# Stochastic Process Algebra and Stability Analysis of Collective Systems

Luca Bortolussi, Diego Latella, Mieke Massink

# Stochastic Process Algebra and Stability Analysis of Collective Systems

Luca Bortolussi[1,2,⋆], Diego Latella[2], and Mieke Massink[2]

Dipartimento di Matematica e Geoscienze, University of Trieste, Italy[1]
Istituto di Scienza e Tecnologie dell'Informazione 'A. Faedo', CNR, Italy[2]

**Abstract.** Collective systems consist of large numbers of agents that coordinate through local behaviour, adapt to their environment and possibly give rise to emergent phenomena. Their formal analysis requires advanced scalable mathematical approximation techniques. We show how Stochastic Process Algebra (SPA) can be combined with numeric analysis tools for the analysis of emergent behavioural aspects of such systems. The approach is based on an automatic transformation of SPA models into ordinary differential equations in a format in which numeric and symbolic computing environments can be used to perform stability analysis of the system. The potential of the approach is illustrated by a crowd dynamics scenario in which various forms of behavioural and topological asymmetry are introduced. These are cases in which analytical approaches to stability analysis are in general not feasible. The analysis also shows some surprising aspects of the crowd model itself.

**Keywords:** Fluid flow, process algebra, crowd dynamics, self-organisation.

## 1 Introduction

A key factor to allow modern cities to reach or maintain a good and sustainable quality of life for their increasingly numerous inhabitants is the development of systems that are relying on a much more decentralised and distributed design that is adapting itself to dynamically changing circumstances [19]. Examples of such future systems are electricity grids that can cope with many local electricity producers and consumers, and a decentralised organisation of transportation and information. Such large scale collective adaptive systems rely on the continuous feedback between vast numbers of participants of different kinds, and, as is well-known, can be expected to show complex dynamic and emergent behaviour or perhaps even exploit such behaviour [8]. Also the formal analysis of such systems poses many new research challenges.

Process algebras have been specifically designed for the compositional and high-level modelling and analysis of distributed concurrent systems. Recently some of them, in particular PEPA [10] and Bio-PEPA [7], have been provided with a fluid semantics based on ordinary differential equations (ODE) [11] providing a scalable approach to the analysis of agent coordination in large collective

⋆ Work partially supported by the project "FRA-UniTS".

systems. In this paper we further exploit this development to study stability aspects of collective adaptive dynamic systems in a symbolic and numeric way. Stability analysis provides important information about the predictability of dynamic systems and their sensitivity to parameter values. Numeric stability analysis is of particular interest for the analysis of distributed adaptive strategies that are applied in *asymmetric* situations. Such situations occur naturally in many real-world, natural or designed, systems. An analytical approach to stability analysis in the presence of asymmetry is infeasible in most cases. To the best of our knowledge this is the first time that a method is proposed for stability analysis that starts from a stochastic process algebra specification of agent coordination in a collective dynamic system. The use of a process algebra greatly facilitates the modelling of variants of a system at a high level. Designers can then focus on the coordination strategies instead of having to manipulate the underlying, possibly large set of non-compositional, ODEs in ad-hoc ways.

We illustrate the approach and related tool chain by analysing a new variant of a collective model of spontaneous self-organisation of drinking parties in the squares of cities in Spain, also known as "El Botellón" [21]. In this variant the parameter of the system is the *level of socialisation*, i.e. the average number of friends people have, instead of the the probability to find a partner to chat with[1]. The model shows a number of surprising behavioural aspects. However, the main contribution of the paper is to illustrate and explore what stochastic process algebra can offer to provide high-level models of coordination in large scale collective systems, in combination with well-established numeric and symbolic analysis techniques for a systematic stability analysis of such systems. In particular in case of models showing various forms of asymmetry.

Related work can be found in several directions. In [15] the original model of Rowe and Gomez was analysed with Bio-PEPA. Both empirical and analytical justification was provided for the good correspondence found between stochastic simulation results and the Bio-PEPA based fluid flow approximation. Moreover, a comparison of analytical results with those obtained numerically via Bio-PEPA was provided. For this reason the focus was on symmetric models that can be handled analytically. It does not include a systematic study of the stability aspects of the model. In [5] a variant based on the socialisation level has been analysed in an analytical and numerical way. The study addresses a stability analysis of a *symmetric* model limited to three squares, but it contains no stability analysis of *asymmetric* models. Recently, the use of Bio-PEPA has also been explored in the field of swarm robotics, where it was used to model a swarm decision making strategy [16, 17]. This collective decision-making strategy has been used as a benchmark for the application of stochastic process algebra, and in particular Bio-PEPA with locations, in this new field of application. It was shown that important aspects of swarm robotics can be addressed such as cooperation and space-time characteristics, but also emergent behaviour.

In the following we first introduce the crowd model and Bio-PEPA followed by a description of the tool pipeline and its application.

