqcup C = (A \sqcup B) \sqcup C = A \sqcup (B \sqcup C)$. And if we know the value of $B \sqcup C$ (for example, $B \sqcup C = R$) then $T = S \sqcup C = A \sqcup R$. Similarly if we need to calculate $A \sqcup R$. Since $R = B \sqcup C$ then $S \sqcup C = A \sqcup R$. So calculating the union of one of those relations will solve also the other. Same reasoning can be applied for the intersection.

Theorem 2: Suppose we have three organisations, A, B and C with A contained into B and B contained into C (thus $A < B < C$). And let R be a fourth organisation. If we want to calculate $B \sqcup R$; and we know that $A \sqcup R = C \sqcup R$ then $B \sqcup R = A \sqcup R$. In other words if we have two organisations, A and B and, when interacting with an external organisation R both will give a particular result S (that is $S = A \sqcup R = C \sqcup R$), then every other organisation in between A and C if united with R will be equal to S.

Theorem 3: Suppose we have two organisations, A, S with $A < S$ then $A \sqcup S = S$ then we can consider that given any other organisations B, such that $A \sqcup B = S$, for every organisation C such that $A < C < S$ then $C \sqcup B = S$. Symmetrically for every D such that $B < D < S$, $A \sqcup D = S$. But since $C \sqcup B = C \sqcup S$, and $B < D < S$ then $C \sqcup D = S$.

So once we calculate a single union we can fill in several entries of the Union Table. The same symmetric reasoning, with exactly the same dual theorem, can be applied for the intersection. We will now use those theorems to find the lattice of all organisations.

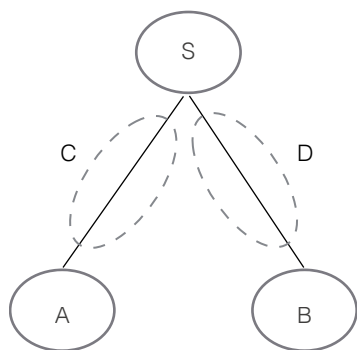


Figure 1: Theorem 3. If $A \sqcup B = S$, any organisation C between A and S united with any organisation D between B and S will always produce S. We do not need to calculate $C \sqcup D$.

Finding the Lattice of Organisations

Let M be a set of Molecules, with a reaction * such that $\forall a, b \in M: a*b \in M \cup \emptyset$. Let $\langle M, * \rangle$ be an Artificial Chemistry. Let $\langle O, \sqcup, \sqcap \rangle$ be the lattice of Organisations over $\langle M, * \rangle$.

We need to find all the organisations in O. Let us start by assuming that we have two basic organisations, the empty set

\emptyset , and M: $\emptyset, M \in O$. Those will be the top and the bottom organisations in the lattice. Every lattice with a finite number of elements has always a top element (the union of all the elements, also called the 1) and a bottom element (the intersection of all the elements, also called the 0). So we start by taking the bottom element as the organisation generated by the empty set, and the top element as the organisation generated by the set of all molecules.

One obvious solution to find all organisations is to check every single subset of M, which means checking $2^{|M|}$ subsets. If M is big this could be impractical or simply impossible. Let's look at ways in which we can exclude some sets without testing them. In fact without even listing them.

To reach O we are going to build a chain of sub-lattices of organisations, $N_0 < N_1 < \dots < N_m = O$, with N_0 the sub-lattice generated by the empty set and by the set of all molecules: $N_0 = \langle \{G_0(\emptyset), G_0(M)\}, \sqcup, \sqcap \rangle$. Note that given a set of organisations P, it is trivial to build a sub-lattice out of it just by taking the closure respect to \sqcup and \sqcap .

Thus once we have the sub-lattice N_i , and we find another organisation H, we can calculate N_{i+1} as $G_{\text{CUP}}(N_i \cup H)$. Where with $G_{\text{CUP}}(P)$ we indicate the set of all organisations that can be generated by recursively applying the organisation union, \sqcup , and the organisation intersection, \sqcap , between organisations in P and the organisations generated in this process.

