

Modelling Non-linear Crowd Dynamics in Bio-PEPA

Mieke Massink¹, Diego Latella¹, Andrea Bracciali^{1,3}, and Jane Hillston²

Istituto di Scienza e Tecnologie dell'Informazione 'A. Faedo', CNR, Italy¹
University of Edinburgh, Edinburgh, U.K.²
University of Stirling, U.K.³

Abstract. Emergent phenomena occur due to the pattern of non-linear and distributed local interactions between the elements of a system over time. Surprisingly, agent based crowd models, in which the movement of each individual follows a limited set of simple rules, often re-produce quite closely the emergent behaviour of crowds that can be observed in reality. An example of such phenomena is the spontaneous self-organisation of drinking parties in the squares of cities in Spain, also known as “El Botellón” [20]. We revisit this case study providing an elegant stochastic process algebraic model in Bio-PEPA amenable to several forms of analyses, among which simulation and fluid flow analysis. We show that a fluid flow approximation, i.e. a deterministic reading of the average behaviour of the system, can provide an alternative and efficient way to study the same emergent behaviour as that explored in [20] where simulation was used instead. Besides empirical evidence, also an analytical justification is provided for the good correspondence found between simulation results and the fluid flow approximation.

1 Introduction

In modern society the formation of crowds, intended as large concentrations of people, is a phenomenon that occurs frequently. Well known examples are crowds at large entertainment events in cities or other open-air facilities such as sport stadiums, but also crowds at large airports and train stations. Fortunately, such crowds usually occur and dissolve without serious problems. However, in some cases accidents happen with possibly major consequences such as loss of lives and a large number of injuries [22]. The recent drama at the open-air festival in Germany is sadly adding to the list of such events [16].

There is an ever stronger interest in being able to prevent such disasters and there exists an extensive literature on numerous approaches to the study of crowd formation, crowd management and emergency egress [21]. Simulation models play an important role in these approaches. In particular, agent based modelling has become popular in recent years because it may provide valuable information about the dynamics of systems that contain non-linear elements, chaos and random cause and effect. Several works in this area, e.g. work by Still [22], show that a crowd of people in which each individual follows a limited

number of simple rules produces quite closely the emergent behaviour that can be observed in real human crowds. Emergent phenomena are known to occur due to the pattern of non-linear and distributed local interactions between the large number of elements of a system over time. His work and that of others have led to the development of several professional tools, based on agent simulation, for the realistic analysis and prediction of crowd behaviour in emergency situations. Such analyses may help to detect architectural or organisational problems that may potentially cause loss of lives in the event of emergency situations.

However, for very large crowds, analysis via detailed simulation may become costly and time-consuming since each execution of the model produces only a single trajectory through the state space whereas many executions are needed to reach statistically relevant conclusions. Such high costs may be justified when a final verification of a system is required, but often become prohibitive in situations in which a quick analysis is required to compare the consequences of different design options or when a large number of slightly different scenarios need to be analysed.

In a completely different field of research, namely that of the analysis of biochemical reactions, it has been shown that under certain conditions, such as the presence of a sufficiently large population, a deterministic continuous interpretation of models composed of many similar small independent components provide a good approximation of the average transient behaviour of the overall model. In the context of stochastic process algebras this insight has led to the development of an alternative formal fluid flow semantics for PEPA first and later for Bio-PEPA, a variant of PEPA originally devised for modelling biochemical processes. Such semantics is based on the generation of sets of ordinary differential equations (ODE) [12, 7]. An application of PEPA with this alternative semantics in the context of emergency egress has been presented in [17, 3]. It has been shown that the approach is an efficient and scalable alternative to simulation when average behaviour is of interest. It is not suitable to analyse, for example, exceptional or oscillatory behaviour. Good correspondence with results from the literature on evacuation times and node profiles—the average number of people present in a particular part of the building over time during egress—has been achieved. In that work only linear differential equations have been considered.