[1] Studied in the original work by Row and Gomez [21] and in [15]

## 2 A Socialisation Level Based Crowd Dynamics Model

Rowe and Gomez [21] analyse the movement of crowds between four city squares, connected in a ring by streets, using an agent based approach. The movement of individual people is simulated by agents following a simple set of rules. At every step each agent tries to find a partner to chat with. If this succeeds it stays where it is; else, it moves randomly to an adjacent square. It is assumed that the probability of the latter is $(1-c)^{p_i-1}$ when this agent is at square $i$, and $p_i > 0$ is the number of agents currently at square $i$. The parameter $c$ (representing the *chat probability*, $0 \leq c \leq 1$) is the probability that an agent finds a partner to talk to and thus remains in the square. Rowe and Gomez showed that the emergence of crowds in their model is directly linked to a critical threshold value of the chat probability. If the value is below the threshold, the population remains evenly distributed over the squares, while walking around. If the value is above the threshold, the population tends to gather into one or a few squares. However, the probability to meet a friend in a crowded city is in general not the same as the probability to find a friend when it is less crowded. People tend to have a fixed number of friends given a city population, and the larger the number of people walking around, the more of them will turn out not to be one of your friends. This consideration leads to an alternative crowd model, introduced in [5], in which the chat probability is defined as $c = s/N$, where $N$ is the size of the population and $s$ is the level of socialisation of the population, i.e. the average number of friends that people have. Using this alternative definition of $c$ and an $n \times n$ routing matrix $Q$ for $n$ squares, i.e. $Q_{i,j}$ is the probability that a person moves to square $j$, given that she decided to leave square $i$, the ODE for population level $N$ of this model is:

$$\frac{dx_i}{dt} = -x_i(1-s/N)^{Nx_i-1} + \sum_j x_j(1-s/N)^{Nx_j-1}Q_{j,i} \tag{1}$$

where $x_i$ denotes the fraction of the population that is in location $i$. Here we assume that $Q$ is symmetric, i.e. $Q_{i,j} = Q_{j,i}$, so $Q$ is a stochastic and symmetric matrix. It is further assumed that $Q$ is irreducible (this is not a limitation since otherwise the city can be split into its connected components.) The above ODE is also the fluid flow interpretation of a Bio-PEPA model of this scenario that will be presented in Section 4 with the only difference that the latter is defined directly on the population sizes itself and not on their fractions. The basic symmetric model has also an interesting fluid limit, i.e. an ODE model resulting from letting $N \to \infty$. For $N$ going to $\infty$ one obtains (see [5] for an analytical derivation):

$$\frac{dx_i}{dt} = -x_i e^{-sx_i} + \sum_j x_j e^{-sx_j} Q_{j,i} \tag{2}$$

This is a non linear ODE, that has all its solutions in the unit simplex $\Delta_n = \{x \in \mathbb{R}^n : x_i \geq 0 \text{ and } \sum_i x_i = 1\}$ if the initial condition is in $\Delta_n$.

## 3 Bio-PEPA and Fluid Flow Analysis

Before presenting a process algebra model of the crowd scenario we briefly recall Bio-PEPA [7], a language that has originally been developed for the modelling and analysis of biochemical systems. The main components of a Bio-PEPA system are the *"species" components*, describing the behaviour of individual entities of a species, and the *model component*, describing the interactions between the various species. The initial amounts of each type of entity or species are given in the model component. The syntax of the Bio-PEPA components is defined as:

$$S ::= (\alpha, \kappa)\ \mathtt{op}\ S \mid S + S \mid C \text{ with } \mathtt{op} = \downarrow \mid \uparrow \mid \oplus \mid \ominus \mid \odot \text{ and } P ::= P \underset{\mathcal{L}}{\bowtie} P \mid S(x)$$

where $S$ is a *species component* and $P$ is a *model component*. In the prefix term $(\alpha, \kappa)\ \mathtt{op}\ S$, $\kappa$ specifies the multiples of an entity of species $S$ involved in an occurring action $\alpha$[2]. The *prefix combinator* "op" represents the role of $S$ in the action, or conversely the impact that the action has on the species. Specifically, $\downarrow$ indicates a reduction of the population involved in the action, $\uparrow$ indicates an increase as a result of the action. The operators $\oplus$, $\ominus$ and $\odot$ play a role in an action without leading to increments or decrements in the involved populations and have a defined meaning in the biochemical context. Except from $\odot$, that will play a role in the shorthand notation introduced below, we will not need these operators in this paper. The operator "+" expresses the choice between possible actions, and the constant $C$ is defined by an equation $C{=}S$. The process $P \underset{\mathcal{L}}{\bowtie} Q$ denotes synchronisation between components $P$ and $Q$, the set $\mathcal{L}$ determines those actions on which the components $P$ and $Q$ are forced to synchronise, with $\bowtie$ denoting a synchronisation on all common actions. In $S(x)$, the parameter $x \overset{*}{\in} \mathbb{R}$ represents the initial amount of the species. The frequency with which an action occurs is defined by its (functional) rate. This rate is the parameter of a negative exponential distribution. Its value may be a function of the size of the populations involved in the interaction.