When we are doing this we also store two tables, the Table of Unions of N_i : $T_{\sqcup, i}$, and the Table of Intersections of N_i : $T_{\sqcap, i}$. In general for every A, B in N_i , $T_{\sqcup, i}$ will store $A \sqcup B$ and $T_{\sqcap, i}$ will store $A \sqcap B$. And being N_i a sub-lattice, we know that both $A \sqcup B, A \sqcap B \in N_i$. Thus we know that if the tables $T_{\sqcup, i}, T_{\sqcap, i}$ are complete with every union and intersection calculated, and the result is an organisation known we do not have to proceed further, and we have found a sub-lattice. It is important that this process of calculating all the organisation union and intersection is made as efficient as possible. In this we are helped by the fact that O is a lattice, and thus we can apply the theorems listed above.

Adding one molecule at a time

Let us start by assuming we have a sub-lattice of organisations N_i , then $\forall B \in N_i$, we need to test all the molecules that are not in B, so $\forall e \in M \setminus B$, we need to study Be^1 . Calculating Be leads to 3 possible outcomes:

- 1) $Be = B$;
- 2) $Be = C$ with $e \notin C$ and $B < C$;
- 3) $Be = C$ with $e \in C$ and $B < C$.

In the first case we will say that e, in the context of B, goes down (case 1); goes by the side (case 2); or goes up (case 3).

So we go through all the molecules and we store for each molecule e, if in the context of B, e goes upward, downward, or sideward. We also store the generated organisation Be .

For each new organisation H that we find, we expand N_i into N_{i+1} by adding the organisation to the list of known organisations, and calculating the union and intersection closure. Of course, if $B \cup \{e\} = C$, with C already a known organisation, we don't need to test Be at all, and we know we are in case 3.

¹Note: In this paper we will indicate briefly $G_0(\{e\} \cup B)$ as Be , for every molecule e, and similarly if we need to test a set adding two molecules, e, f, we will indicate $G_0(\{e\} \cup \{f\} \cup B)$ as Bef .

cases		e goes		
		Upward	Sideward	Downward
f goes	Upward	1	2	3
	Sideward	2	4	5
	Downward	3	5	6

Table 2: Will adding two molecules at the same time to an organisation produce novelty? Depends on what each molecule does by itself. Six cases are possible.

Once we have finished exploring what happens by adding one molecule to a specific organisation, we continue with the next organisation, still maintaining the memory of all the molecules that added to B go up, on the side, or down. Also note that if Be goes sideways generating C , then Ce will go downward, so we do not need to test it. Once we have finished exploring for each known organisations (which included the new ones we might have found in the meantime), what happens when we add one molecule, we need to consider what happens when we add two molecules. This is where some shortcuts will be possible.

Adding more than one molecule at a time

Let us start by saying that if we have two sets of molecules x , y , and two organisations B and C , with $B < C$, we do not need to test Bx if we can prove that $Bx = Cy$. Because we will already consider the organisation generated by Bx , when we consider the organisation generated by Cy . Let us suppose we take an organisation B , and two molecules e, f in $M \setminus B$. Do we need to test Bef ? We need now to look at the combinations depending if e or f of upward, downward or sideward: here we find a symmetric matrix with 6 different cases (see table 2): If e goes up, than $Be = C$ thus for every f , $Bef = Cf$. So any organisation that would be found through Bef will be found through Cf . So if either e or f goes upward, we do not need to test Bef . This clarifies cases 1, 2 and 3. We will immediately state that we are going to test Bef in case that both e, f go downward. And this clarifies case 6.

cases		e goes		
		Upward	Sideward	Downward
f goes	Upward	✓	✓	✓
	Sideward	✓	✓	✓
	Downward	✓	✓	to test

Table 3: If we consider two molecules at the same time and add them to an organisation, this can produce novelty that would not be found through other ways only in one case.

We need now to explore case 4 and 5. Let us suppose that e goes sideward, while f goes downward. So we are in case 5; do we need to test Bef ? Since Be goes sideward, there exist an organisation C such that $Be = C$ with $e \notin C$ and $B < C$. So Bef will be equivalent to test Cef . And since $C > B$ than we will simply test this as part of testing C .