In this paper we revisit the case of self-organisation of crowds in a city as described by Rowe and Gomez in [20]. The work was inspired by a typical social phenomenon observed in Spanish cities, on summer nights, called “El Botellón”, when crowds of youngsters wander between city squares in search of a party. Such self-organising parties sometimes lead to heavy drinking and noisy behaviour until late at night. It turned out to be hard to predict when and where a large party would take place. The aim of the work by Rowe and Gomez was to gain insight into the general principles due to which parties self-organise, abstracting from specific details of individual cases. We show that with Bio-PEPA a fluid flow approximation can provide an alternative way to study the same emergent behaviour as that explored in [20] where simulation was used instead. This case differs from that of emergency egress mentioned before because of the presence of

non-linear aspects where the behaviour of the individual agents depends directly on other, similar agents present in the same environment. In [20] the movement of a crowd in a city is studied under various assumptions about the likelihood that people remain in a square. Agents follow two basic rules. The first rule defines when agents remain in a square, which depends on the “chat-probability”, i.e. the likelihood to meet someone in the square to chat with. The second rule defines how agents move between squares.

Rowe and Gomez, by developing an analytical model, determined a threshold of the chat-probability below which people are freely moving through the city and above which large crowds start to form. They validated their theory by the simulation of a multi-agent model for a ring topology of 4 squares and up to 80 agents. Both the theory and the simulation results show that for a value of the chat probability $c = n/N$, where n is the number of squares and N the number of agents, a clear phase-transition can be observed between a steady-state situation in which agents are evenly distributed over the squares (when c is below the threshold) and a situation in which agents spontaneously gather in one or a few squares (when c is above the threshold).

In this paper, we approach the modelling of crowds by adopting the Bio-PEPA stochastic process algebra [8]. Bio-PEPA embeds a notion of spatial location, intended to model compartments, suitable to describe the city topology and locate agents within it. Moreover, the agent behaviour can be expressed as a function of the current state of the system, such as the number of people present in a square, an abstraction of the act of sensing the environment, common to standard agent models. Such a function may contain non-linear elements, which makes Bio-PEPA also particularly interesting for the analysis of some forms of emergent behaviour, as we will see in later sections. The fluid flow results obtained with the Bio-PEPA model correspond surprisingly well to the simulation results obtained with the the same model and to those published by Rowe and Gomez [20]. Informally speaking, for models where the rates can be expressed as functions of the average density of the population this phenomenon is well known (assuming that the populations are sufficiently large), see e.g. Kurtz [14] and in the context of mean field analysis Le Boudec et al. [2]. However, the rate functions in the crowds model addressed in this paper cannot be expressed this way. We provide an alternative analytical explanation for the observed correspondence which is partially based on recent work by Hayden and Bradley [11] on PEPA.

The outline of the paper is as follows. Section 2 recalls the crowd model used in the case study by Rowe and Gomez [20]. Section 3 briefly presents Bio-PEPA and its analysis environment. Section 4 describes the Bio-PEPA model of the collective behaviour of crowds in a city followed by a selection of the analyses results in Section 5. Section 6 provides insight in the close correspondence between the simulation results and fluid flow approximation for this case. Finally, in Section 7 conclusions are presented and future research is outlined. A preliminary version of our results has been discussed at the PASTA 2010 workshop [19], whereas further results and details can be found in [18].

2 Rowe and Gomez Model of Crowd Dynamics

In this section we briefly recall the model of movement of crowds between squares in a city as presented by Rowe and Gomez in [20]. Assume a city with n squares represented as a graph with vertices $\{1, 2, \dots, n\}$. People are simulated by “agents” that are following a simple set of rules. The number of agents in square i , with $i \in \{0, 1, \dots, n\}$, at time t , with t representing discrete time steps, is represented by $p_i(t)$. The state of the system at t is given by the number of agents present in each square modelled by the vector $\mathbf{p}(t) = (p_1(t), p_2(t), \dots, p_n(t))$. The total number of agents N at any time t is constant: $N = \sum_{i=1}^n p_i(t)$.

Agents are located in squares. The rules guiding agents’ behaviour are the following. The probability that an agent decides to remain in a square depends on how many other agents are present in the same square. If a square i contains $p_i > 0$ agents, the probability that an agent *leaves* the square is given by $(1 - c)^{p_i - 1}$. The parameter c (representing the *chat probability*, $0 \leq c \leq 1$) is the probability that an agent finds another one to talk to and thus remains in the square. Consequently, if the population of square i is p_i then the probability for an agent to find nobody to talk to is $(1 - c)^{p_i - 1}$. Note that when there is only one agent in the square, it decides to leave with probability 1, since there is nobody else to talk to. If an agent decides to move, it moves with equal probability to any neighbouring square reachable by a street. Considering an analytical model of the above discrete behaviour the *expected* number of agents that will leave square i at a given time step t is given by the function:

$$f_i(t) = p_i(t)(1 - c)^{p_i(t) - 1}$$

This models the part of the population in square i that does not find anyone to talk to in that square¹. The probability that an agent, which decided to leave square j , moves to the adjacent square i is given by the matrix A_{ij} :

$$A_{ij} = \text{con}_{ij} / d_j$$

where d_j is the degree of vertex j , i.e. the number of streets departing from square j , and con_{ij} denotes that square i is connected to square j :

$$\text{con}_{ij} = \begin{cases} 1 & \text{if } i \text{ is connected to } j \\ 0 & \text{otherwise} \end{cases}$$