Bio-PEPA comes with a notion of discrete locations that may contain species. A Bio-PEPA *system* with *locations* consists of a set of species components, also called sequential processes, a model component, and a context (locations, functional rates, parameters, etc.). The prefix term $(\alpha, \kappa)\ \mathtt{op}\ S@l$ is used to specify that the action is performed by $S$ in location $l$. The notation $\alpha[I \to J] \odot S$ is a shorthand for the pair of interactions $(\alpha, 1)\downarrow S@I$ and $(\alpha, 1)\uparrow S@J$ that synchronise on action $\alpha$. This shorthand is very convenient when modelling agents migrating from one location to another as we will see in the next section. Bio-PEPA is given an operational semantics, based on Continuous Time Markov Chains (CTMCs), and a fluid semantics, based on ordinary differential equations (ODE) [7]. The Bio-PEPA language is supported by a suite of software tools which automatically process Bio-PEPA models and generate internal representations suitable for different types of analysis [7, 6]. These tools include mappings from Bio-PEPA to differential equations (supporting a fluid flow approximation), stochastic simulation models [9], and PRISM models [14].

[2] The default value of $\kappa$ is 1 in which case we simply write $\alpha$ instead of $(\alpha, \kappa)$.

## 4 Bio-PEPA Crowd Model

In this section we define a Bio-PEPA specification of the crowds scenario presented in Section 2. Let us consider a small ring topology with 4 city squares in a $2 \times 2$ grid, denoting them by 00, 01, 10 and 11, allowing bi-directional movement between squares. In Bio-PEPA the city squares are modelled as locations called $sq00, sq01$, $sq10$ and $sq11$. Parameter $c$ defines the chat-probability and parameter $d$ the degree or number of streets connected to a square. In the symmetric topology $d = 2$ for each square. The chat-probability is defined as the fraction of the socialisation factor $s$ w.r.t. the total population $N$, i.e. $c = s/N$. The actions modelling agents moving *f*rom square X *t*o square Y are denoted by $fXtY$. The associated functional rate for *f00t01* with $P@sq00$ denoting the population in $sq00$ at time $t$ is defined as:

$$f00t01 = (P@sq00 * (1 - c)^{(P@sq00-1)})/d;$$

the other rates are defined similarly. The behaviour of a typical agent moving between squares is modelled by sequential component $P$. For example, $f00t01[sq00 \rightarrow sq01] \odot P$ models that an agent present in $sq00$ moves to $sq01$ according to the functional rate defined for the action *f00t01*.

$$\begin{aligned} P = {} & f00t01[sq00 \rightarrow sq01] \odot P + f01t00[sq01 \rightarrow sq00] \odot P + \\ & f00t10[sq00 \rightarrow sq10] \odot P + f10t00[sq10 \rightarrow sq00] \odot P + \\ & f01t11[sq01 \rightarrow sq11] \odot P + f11t01[sq11 \rightarrow sq01] \odot P + \\ & f01t11[sq10 \rightarrow sq11] \odot P + f11t10[sq11 \rightarrow sq10] \odot P; \end{aligned}$$

Finally, the model component defines the initial conditions of the system, i.e. in which squares the agents are located initially, and the relative synchronisation pattern. If, initially, there are 1000 agents in $sq00$ this is expressed by $P@sq00[1000]$. The fact that moving agents need to synchronise follows from the definition of the shorthand operator $\rightarrow$.

$$((P@sq00[1000] \underset{*}{\bowtie} P@sq01[0]) \underset{*}{\bowtie} (P@sq10[0])) \underset{*}{\bowtie} (P@sq11[0])$$

The total number of agents $P@sq00 + P@sq10 + P@sq01 + P@sq11$ is invariant and amounts to 1000 in this specific case. The occupancy measure, i.e. the fraction of the population in $sq00$ can be defined as $Psq00 = P@sq00/N$ and similarly for the other squares. The fluid semantics of Bio-PEPA leads to an ODE that is very similar to Eq. (1) except that it is defined on the actual population sizes rather than their fractions:

$$\frac{dP@sq_i}{dt} = -P@sq_i(1 - s/N)^{P@sq_i - 1} + \sum_j P@sq_j(1 - s/N)^{P@sq_j - 1} Q_{j,i} \quad (3)$$

Using the Bio-PEPA plugin tool suite a first insight in the behaviour of the above model for different values of the socialisation factor $s$ can be obtained using e.g. stochastic simulation [9] or one of the built-in ODE solvers. For example, for $s = 5$ an interesting so called 'metastable' behaviour can be observed in

ODE trajectories such as the one shown in Fig. 2(a), where the fractions of the population present in squares $sq00$ through $sq11$, are denoted by $x_i$, for $i \in \{1, .., 4\}$. Until time 150 just over 40% of the population is in each of the non-adjacent squares $x_1$ and $x_3$, and slightly less than 10% in each of the remaining squares. Then suddenly this situation changes and square $x_1$ gets almost all of the population. This is just one example of the typical kind of behaviours that may occur in non-linear systems such as these.

To get a more complete overview of potential emergent behaviour of a collective dynamic system it may be useful to construct a bifurcation diagram of the system. This is a diagram that shows for each value of a selected parameter of the model its stationary points for that value. For each stationary point it also shows whether it is stable or unstable, i.e. whether a system would remain in a state forever once it is reached, or whether it could still move on from there reaching other states. The selected parameter of interest in our case is the socialisation value $s$. Fig. 1 shows the bifurcation diagram for square $x_1$ of the crowds model with four squares and for $s$ ranging from 2.75 to 6. For example, for $s = 3$ we see that the model has one stable stationary point with value 0.25. This means that for $s = 3$ the system ends up in a stable state in which 25% of the population is expected to be in square $x_1$.