Finally let us consider the case where both e and f go sideward. In this case $Be = C$; $Bf = D$. Thus there exist an organisation H such that $C \sqcup D = H$. Now $Bef = Hef$, but $B < H$, so we do not need to test it as part of B . And the result is that we only need to test sets of molecules such that all the molecules in this set go downward (table 3).

To summarise: So far the algorithm proceeds as follows:

```

Let  $L_0$  be the starting sub lattice of organisations;
let  $O$  be the set of organisations known;
let Up, Down, Side be empty dictionaries;
for each organisation  $B$  in  $O$ :
  for each  $m$  not in  $B \cup \text{Down}[B] \cup \text{Up}[B] \cup \text{Side}[B]$ :
    calculate  $Bm$ ;
    if  $Bm = B$ :
      Down[B].add( $m$ ); continue;
    if  $Bm$  not in  $O$ :
      calculate  $L_{i+1}$  from  $L_i$ ,  $Bm$ , using the theorems above;
      expand  $O$  with  $L_{i+1} \setminus L_i$ ;
    if  $m$  not in  $Bm$ :
      for each organisation in  $L_{i+1}$ ,  $C$  between  $B$  and  $Bm$ :
        Up[C].add( $m$ )
        Down[Bm].add( $m$ )
        Side[B].add( $m$ )
      else: Up[B].add( $m$ )

```

At this point a sub lattice L_n will be discovered with a general structure with all the organisations that can be reached by adding one molecule at a time, and applying the union and intersection operations. This is not the complete lattice yet. For each organisation O we need to add the organisations generated by adding to O to every possible set S made up with molecules from $\text{Down}[O]$. In this case we can distinguish in four possible results (one more than before):

- 1) downward case: $OS = B$;
- 2) upward case: $S \subset OS$;
- 3) sideward case: $S \not\subset OS$, $S \cap OS = \emptyset$;
- 4) diagonal case: $S \not\subset OS$, $S \cap OS \neq \emptyset$.

The fourth case, the diagonal case, is the new one and happens when S is *partially* contained in OS . Again, as we build the sets S from the smallest to the biggest we do not need to check the organisation generated by any set that has a subset T such that $O < OT$. The proof follows the exact same structure as the proof above.

So given any organisation discovered we need to test it by trying to expand it with the molecules not in it, one by one. And with every subset of the “downward” molecules. While we do not need to test a set of “downward” molecules if it contains any subset which is upward, sideward, or diagonal. So the obvious thing to do, is to start with the smaller organisations, with the smaller subsets, and slowly build our way up.

The “No Organisation Left Behind” theorem.

We need now to prove that once the algorithm is followed, every organisation in the lattice of organisations will be found. This can easily be shown.

Suppose by contradiction that C is an organisation in the Lattice that has not been found with the algorithm. Suppose that B is the maximal organisation contained in C found by the algorithm. Such organisation is unique because if there were multiple maximal organisations, the union of them would also be contained in C and would be known because (as part of the algorithm) we are studying the union of all the known organisations. Similarly we assume that C is the minimal organisation unknown above B . So $B < C$, A maximal organisation known under C , C minimal organisation unknown above B . Now we need to show that following the algorithm we would discover C , against the hypothesis.

Let S be the set of molecules in C , but not in B ; $S = C \setminus B$. Thus obviously $BS = C$.

If we consider any subset T of S , $B < BT < C$. Since B is self-maintaining and C is closed, $B \leq G_0(BT) \leq C$.

Now, as part of the algorithm we explore one by one the subsets of S starting with the smaller one. For each subset T :

either T goes upward letting C be discovered;
or T goes sideways or diagonal. But in this case there would exist an organisation $D = BT$, with $B < D < C$. Now if D is known, it would be against the hypothesis that B is Maximal, and if it is unknown, it would be against the hypothesis that C is minimal. So there cannot be a D such that $D = BT$.

Thus for all T , T must go downward, eventually letting us test S and discovering C . Against the hypothesis. So this algorithm explores the full lattice of organisations.