Clearly, $\text{con}_{ij} = \text{con}_{ji}$ and we assume that adjacent squares are connected by at most one street. The expected distribution of agents over squares at time $t + 1$ can now be defined as:

$$\mathbf{p}(t + 1) = \mathbf{p}(t) - \mathbf{f}(t) + \mathbf{A}\mathbf{f}(t)$$

¹ Note that in this analytical model the number of agents $p_i(t)$ in square i is now approximated by a real number: the *expected* number of agents in square i at time t .

Clearly, from this formula it follows that a steady-state behaviour is reached when $f(t) = Af(t)$. In other words, when the number of people entering a square is equal to the number leaving the square. Rowe and Gomez show that there are two possibilities for such a stable state. In one case the agents freely move between squares and their distribution is proportional to the number of streets connected to each square. In the second case agents gather in large groups in a small number of squares corresponding to emergent self-organisation of parties. Which of the two situations will occur depends critically on the value of the chat probability c . When all squares have the same number of neighbouring squares a phase shift occurs at about $c = n/N$ where n is the number of squares and N the number of agents. For $c < n/N$ people freely move between squares whereas for $c > n/N$ agents self-organise into large groups. Simulation of the model confirms in an empirical way that this estimate for c is quite accurate when the population is large enough where large means about 60 agents or more in a 4-square topology.

For topologies where each square has the same number of streets the critical value of c can be estimated in an analytical way. For less regular topologies and when different squares have different chat probabilities and not all directions leaving from a square are equally likely to be taken by people it is very difficult to identify such critical values in an analytical way. Usually, in such cases simulation is used to analyse the models. However, when a large number of agents is involved, simulation may be extremely time consuming.

3 Bio-PEPA and Fluid Flow Analysis

In this section we briefly describe Bio-PEPA [7, 8, 6], a language that has recently 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, 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:

$$S ::= (\alpha, \kappa) \text{ op } S \mid S+S \mid C \quad \text{with op} = \downarrow \mid \uparrow \mid \oplus \mid \ominus \mid \odot \quad P ::= P \underset{c}{\boxtimes} P \mid S(x)$$

where S is a *species component* and P is a *model component*. In the prefix term $(\alpha, \kappa) \text{ op } S$, κ is the *stoichiometry coefficient* of species S in action α . This arises from the original formulation of the process algebra for modelling biochemical reactions, where the stoichiometric coefficient captures how many molecules of a species are required for a reaction. However it may be interpreted more generally as the multiples of an entity involved in an occurring action. The default value of κ is 1 in which case we simply write α instead of (α, κ) . 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 *reactant* which will be consumed in the action, \uparrow a *product* which is produced as a result of the action, \oplus an *activator*, \ominus an *inhibitor* and \odot a generic *modifier*, all of which

play a role in an action without being produced or consumed and have a defined meaning in the biochemical context. 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 $x \in \mathbb{R}$ represents the initial amount of the 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/kinetics rates, parameters, etc.). The prefix term $(\alpha, \kappa) \text{ op } S@l$ is used to specify that the action is performed by S in location l . The notation $\alpha[I \rightarrow J] \odot S$ is a shorthand for the pair of reactions $(\alpha, 1) \downarrow S@I$ and $(\alpha, 1) \uparrow S@J$ that synchronise on action α ². 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 [8] which is based on Continuous Time Markov Chains (CTMCs).

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 [8, 5]. These tools include mappings from Bio-PEPA to differential equations (supporting a fluid flow approximation), stochastic simulation models [10], CTMCs with levels [7] and PRISM models [15].