In fact, the model has a stationary point at 0.25 for all values of $s$ considered, i.e. the vector of four squares $\mathbf{x}_{sym} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ is always a stationary point, but it is not stable for all values of $s$; its stability changes at $s = 4$. The stability of this equilibrium has been analysed in an analytical way in [5] for a model with an arbitrary number $n$ of squares, but with symmetric routing matrix $Q$. There it was proven that the stability status of this stationary point changes from stable into unstable when $s$ becomes equal to $n$. It is stable for models in which $s < n$ and unstable for those in which $s > n$. However, in [5] no analytic results have been given for the other stationary points, due to the difficulty in identifying them in general. Stability of all fixed points has been discussed in [5] only for a symmetric, fully connected, model with three squares. Such difficulties are also common for models that show irregular structure or other forms of asymmetry. When investigating the behaviour of the model in more detail using stochastic simulation (not shown) it turns out that in the model with 4 squares and $s < 3.25$ the population is migrating between squares in such a way that it is evenly distributed over the squares. This means that at any point in time one would see approximately 25% of the population in each of the squares. However, this situation changes for $s > 3.25$. Stochastic simulation shows that in that case most of the population eventually gathers at random in one of the squares.

In the following we propose an analysis pipeline, starting from a Bio-PEPA specification, that can be used to compute the stationary points *numerically* instead of analytically for various values of $s$, such that a bifurcation diagram of the kind shown in Fig. 1 can be generated. As we can see in the figure, the change of stability of $\mathbf{x}_{sym}$ is correctly identified by the numerical method, which also predicts other stationary points, corresponding to three different configurations:

1. Most agents in a single square. This is the top branch in the diagram. These are stable stationary points, but emerge for $s$ greater than 3.5, approximately. For instance, for $s = 5$, in this case 98% of agents are in a single square.
2. Agents evenly split between two opposite squares. This is an unstable stationary point, emerging for $s > 4$, which is in fact a saddle node [3]. For $s = 5$, this configuration corresponds to 43% of agents in two opposite squares (for instance, $sq00$ and $sq11$), and 7% in the remaining ones. Notice that this information can be deduced by observing the numeric values of fixed points.
3. Agents evenly split between three adjacent squares. This is again an unstable equilibrium of the system. for $s = 5$, we have 29% of agents in each of the three adjacent squares and 0.13% in the remaining one.

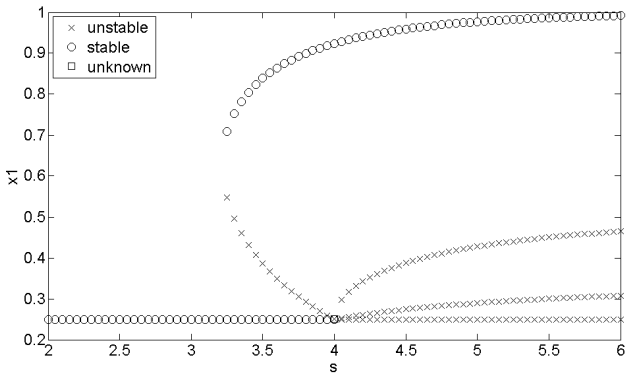
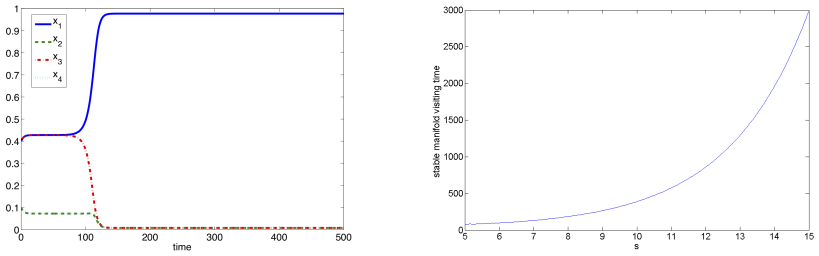



**Fig. 1.** Bifurcation diagram for the symmetric 4 square model.

Interestingly, configuration (2) above, despite being a saddle node, hence unstable, has a quite strong stable manifold. The effect is that solutions starting nearby the stable manifold (e.g. close to the plane $x_1 = x_3$) remain for some time close to it, before escaping to one of the stable fixed points. This gives rise to the typical metastable behaviour that is shown in the ODE trajectory in Fig. 2(a). As could be expected, the higher the socialisation coefficient $s$, the more intense is the attractive behaviour of the stable manifold, hence the longer solutions of the ODE of the model, that start from initial values in the neighbourhood of such a point, remain attracted to it. This can be seen in Figure 2(b), where the time spent nearby the stable manifold is shown as a function of $s$. In Section 6 we will apply the analysis pipeline proposed in the next section to several variants of the crowd model. Each of these variants is characterised by a different form of asymmetry e.g. caused by squares with different attractivity, or by asymmetry in topology.

[3] Unstable stationary points come in different kinds. One of these kinds is called a 'saddle node', informally it is attracting from two opposite sides and repellent from the two other sides of the point.