A preliminary version of the code is available at (<https://github.com/pietrosperoni/LatticeOfChemicalOrganisations>)

Testing the Algorithm on the NTop

To test out the system it was used the NTop Artificial Chemistry (Banzhaf, 1993, 1994). This artificial Chemistry uses boolean vectors of size 4 as molecules, which are then folded into 2×2 matrixes, to react. This results are 16 reacting molecules. One of those acts as an algebraic zero ($\mathbf{0} * \mathbf{a} = \mathbf{a} * \mathbf{0} = \mathbf{0}$, for every \mathbf{a}) and it is usually eliminated. With the 15 remaining molecules it is possible to obtain 54 organisations out of a space of 32.768 possible subsets. The Brute Force algorithm tests all those subsets. Instead the algorithm

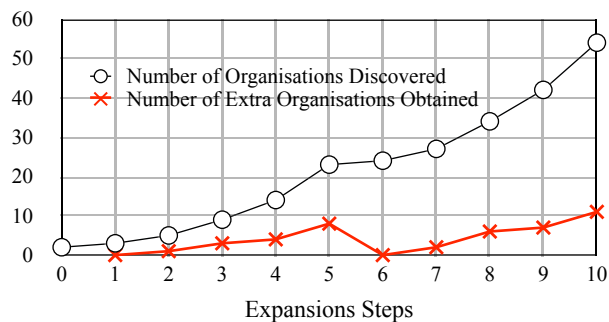


Figure 2: NTop Original: Number of organisations in each subsequent sub-lattice; the complete lattice, 54 organisations, was found in 10 subsequent expansions.

developed above was applied. The first step is to take two trivial organisations and the top, o_1 , and the bottom, o_2 , were taken. Then the bottom one was expanded, by adding one by one the 15 molecules. The first molecules led to o_3 . But the 3 organisations (o_1 , o_2 , and o_3) formed a sub-lattice so it was not possible to expand this further. Second, and third molecules also generated o_3 . The fourth molecule, added to o_2 , generated o_4 . But now it was possible to find o_5 as $o_5 = o_3 \sqcup o_4$. Again o_1 - o_5 formed a sub-lattice. The next molecule generated o_6 and permitted to find o_7 , o_8 and o_9 . o_{10} lead to o_{11} - o_{14} ; o_{15} all the way to o_{23} . So each organisation found would usually bring others with it, easily calculated. Once the o_1 - o_{23} sub-lattice of organisations was found, all the organisations, that could be found by adding a single molecule to o_2 had been discovered. Also each molecule tested was divided into downward, upward and sideward, thus simplifying the tests to do later on. Then the algorithm started expanding on those organisations. Expanding o_3 did not discover any new organisation. As did o_4 , o_5 , ..., o_9 . expanding o_{10} lead to o_{24} (and nothing else). Expanding o_{15} lead to o_{27} , then to o_{32} and finally to o_{42} . And then nothing else. At this point a sub-lattice was found where adding a single molecule to each of the known 42 existing organisation would always lead back to a known organisation. o_1 - o_{42} was not just any sub-lattice, but a sub lattice that could not be pierced by adding a single molecule at a time. Then the algorithm started adding 2 molecules at a time. Expanding o_6 it was possible to find o_{43} which then interacting with the other organisations generated o_{44} to o_{54} .

It was important to follow the algorithm, not just to understand it better, but because it showed a number of informations about the lattice. First and foremost the fact that the lattice has indeed a number of sub-lattices. Each new organisation found permit us to expand the space of the known organisations to the next sub-lattice in a chain that leads to the complete lattice. Although only 10 basic organisations calculated were necessary to generate the whole lattice, nothing tells us how to find those generators. Indeed finding a minimum number of generators, or just even any set of generators of organisations, is an open problem.