A Bio-PEPA model describes a number of sequential components each of which represents a number of entities in a distinct state. The result of an action is to increase the number of some entities and decrease the number of others. Thus the total state of the system at any time can be represented as a vector with entries capturing the counts of each species component (i.e. an aggregated CTMC). This gives rise to a discrete state system which undergoes discrete events. The idea of fluid flow analysis is to approximate these discrete jumps by continuous flows between the states of the system.

4 Modelling Crowd Movement with Bio-PEPA

Let us consider the same small ring topology with 4 squares, A , B , C and D , as in Rowe and Gomez, allowing bi-directional movement between squares. The excerpt from the Bio-PEPA specification below defines this topology. The default compartment *top* contains all other compartments. The next line defines square A . Definitions for the other squares are similar and have been omitted. In this context *size* is used to denote a capacity in terms of number of agents.

$$\begin{aligned} & \text{location } top : \text{size} = 1000, \text{type} = \text{compartment}; \\ & \text{location } sqA \text{ in } top : \text{size} = \text{normal_square}, \text{type} = \text{compartment}; \end{aligned}$$

The size of the squares, defined by parameter *normal_square* = 100, is defined in such a way that all agents, 60 in this case, would fit in any single square and

² The concrete syntax for writing this in the Bio-PEPA tool set differs somewhat.

does not impose any further constraints. The Bio-PEPA specification, which we will henceforth refer to as the ‘crowd model’, has two further parameters. The parameter c defines the chat-probability and the parameter d the degree or number of streets connected to a square. In the considered topology $d = 2$ for each square. The actions modelling agents moving from square X to square Y will be denoted by $fXtY$. The associated functional rate (indicated by the keyword “kineticLawOf”) is defined in analogy to [20] (see Section 2). Since the only information on the probability distribution available is the expected number of agents leaving a square per time unit, this same information can also be modelled as the rate parameter of an exponential distribution. If one also considers the uniform distribution of people over the outgoing streets of the square then this rate needs to be divided by its degree d when agents leaving through a particular street are considered.

So the general rate with which agents leave square X via a particular street is:

$$(P@sqX * (1 - c)^{(P@sqX-1)})/d$$

This leads to the following functional rates for the crowd model, one for each direction of movement. Only the one for $fAtB$ is shown, the others being similar:

$$kineticLawOf fAtB : (P@sqA * (1 - c)^{(P@sqA-1)})/d;$$

The sequential component P below specifies the possible movements of a typical agent between squares. For example, $fAtB[sqA \rightarrow sqB] \odot P$ means that an agent present in square A moves to square B according to the functional rate defined for the action $fAtB$.

$$\begin{aligned} P = & fAtB[sqA \rightarrow sqB] \odot P + fBtA[sqB \rightarrow sqA] \odot P + \\ & fAtC[sqA \rightarrow sqC] \odot P + fCtA[sqC \rightarrow sqA] \odot P + \\ & fBtD[sqB \rightarrow sqD] \odot P + fDtB[sqD \rightarrow sqB] \odot P + \\ & fCtD[sqC \rightarrow sqD] \odot P + fDtC[sqD \rightarrow sqC] \odot P; \end{aligned}$$

Finally, the model component defines the initial conditions of an experiment, i.e. in which squares the agents are located initially, and the relative synchronisation pattern. Initially, there are 60 agents in square A . This is expressed by $P@sqA[60]$ in the composition shown below. All other squares are initially empty (i.e. $P@sqX[0]$ for $X \in \{B, C, D\}$). The fact that moving agents need to synchronise follows from the definition of the shorthand operator \rightarrow .

$$(P@sqA[60] \bowtie_* P@sqB[0]) \bowtie_* (P@sqC[0] \bowtie_* P@sqD[0])$$

The total number of agents $P@sqA + P@sqB + P@sqC + P@sqD$ is invariant and amounts to 60 in this specific case.

5 Selected Results for a Model with Four Squares

This section presents a selection³ of the analysis results for the model with four squares. The figures report both analysis via Gillespie stochastic simulation

³ Further results can be found in [18].