(a) Metastable trajectory (b) Metastable equilibrium visiting time

**Fig. 2.** (2(a)) A trajectory for $s = 5$ showing metastable behaviour. The initial conditions are $x_1 = 0.4 + \delta$, $x_2 = 0.1$, $x_3 = 0.4 - \delta$, $x_4 = 0.1$, with $\delta$ small. (2(b)) Time spent close to the metastable equilibrium, as a function of $s$. Initial conditions are as above, with $\delta = 10^{-7}$.

## 5 Numeric/Symbolic Stability Analysis

Before discussing the tool pipeline, we briefly sketch the methods we are using in Octave/Matlab to investigate the stability behaviour of fixed points and to generate bifurcation diagrams for different asymmetric variants of the crowd model that will be presented in next section. The idea, as discussed previously, is to combine the features of stochastic process algebras as modelling languages, Bio-PEPA specifically, with the numerical algorithms and analysis routines of a platform like Matlab [18] or Octave [1].

Assume that we have obtained an ODE function in the input format of Matlab or Octave, i.e. an m-file. Such an ODE function models the changes in the behaviour of the system over time. The idea is then to use numerical routines to look for the set of zeros of this function, i.e. points for which the system is stationary, for a fixed set of parameters (e.g. in our case for the socialisation coefficient $s$). More precisely, we can use the Octave/Matlab method *fsolve* which looks for zeros of non-linear equations and which incorporates different solvers (in our case we used the Levenberg-Marquardt algorithm [20]). As *fsolve* finds one zero, depending on the initial values, we run the algorithm many times to look for more zeros, starting from random initial coordinates (and from some predefined fixed points, like the unit vectors). In order to avoid considering all possible symmetric solutions of the problem, we post-process the set of zeros found in this way, to keep only one zero among those equivalent under symmetry. For instance, in a fully symmetric model with ring topology as the one in Section 4, there is a zero which puts almost all the population in one square, for each square. The post-processing then removes all but one of those zeros. Finally, for each such stationary point, we investigate its stability by computing numerically the Jacobian matrix and its eigenvalues, according to standard methods in dynamic systems theory [22, 13]. After discarding one zero eigenvalue (which is always present due to the conservation of the total number of agents), we check whether

the real part of the remaining ones is positive or negative. If all eigenvalues have a negative real part (and are sufficiently far away from zero to account for numerical errors), we declare the point stable, otherwise we mark it unstable. If there is an eigenvalue too close to zero, we mark its status as unknown. In order to generate bifurcation diagrams, we perform this operation for a range of values of the parameter of interest, e.g. in this case the socialisation coefficient $s$.

*Tool Pipeline.* In order to link the above described procedure to Bio-PEPA, we export the model in the SBML format [4]. Such an export is already available via the Bio-PEPA plugin tool suite [6]. Both Matlab and Octave have a toolbox importing from SBML files [12], generating an m-file computing the vector field (ODE) corresponding to the fluid semantics of Bio-PEPA [7]. Once an m-file has been obtained, it can be used within the routines explained above by creating a function handle. However, following the above procedure, we have not yet obtained the limit model discussed in Section 2, as from Bio-PEPA we export the $N$-dependent model (see Sect. 4). Even if the two sets of ODE will produce very similar solutions when $N$ is large enough, working with the limit model would be preferable, as this seems to reduce the numerical errors caused by the exponentiation. A limit model can be obtained by exploiting a computer algebra system such as the symbolic toolbox in Matlab [2] or the open source software Maxima [3]. This requires that the Bio-PEPA specification is exported to an m-file containing a symbolic definition of the ODE equations in Matlab or an equivalent format suitable for use with Maxima. This is not very difficult to automatise. Once this operation is performed, the symbolic calculus routines can be exploited to compute the limit of the vector field and to compute the Jacobian matrix symbolically, increasing the precision of the method and speeding up the numerical analysis phase. The resulting functions can be either exported from Maxima to Matlab/Octave by a suitable script that generates an appropriate m-file, or by converting a symbolic function into a numerical one in Matlab.

## 6 Results

In this section we present some results for variants of the basic crowd model enriching it either by adding new behaviours or by breaking the symmetry between squares. Due to space limitations, we will mainly report on results for a model with four squares connected in a ring topology. However, we have also obtained interesting results for a larger a model with 9 squares disposed in a 3 $\times$ 3 grid-like topology. The modifications in the basic model are essentially of four types:

– Breaking the symmetry between squares (in a symmetric topology like the ring one) by assigning to each square an uneven attractiveness coefficient. Each agent then chooses the next square to go according to their relative attractiveness.

- Breaking the symmetry in the routing, by assuming that certain connections between squares can be crossed only in a given direction, i.e. by introducing one way pedestrian streets (which can be enforced for instance by the presence of police).
- Breaking the symmetry in the topology, by having different degrees of connection between different squares, e.g. the 9 squares model (not shown).
- Breaking the symmetry in the behaviour of agents, assuming that an agent first decides if she wants either to leave the square or look for somebody to chat to. In this second case, we assume that she leaves the square with the same probability as in the standard model.

In particular, in each of these cases, we will discuss the stability of stationary points as a function of the socialisation parameter $s$, or of other parameters, like attractiveness or the probability to leave.

As discussed in Section 5, the results always consider the limit fluid model (see Eq.(2)), for a population going to infinity. Due to convergence of vector fields, the results for the limit model can be expected to be the same as those of the ODE with explicit dependence on $N$, if $N$ is reasonably large. In particular, we have seen basically no differences between results for the limit model and those for a model with a total population of $N = 1000$.