It was also interesting to count how many relations among organisations needed to be calculated, and how many could be derived from the theorems. The results (figure 2) suggests that as the number of organisations grows the number of organisations that needs to be calculated drops following a

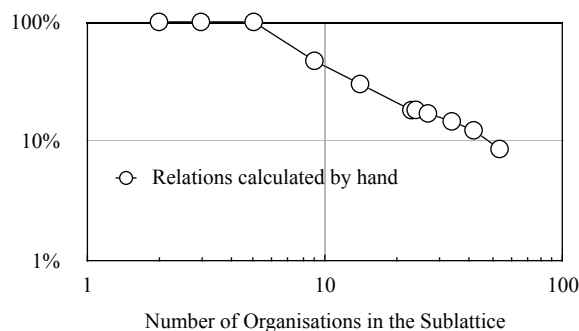


Figure 3: NTop Original: Percentage of Union or Intersection calculated by hand as opposed to extrapolated with the theorems.

power law (making a straight line on a log-log plot). While the remaining relations can all be derived theoretically.

Testing the Algorithm on the Expanded NTop

To try the algorithm on a more challenging artificial chemistry, it was applied to an expansion of the NTop. This time molecules of size 9 were used, which then were folded as 3x3 molecules. This gave 512 molecules. The folding can be done in 9! ways, and this both for the molecule on the right side and on the left side, thus producing 9! * 9! possible Artificial Chemistries. Then the resulting matrix will contain numbers between 0 and 3, and this will be mapped onto the set {0, 1}. This mapping can be done in 2⁴ possible ways. The resulting boolean 3x3 matrix must then be unfolded, and this also can be done in 9! ways. So in total there are 2⁴ * (9!)³ = 1.9 * 10¹⁷ possible Artificial Chemistries. Many of those chemistries produce only a trivial lattice of organisation. For example a lattice that would only contain few organisations, or where every set was an organisation.

In our case, if we consider the molecule

$a = [a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9]$,

the molecules was folded as

$$\begin{bmatrix} a_1, a_4, a_7 \\ a_2, a_5, a_8 \\ a_3, a_6, a_9 \end{bmatrix}$$

when it would react on the left side, and as

$$\begin{bmatrix} a_1, a_2, a_3 \\ a_4, a_5, a_6 \\ a_7, a_8, a_9 \end{bmatrix}$$

when it would react on the right side of the reaction. The result is then transformed according to the map:

$f(x) = \{0 \rightarrow 0; 1 \rightarrow 1; 2 \rightarrow 1; 3 \rightarrow 0\}$

and the resulting boolean matrix was unfolded so that from the matrix:

$$\begin{bmatrix} a_{1,1}, a_{1,2}, a_{1,3} \\ a_{2,1}, a_{2,2}, a_{2,3} \\ a_{3,1}, a_{3,2}, a_{3,3} \end{bmatrix}$$

the vector $[a_{1,1}, a_{1,2}, a_{1,3}, a_{2,1}, a_{2,2}, a_{2,3}, a_{3,1}, a_{3,2}, a_{3,3}]$ was produced. As with the NTop, the algebraic zero, $\mathbf{0} = [0,0,0,0,0,0,0,0,0]$ was excluded. And in this case also the molecule $\mathbf{1} = [1,1,1,1,1,1,1,1,1]$ was excluded. The result is an AC with 510 possible molecules, 2⁵¹⁰ possible sets of molecules; more than 10¹⁵³ sets to test. Obviously too many to directly test them all.

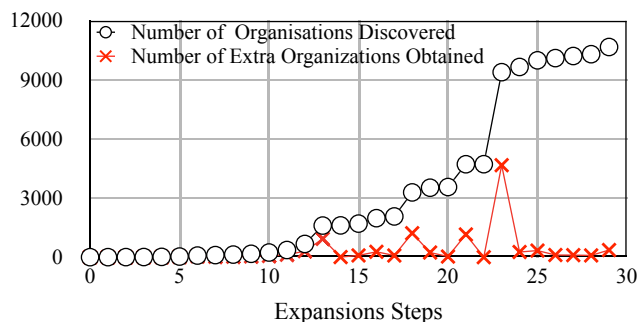


Figure 4: Number of organisations in each subsequent sub lattice.

Results

The algorithm described in this paper was applied. For now it was not possible to find all organisations.