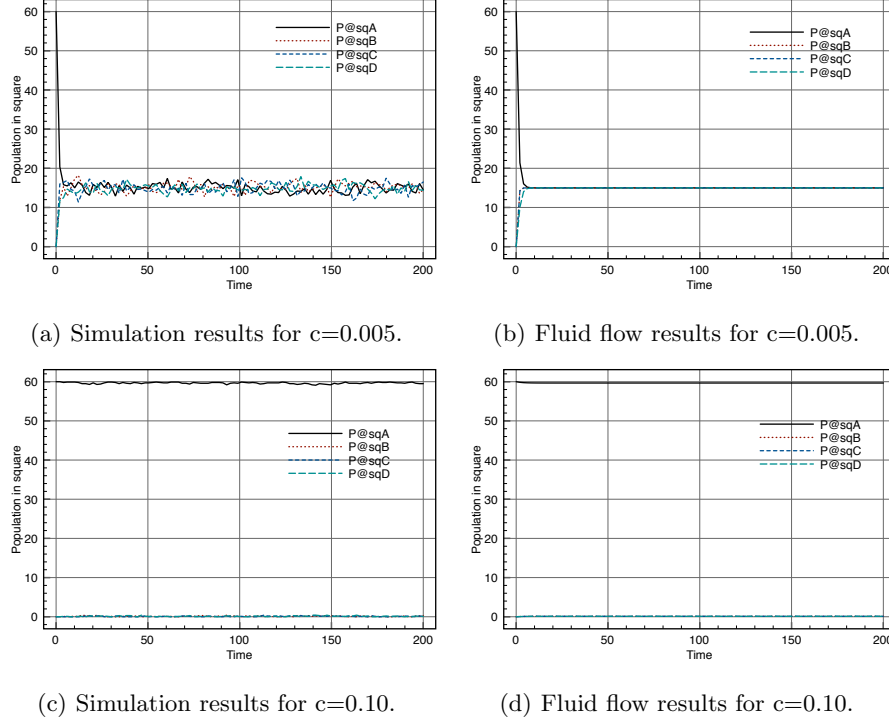
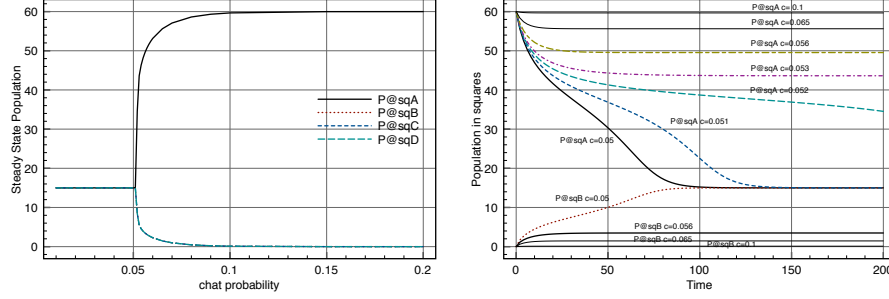


Fig. 1. Results for four squares with 60 agents in A initially

(G) [10], averaged over 10 independent runs, and fluid flow analysis based on the adaptive numeric solution of sets of ODEs based on the adaptive step-size 5th order Dormand-Prince ODE solver [9].⁴

Figure 1(a) and Fig. 1(b) show stochastic simulation and fluid flow results for a model with 60 agents initially in square A and for $c = 0.005$, which is below the analytically estimated threshold of $c = n/N = 4/60 = 0.0666$. The results show that a dynamic equilibrium is reached, i.e. all agents distribute evenly over the four squares. This confirms the discrete event simulation results reported by Rowe and Gomez. Figure 1(c) shows the results of stochastic simulation for the same model, but for $c = 0.10$, a value above the threshold. Corresponding fluid flow results for the same value of c are shown in Fig. 1(d). The figures show that the population settles rather quickly in a steady state in which almost all agents remain in square A. This is the second type of steady state observed also by Rowe and Gomez. Interestingly, the fluid flow analyses of the same model, for both values of c (Fig. 1(b) and Fig. 1(d)) show very good correspondence to the respective simulation results (Fig 1(a) and Fig. 1(c), resp.).

⁴ All analyses have been performed with the Bio-PEPA Eclipse Plug-in tool [5] on a Macintosh PowerPC G5.



(a) Fluid flow results about “Steady state” (b) Fluid flow results for square A (and population levels in each square. partially B) for varying chat probabilities.

Fig. 2. Results at $t=200$ for varying chat probabilities.