**4 square model with attractive squares** The first kind of asymmetric model we consider is one in which an attractiveness coefficient is assigned to each square. Higher attractiveness is modelled by a higher value of the coefficient. When agents decide to leave a square, their decision to which of the adjacent squares to go is now proportional to the relative attractiveness of the adjacent squares. In particular, we consider a situation in which only one square (by convention, square $sq00$), has a higher attractiveness than the others. Hence, the attractiveness coefficient of square $sq00$ is equal to $a \geq 1$, while that of the other squares is set to 1. In Bio-PEPA this is modeled by replacing the transition rates in the symmetric model of Sect. 4 by the following ones, assuming attractiveness coefficients $a00$ for square $sq00$, $a01$ for square $sq01$ and so on, e.g.:

$$f00t01 = (P@sq00 * (1-c)^{(P@sq00-1)}) * (a01/(a01+a10));$$

In Fig. 3 we show a bifurcation diagram as a function of the socialisation factor $s$, for the fraction of people in the first square, $x_1$, for a model with $a = 3$. As we can see, the solution in which the more attractive $sq00$ gets the larger amount of people is always stable. However, for $s$ around 3.5, a new set of stationary points appears, with a stable and an unstable branch. The stable branch corresponds to situations in which most of the people stay in one of the two squares adjacent to the attractive one. This stable equilibrium is quite surprising, as one would expect that people always move towards the attractive square. Note that this happens for $s < 4$, i.e. the predicted threshold for $s$ in the symmetric model.

The other stable equilibrium that emerges around $s = 4.5$ corresponds to the situation in which all agents are in the square opposite to the attractive one.

Hence, even in the presence of asymmetric attractiveness, we may still obtain a pattern of emergent behaviour in which all people are gathered at a random single square. So, contrary to intuition, for sufficiently large values of $s$, people may concentrate also in non-attractive squares. Furthermore, as expected, we loose the symmetric equilibrium in which agents are uniformly distributed over the squares and find no replacement for it.

**4 square model with one-way streets** In this model we assume that streets between squares are unidirectional rather than bi-directional. In particular, we assume that the ring can be traversed only clockwise, with no square having a higher attractiveness than others. This model is still ergodic, and moreover the four squares still behave symmetrically. However, the routing matrix is no longer symmetric. In the Bio-PEPA model this can be easily obtained by removing the related directions of movement and their corresponding rate definitions.

The pattern of the bifurcation diagram we obtain is very similar to the one for the symmetric model in Fig. 1. Indeed, the only difference is that now some eigenvalues of the Jacobian in steady states have imaginary values, suggesting that we may in fact have stable and unstable foci rather than single nodes [22], i.e. convergence to the fixed point happens by damped oscillations. In any case, this effect is very weak, and cannot be observed at the resolution scale at which we plot trajectories.

**4 square model with independent leaving probability** In this model we change the behaviour of single agents. In particular, we assume that each agent first chooses if it wants to leave the square (with a "boredom" probability $p$) or look for another one to chat with. In the latter case, it behaves like in the original model. In Bio-PEPA this is modeled by replacing the transition rates in the model with attractiveness by:

$$f00t01 = (1-p) * (P@sq00 * (1-c)^{(P@sq00-1)}) * (a01/(a01+a10)) + p * (P@sq00);$$

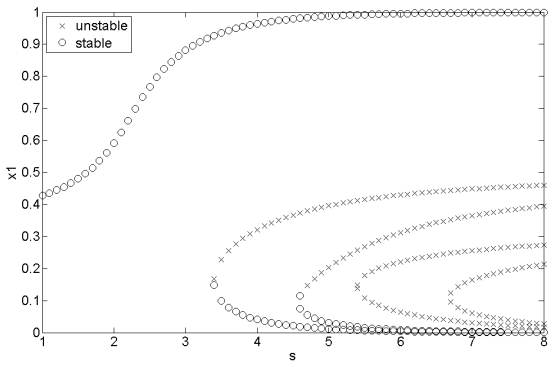



**Fig. 3.** Bifurcation diagram for the 4 square model with asymmetric attractiveness, as a function of $s$, holding $a$ fixed to 3 .

Note that in case all squares have an attractiveness coefficient equal to 1 we can study the effect of the boredom probability separately from the effect of attractiveness of squares. This is what is assumed in the following.

The model with independent leaving probability has a second parameter in addition to $s$, namely the boredom probability $p$. Intuitively, this probability should influence the overall behaviour of the model quite radically. If it is large enough (maybe in case of a meeting of eremites), then we do not expect people to gather in a single square, but rather to see them uniformly distributed over the four squares while moving between them. This indeed happens for $p = 0.15$ and for any value of $s$ considered. On the other hand, for small values of $p$, the behaviour manifested by the system becomes much more complicated, as shown in Fig. 4. In this case, for $p = 0.05$, we can observe a frequent change in the stability status of fixed points, due to bifurcation events in which stationary points split, change stability status, and move towards other fixed points, generating a cascade of bifurcations.