As with the NTop the algorithm started with the lattice which just included the empty and the complete organisations. Then it expanded the set of organisations going through 29 sub-lattices of respectively of 3, 5, 9, 17, 33, 84, 107, 133, 173, 238, 365, 672, 1604, 1612, 1703, 1978, 2066, 3284, 3522, 3557, 4711, 4713, 9377, 9641, 10090, 10196, and 10288 organisations (figure 4). After which no new organisation was found expanding the empty organisation. So 10288 organisations were found just by adding 29 times a single molecule to the empty organisation. This created also two symmetric tables with all the intersection and unions. As the algorithm went on those tables were completed more and more using only the theorems. Again as the number of known organisations grew the percentage of organisations that needed to be calculated by hand decreased following a power law (figure 5). As such the more the lattice is known, the more powerful those theorems are to find the remaining organisations.

Of the 510 molecules 29 permit us to find all the organisations that could be generated by the empty set. Then those organisations started to be expanded themselves.

After the 10288 organisations sub lattice was found the algorithm tested one by one each of those organisations, and for the first 83 organisations, adding a single molecule would keep generating well known organisations. Then on the 84th organisation a molecule was added that expanded the sub lattice, from 10288 to over 69000. And then at 69779 the system could not handle the data using more than 150 gigabyte of RAM and crashed. It should be noted that in a space of 2⁵¹⁰ molecules it was realistically impossible to map the whole space of all the organisations.

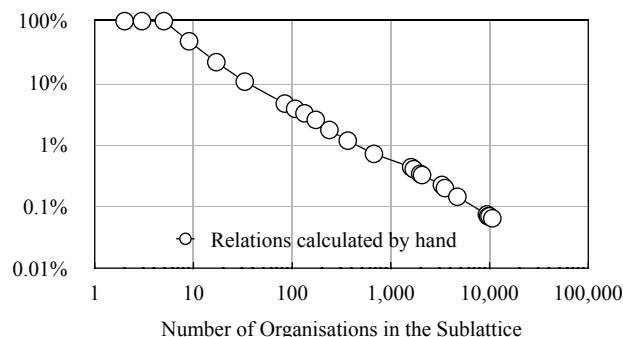


Figure 5: Percentage of Union or Intersection calculated by hand as opposed to extrapolated with the theorems.

Testing the Algorithm against the Brute Force

As a final test the Brute Force Algorithm was applied to this Artificial Chemistry. On a home laptop it only found 263 Organisations before crashing. And the average time to find each organisation was 42 seconds. While the home laptop could find 3000 organisations keeping an average of 0.2 seconds per organisation. In figure 6 the two averages are compared for the first 263 organisations. It should nevertheless be noted that both averages were growing

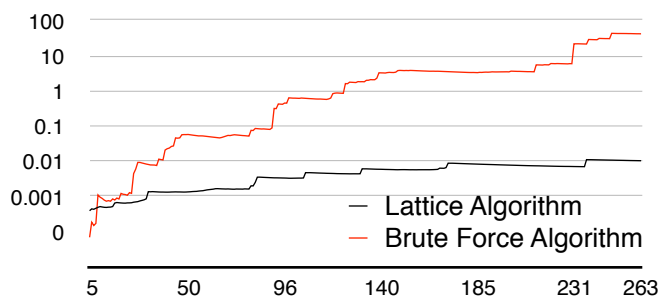


Figure 6: Average time needed to find an organisation, for the first 263 organisations as the number of organisations grows.

although the difference kept increasing. Also it is important to remember that the Brute Force only returns a set of organisations. No informations about the relative relations between the organisations, is returned. What organisations are above or below which others; what is the union or intersections of two organisations, is an information which is equally missing. All information that the Lattice Algorithm can easily return as it needs them to compute the lattice. So not only is the Brute Force much slower and generally inefficient algorithm. But it also is insufficient to really let us know an artificial chemistry.

Consequences and Conclusions

Reaction networks appear everywhere. And in their exploration the study of organisations, and their lattice is a necessary step to really understand their global behaviour as constructive dynamical system. The fact that organisations form a Lattice permit us to compute them faster. Although several results claim to be able to find all the organisations of a reaction network (Centler at al 2008, Centler at al 2010) they never used the algebraic properties of the Artificial Chemistry (namely that it is a lattice).