Since these results show that both types of steady state emerge in this stochastic version of the crowd model and for both types of analysis, the question naturally arises whether fluid flow could be used as an efficient technique to investigate the behaviour of the model for various values of the chat-probability c , in particular those close to the critical threshold. Fig. 2(a) shows the expected number of agents in the squares at $t = 200$, starting with 60 agents in square A initially, for different chat probabilities ranging from 0.01 to 0.2 with steps of 0.01 (except between 0.05 and 0.065 where the steps are 0.001) in a *fluid flow analysis*. The figure shows clearly that for a chat probability below 0.05 in the steady state the population is evenly distributed over the four squares. For $c > 0.05$ the situation changes sharply. For these higher values of c the agent population tends to concentrate in square A, the square from which they all started. All other squares remain essentially empty. The results in Fig. 2(a) illustrate a clear case of spontaneous *self-organisation* or *emergent* behaviour. The results in Fig. 2(a) closely correspond to those obtained by Rowe and Gomez [20] by discrete event simulation. The ease and computational efficiency with which these results can be produced by means of fluid flow analysis opens up a promising perspective on how process algebraic fluid flow analysis could be used as an alternative, efficient, scalable and formal approach to investigate emergent behaviour in the vicinity of critical parameter values for this class of models.

An impression of how the distribution of agents over the four squares evolves for values of c that are close to the threshold of $c = 0.05$ is shown in Fig. 2(b). For $c = 0.05$ and $c = 0.051$ the agents still distribute uniformly over the four squares, though this takes a bit more time than for lower values of c . For $c = 0.052$ this situation is changing, and for $c = 0.053$ and higher clearly a different steady state is reached in which most agents group in the single square A. Note that for $c = 0.052$ a stable state has not yet been reached at time $t = 200$. However, this does not influence the overall picture. The theory on the stability of fixed-points predicts that the steady state behaviour for values of $c > n/N = 4/60 = 0.066666$ is unstable (see [20]). Instability in this case means extreme sensitivity

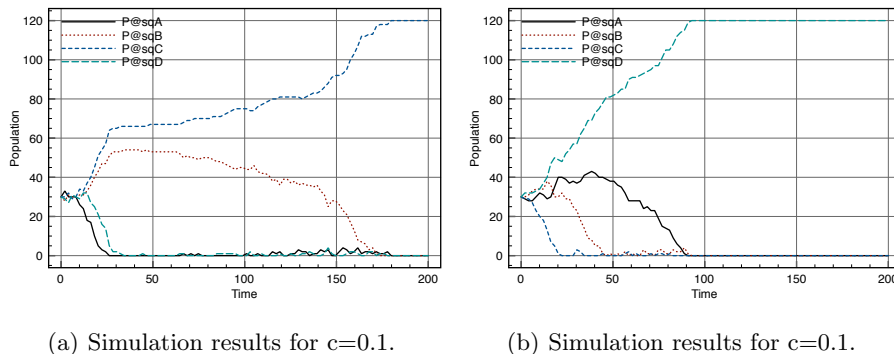


Fig. 3. Two single simulation runs for the same model with initially 30 agents in each square

to the initial values of the number of agents in each square. This phenomenon is illustrated by the results in Fig. 3 where single simulation runs are shown for the same model and the same initial conditions, i.e. 30 agents per square and $c = 0.1$. In every run the agents may all gather in one of the squares without leaving again. However, no prediction is possible on which square this will be except that they are all equally likely in this particular topology. Scalability of the fluid flow analysis is well illustrated by comparing the time of a *single* simulation run for 600,000 agents over 500 time units in a 3 by 3 grid topology, taking approx. 27 min., and a fluid flow analysis of that model taking 20 ms.⁵

6 Analytical Assessment of the Fluid Flow Approximation

The results in the previous sections show a very good correspondence between the results obtained via fluid flow approximation and those obtained, on the one hand, with Gillespie’s stochastic simulation algorithm applied on exactly the same Bio-PEPA specification and, on the other hand, with discrete event simulation results found by Rowe and Gomez [20]. In this section we provide a justification for this correspondence from an analytical perspective.

There exist several theories that address the relation between the interpretation of the model as a large set of individual, independently behaving and interacting agents, as is the case for simulation, and a continuous deterministic interpretation of the same model as occurs with a fluid flow approximation. Perhaps the most well-known is the theory by Kurtz [14]. Informally speaking, Kurtz shows an exact relation (in the limit when the population goes to infinity) between the two above mentioned interpretations when the rate-functions can be expressed in terms of the average *density* of the population. A similar requirement needs to be satisfied in the context of the theory of mean field analysis,

⁵ These timing results have been obtained with the Bio-PEPA plugin tool for Eclipse Helios on a MacPro4,1.

for example in recent work by Le Boudec et al. [2]. Unfortunately, in the crowd model the rate-functions cannot be expressed in terms of the density of the population because the exponent of the factor $(1 - c)^{(p_X - 1)}$ requires the absolute number p_X of agents present in square X .