The most interesting feature is the stability of the symmetric equilibrium. Initially it is stable, but then, as $s$ increases, it undergoes a bifurcation event and becomes unstable (for $s$ around 5 here), as in the original model. At the same time, a new stable branch emerges, corresponding to the emergent behaviour in which most people are concentrated in a single square. However, for $s$ around 12, this stable behaviour undergoes a new bifurcation event and becomes unstable. Furthermore, as $s$ increases further, this unstable equilibrium approaches the symmetric one and hits it around $s = 16$. When this happens, the unstable symmetric equilibrium becomes stable again, and as $s$ increases even further, it becomes the only (stable) fixed point. This is a counter-intuitive behaviour of the model: as the socialisation factor increases, instead of obtaining a higher probability of people concentrating in a single square, exactly the opposite effect emerges.

The assumption that agents can leave a square with a small fixed probability, no matter whether friends are present or not, is quite reasonable. For example one may receive a text message from a friend in another square or have other things to do. The behaviour of this model shows that, when the socialisation coefficient is large enough (in reality, for a population of a few thousand people, we can expect a value of $s$ well above 30), the chat probability mechanism is not able to fully explain the emergent behaviour of 'El Botellon'. Other mechanisms have to be taken into account and also their potential interference. Among these are most likely some asymmetry breaking phenomena such as the different degrees of attractiveness of the squares.

To obtain deeper insight in why we observe the effect shown in Fig. 4, we can compare the total exit rate of an agent from a square for the symmetric model and the model with boredom probabilities. In particular, we compare these rates as a function of $s$, for a fixed fraction $\alpha$ of the total population in the square. In the symmetric case, we obtain $e^{-\alpha s}$ (see Eq.(2)), which is a decreasing function of $s$, exponentially approaching zero for large values of $s$. Hence, for a sufficiently large value of $s$, the rate of leaving a square becomes

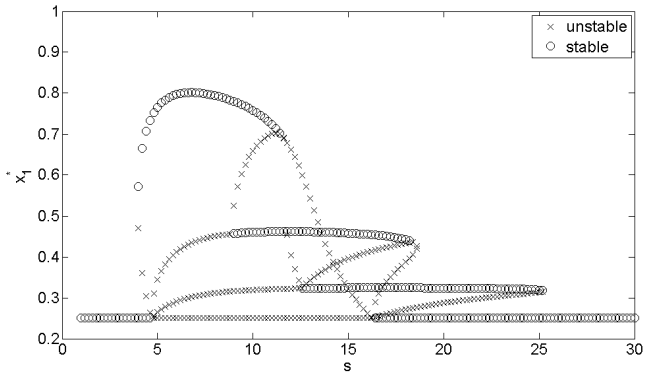



**Fig. 4.** Bifurcation diagram for the 4 square model with boredom probability $p = 0.05$.

very small: if there are many agents in the square ($\alpha$ close to 1), then we can expect nobody will move, as the exit rate from the crowded square will be much smaller than the exit rate of the other squares. On the other hand, for the model with boredom probability, the probability at which a *single* agent leaves a square is $p + (1 - p)e^{-\alpha s}$, which for large $s$ converges to $p$. This means that for large $s$, the effect of the chat probability to remain in a square is negligible compared to the effect of the boredom probability with which agents leave a square. This, in turn, implies that the ODE we obtain for large $s$ is essentially a set of linear ODE with a small non-linear perturbation term, hence they converge to the unique equilibrium of the linear system, which is the symmetric one in which the population is uniformly distributed over the squares.

In order to break this effect, we may want to find out whether the introduction of asymmetric attractiveness can counterbalance the disruptive effect of boredom probability on the emergent behaviour of the model. Fig. 6 (left) shows the bifurcation diagram for the model which $sq00$ is three times as attractive as the other squares, $a = 3$, and boredom probability with $p = 0.05$. We observe the same pattern as in the symmetric model: for large enough values of $s$, the system converges to the perturbed symmetric equilibrium (due to asymmetric attractiveness). An additional variation can be to consider a boredom probability which is inversely proportional to the attractiveness of squares, for instance equal to $p/a^2$, in order to take into account the effect that it is more unlikely that people just leave a square where interesting events are going on. In this case, in the presence of the compensating effect of boredom probability for large $s$, we can enlarge the range of $s$ for which we observe an emergent party (see Fig. 6 (right)). So, a combination of effects of the attractiveness on the choice of the next square and on the boredom probability can still be used to explain the emergent behaviour of 'El Botellon'.

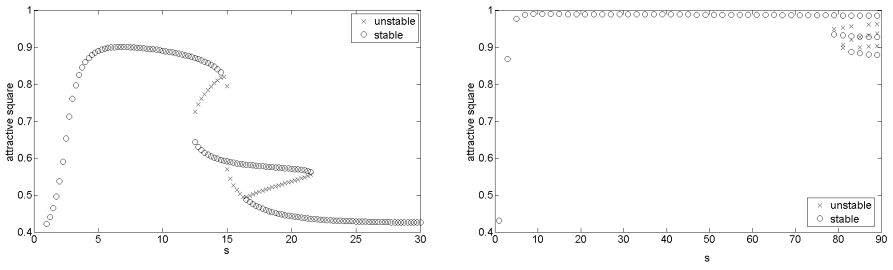



**Fig. 5.** Bifurcation diagram for the attractive square in 4 square model with boredom probability. for $p = 0.05$ and attractiveness equal to 3 (left), and attractiveness dependent $p$ (right).