Another important aspects that was uncovered in this work is the concept of sub-lattice, as a subset of organisations such that each union and intersection is an organisation inside the sub-lattice. From mathematics we know that if we consider the space of all sub-lattices of a lattice, they form a lattice, too: the lattice of sub-lattices of the lattice of organisations. Such lattice is not explored here, instead we merely extract a chain inside such lattice of sub-lattices and use this to build the lattice of organisations. Interestingly on the NTop system we could identify a sub lattice of organisations that was closed respect to the operation of adding a single molecule to any organisation. This shows that if we must find all the organisations, we cannot simply consider adding 1 molecule.

But it also means that in some situations we are not interested in the complete lattice of organisations, but only in a sub-lattice since the the artificial chemistry will only explore that. For example in a Flow System both organisations and semi organisations form a lattice. But while each organisation is a semi-organisation the opposite is generally not true. Thus the lattice of organisations is a sub-lattice of the lattice of semi-organisation. And this fact could be used to map it.

Similarly, Artificial Chemistries are not the only case of lattice present in the fields of Bioinformatics, Artificial Life and

Systems Biology. Researchers looking at autocatalytic cycles and closed sets also are looking at sets of molecules that form lattices. Thus the same procedures that were exposed here, and the same theorems, can be applied over there, with comparable results.

Every research clarifies some elements, while opening new questions.

- In particular it is still unclear what is the most effective way to apply the lattice theorems to study the lattice of organisations. Yes, theorems can shortcut the calculations, and we could see that the algorithm was at least three orders of magnitude faster than the Brute Force algorithm (and the difference was increasing), but finding which theorems can be applied can be time consuming as well. So a smart strategy might need to be applied to chose when to try to apply the theorems, as a further improvement of the algorithm.

- Also it is unclear why the number of times the theorems are not applicable follows so closely a power law. This might be related to the nature of the Lattice of Organisations as a graph. But the details are still missing.

- The roles of sub-lattices, what is the sub-lattice of sub-lattices, and how the sub-lattices of the lattice or organisations can be used to study an ecology of different experiments on one artificial chemistry is also an open question.

- And finally the artificial chemistry presented here is very vast and exploring it all is at the moment impossible. It could as such be used as a benchmark for future work.

Acknowledgement: The author acknowledges support from the European Union through funding under FP7-ICT-2011-8 project HIERATIC (316705). Also thanks Peter Dittrich for personal remarks and the two anonymous reviewer for fundamental comments.

References

- Banzhaf W. (1993) Self-replicating sequences of binary numbers. Foundations II: Strings of length N- 4 *Biol. Cybern.*, 69, pg 275-281
- Banzhaf W. (1994) Self-replicating sequences of binary numbers. Foundations I: The Built Up of Complexity. *Complex Systems*, 8, pg 215-225
- Centler, F, Kaleta C., Speroni di Fenizio, P; Dittrich, P (2008) Computing chemical organizations in biological networks. *Bioinformatics* 24 (14): 1611-1618.
- Center, F.; Kaleta C.; Speroni di Fenizio, P.; Dittrich, P. (2010) A parallel algorithm to compute chemical organizations in biological networks. *Bioinformatics Applications Note* 26 (14), pg 1788 - 1789
- Dittrich, P., Speroni di Fenizio, P (2007). Chemical organization theory. *Bull. Math. Biol.*, 69(4): pg 1199-1231
- Fontana, W. and Buss, L. W. (1994). 'The arrival of the fittest': toward a theory of biological organization. *Bull. Math. Biol.*, 56: pg 1-64
- Fontana, W. and Buss, L. W. (1994). What would be conserved if 'the tape were played twice'?. *Proc. Natl. Acad. Sci. USA*, 91: pg 757-761
- Speroni di Fenizio, P.; (2007). *chemical organizations theory*. Doctoral thesis, Chair of Systems Analysis, Department of Computer Science, University of Jena, Germany, 2007
- Fontana, W. and Buss, L. W. (1996). The barrier of objects: from dynamical systems to bounded organizations. Casti J. and Karlqvist A., editors, *Boundaries and Barriers*, pg 56-116, Addison-Wesley