A third approach, by Hayden and Bradley [11], has recently been applied to assess the quality of the fluid approximation for (grouped) PEPA models. In that approach the Chapman-Kolmogorov forward equations (C-K) are derived from a typical central state of the aggregated CTMC⁶ associated with a PEPA model. These equations are then used in the moment generating function from which, by partial differentiation, ordinary differential equations are obtained for the expected value over time of each sequential component in the PEPA model. In this section we adapt the approach to the Bio-PEPA crowds model which is characterised by non-linear rate functions. Let p_A , p_B , p_C and p_D

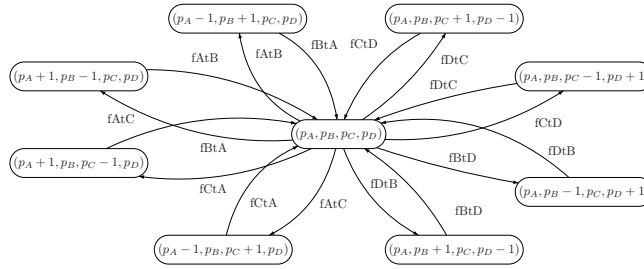


Fig. 4. A central state of the crowds model.

denote the number of agents in square A , B , C and D , respectively. A central state of the aggregated CTMC of the crowds model is shown in Fig. 4. Let $prob_{(p_A, p_B, p_C, p_D)}(t)$ denote the transient probability of being in the aggregated CTMC state with p_X agents in square X , for X in $\{A, B, C, D\}$ at time t . From the crowd model in Sect. 4 we obtain the following C-K, governing the evolution of the state probabilities over time of the underlying aggregated CTMC:

$$\begin{aligned} \frac{d prob_{(p_A, p_B, p_C, p_D)}(t)}{dt} = & ((p_A + 1) \cdot (1 - c)^{p_A})/2 \cdot prob_{(p_A + 1, p_B - 1, p_C, p_D)}(t) \\ & + ((p_B + 1) \cdot (1 - c)^{p_B})/2 \cdot prob_{(p_A - 1, p_B + 1, p_C, p_D)}(t) \\ & + ((p_C + 1) \cdot (1 - c)^{p_C})/2 \cdot prob_{(p_A, p_B, p_C + 1, p_D - 1)}(t) \\ & + ((p_D + 1) \cdot (1 - c)^{p_D})/2 \cdot prob_{(p_A, p_B, p_C - 1, p_D + 1)}(t) \\ & + ((p_A + 1) \cdot (1 - c)^{p_A})/2 \cdot prob_{(p_A + 1, p_B, p_C - 1, p_D)}(t) \\ & + ((p_C + 1) \cdot (1 - c)^{p_C})/2 \cdot prob_{(p_A - 1, p_B, p_C + 1, p_D)}(t) \\ & + ((p_B + 1) \cdot (1 - c)^{p_B})/2 \cdot prob_{(p_A, p_B + 1, p_C, p_D - 1)}(t) \\ & + ((p_D + 1) \cdot (1 - c)^{p_D})/2 \cdot prob_{(p_A, p_B - 1, p_C, p_D + 1)}(t) \\ & - ((p_A) \cdot (1 - c)^{p_A - 1}) \cdot prob_{(p_A, p_B, p_C, p_D)}(t) \\ & - ((p_B) \cdot (1 - c)^{p_B - 1}) \cdot prob_{(p_A, p_B, p_C, p_D)}(t) \\ & - ((p_C) \cdot (1 - c)^{p_C - 1}) \cdot prob_{(p_A, p_B, p_C, p_D)}(t) \\ & - ((p_D) \cdot (1 - c)^{p_D - 1}) \cdot prob_{(p_A, p_B, p_C, p_D)}(t) \end{aligned}$$

⁶ For a formal definition see for example [11].

Each of the eight first summands appears only when the state (p_A, p_B, p_C, p_D) has the corresponding incoming transitions in the aggregated state space.