## 7 Discussion and Further Work

The engineering of large scale collective dynamic systems is a relatively new domain of software engineering and requires effective and scalable formal analysis methods. We illustrated how the exploitation of the combination of process algebras, designed specifically for concurrent systems, and techniques typical of dynamic systems analysis, such as stability analysis, could provide valuable tools for such engineering approaches. In this paper we proposed a tool pipeline that, starting from a Bio-PEPA specification of non-linear asymmetric collective coordination, can produce bifurcation diagrams. These can be used to analyse the effect of potentially interfering coordination mechanisms on the stability properties of a system. We have illustrated the combined numeric and symbolic approach on a number of variants of a model of crowd dynamics that represents various kinds of *asymmetry*. The automatisation of the approach is feasible and part of future work. An issue that needs further investigation is the scalability of the method. Fluid ODEs are independent on the population size, but not on the number of states of the agents. Finding numerically all the zeros of a vector field can be challenging in large dimensions. One approach would be to parallelise the code and use more efficient zero finding numerical routines. An alternative and more promising direction is to exploit the formal nature of process algebras to reduce the agent's state space by using behavioural equivalences or abstract interpretation. Furthermore we plan to integrate the approach with the other already available analysis tools for Bio-PEPA, such as fluid flow analysis and stochastic simulation, via the Bio-PEPA tool suite.

**Acknowledgments.** This research has been partially funded by the EU-IP project ASCENS (nr. 257414), the EU-FET project QUANTICOL (nr. 600708) and the Italian MIUR-PRIN project CINA.

## References


1. GNU Octave, http://www.gnu.org/software/octave/
2. Matlab symbolic toolbox, http://www.mathworks.it/products/symbolic/
3. Maxima, a computer algebra system, http://maxima.sourceforge.net/
4. SBML: Systems Biology Markup Language, http://sbml.org
5. Bortolussi, L., Le Boudec, J.Y., Latella, D., Massink, M.: Revisiting the limit behaviour of "El Botellon". Tech. Rep. EPFL-REPORT-179935, École Polytechnique Fédérale de Lausanne - INFOSCIENCE (July 2012)
6. Ciocchetta, F., Duguid, A., Gilmore, S., Guerriero, M.L., Hillston, J.: The Bio-PEPA Tool Suite. In: Proc. of the 6th Int. Conf. on Quantitative Evaluation of SysTems (QEST 2009). pp. 309–310 (2009)
7. Ciocchetta, F., Hillston, J.: Bio-PEPA: A framework for the modelling and analysis of biological systems. TCS 410(33-34), 3065–3084 (2009)
8. Frei, R., Di Marzo Serugendo, G.: Advances in complexity engineering. International Journal of Bio-Inspired Computation 3, 199–212 (2011)
9. Gillepie, D.T.: Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry 81(25), 2340–2361 (1977)
10. Hillston, J.: A compositional approach to performance modelling (1996), distinguished Dissertation in Computer Science. Cambridge University Press.
11. Hillston, J.: Fluid flow approximation of PEPA models. In: Proceedings of QEST'05. pp. 33–43. IEEE Computer Society (2005)
12. Keating, S.M., Bornstein, B.J., Finney, A., M., H.: SBMLToolbox: an SBML toolbox for MATLAB users. Bioinformatics 22(10), 12751277 (2006), http://sbml.org/Software/SBMLToolbox
13. Khalil, H.K.: Nonlinear systems. MacMillan Pub. Co. (1992)
14. Kwiatkowska, M., Norman, G., Parker, D.: PRISM: Probabilistic model checking for performance and reliability analysis. ACM SIGMETRICS Performance Evaluation Review (2009)
15. Massink, M., Latella, D., Bracciali, A., Hillston, J.: Modelling non-linear crowd dynamics in Bio-PEPA. In: Fundamental Approaches to Software Engineering, LNCS, vol. 6603, pp. 96–110. Springer, Heidelberg (2011)
16. Massink, M., Brambilla, M., Latella, D., Dorigo, M., Birattari, M.: Analysing robot swarm decision-making with Bio-PEPA. In: Dorigo, M., Birattari, M., Blum, C., Christensen, A.L., Engelbrecht, A.P., Groß, R., Stützle, T. (eds.) Swarm Intelligence - 8th International Conference, ANTS 2012. LNCS, vol. 7461, pp. 25–36. Springer (September 2012)
17. Massink, M., Latella, D.: Fluid analysis of foraging ants. In: Sirjani, M. (ed.) Coordination Models and Languages (COORDINATION 2012). LNCS, vol. 7274, pp. 152–165. Springer-Verlag (June 2012)
18. MATLAB: v. 7.10.0 (R2010a). The MathWorks Inc., Natick, Massachusetts (2010)
19. Naphade, M., Banavar, G., Harrison, C., Paraszczak, J., Morris, R.: Smarter cities and their innovation challenges. Computer 44(6), 32–39 (june 2011)
20. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd Edition. Springer. (2006)
21. Rowe, J.E., Gomez, R.: El Botellón: Modeling the movement of crowds in a city. Complex Systems 14, 363–370 (2003)
22. Strogatz, S.H.: Non-linear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. Perseus Books Publishing (1994)