Let us now consider, following the approach outlined in [11], how an ODE for the function $P_A(t)$ can be obtained. First note that the expected value of $P_A(t)$ is given by:

$$\mathbb{E}[P_A(t)] = \sum_{(p_A, p_B, p_C, p_D)} p_A \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t)$$

So this leads to the following derivation:

$$\begin{aligned} \frac{d \mathbb{E}[P_A(t)]}{dt} &= \{\text{By def. of expected value}\} \\ \frac{d \sum_{(p_A, p_B, p_C, p_D)} p_A \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t)}{dt} &= \{\text{By distribution of differentiation}\} \\ \sum_{(p_A, p_B, p_C, p_D)} p_A \cdot \frac{d \text{prob}_{(p_A, p_B, p_C, p_D)}(t)}{dt} &= \{\text{By definition of C-K equations}\} \\ \sum_{(p_A, p_B, p_C, p_D)} [& ((p_A - 1) \cdot p_A \cdot (1 - c)^{(p_A-1)})/2 \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t) \\ & + ((p_A + 1) \cdot p_B \cdot (1 - c)^{(p_B)})/2 \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t) \\ & + (p_A \cdot p_C \cdot (1 - c)^{(p_C-1)})/2 \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t) \\ & + (p_A \cdot p_D \cdot (1 - c)^{(p_D-1)})/2 \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t) \\ & + ((p_A - 1) \cdot p_A \cdot (1 - c)^{(p_A-1)})/2 \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t) \\ & + ((p_A + 1) \cdot p_C \cdot (1 - c)^{(p_C-1)})/2 \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t) \\ & + (p_A \cdot p_B \cdot (1 - c)^{(p_B-1)})/2 \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t) \\ & + (p_A \cdot p_D \cdot (1 - c)^{(p_D-1)})/2 \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t) \\ & - (p_A \cdot p_A \cdot (1 - c)^{(p_A-1)}) \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t) \\ & - (p_A \cdot p_B \cdot (1 - c)^{(p_B-1)}) \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t) \\ & - (p_A \cdot p_C \cdot (1 - c)^{(p_C-1)}) \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t) \\ & - (p_A \cdot p_D \cdot (1 - c)^{(p_D-1)}) \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t)] \end{aligned}$$

If $(P_A(t), P_B(t), P_C(t), P_D(t))$ is the state of the aggregated CTMC at time t , then by cancelling terms in the above equation one obtains:

$$\begin{aligned} \sum_{(p_A, p_B, p_C, p_D)} [& -p_A \cdot (1 - c)^{(p_A-1)} \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t) \\ & + (p_B \cdot (1 - c)^{(p_B-1)}) \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t)/2 \\ & + (p_C \cdot (1 - c)^{(p_C-1)}) \cdot \text{prob}_{(p_A, p_B, p_C, p_D)}(t)/2 \end{aligned}$$

This yields in terms of expectations the following ODE for $\mathbb{E}[P_A(t)]$:

$$\begin{aligned} \frac{d \mathbb{E}[P_A(t)]}{dt} &= -\mathbb{E}[P_A(t) \cdot (1 - c)^{(P_A(t)-1)}] \\ &+ (\mathbb{E}[P_B(t) \cdot (1 - c)^{(P_B(t)-1)}])/2 \\ &+ (\mathbb{E}[P_C(t) \cdot (1 - c)^{(P_C(t)-1)}])/2 \end{aligned}$$

If at this point, as shown in [11], the functions $(1 - c)^{(X-1)}$ were just constant rates, then expectation would just distribute over multiplication and one would obtain an equation in terms of expectations of populations. However, in our case the rate is a more complicated function and, in general, expectation does

not distribute over an arbitrary function, i.e. $\mathbb{E}[\phi(X)] \neq \phi(\mathbb{E}[X])$. This means that exact equality cannot be obtained this way. As an alternative we consider whether $\mathbb{E}[P_A(t)](1-c)^{(\mathbb{E}[P_A(t)]-1)}$ could be expected to approximate $\mathbb{E}[P_A(t) \cdot (1-c)^{(P_A(t)-1)}]$. To address this question we recall Jensen's work [13] from which it is known that for strictly convex functions ϕ the following inequality holds:

$$\mathbb{E}[\phi(X)] \geq \phi(\mathbb{E}[X])$$

with *equality* when ϕ is linear or when it is a constant. The reverse inequality holds in case ϕ is strictly concave. So this requires a closer investigation of the particular function at hand. In Fig. 5 graphs are shown of the function $\phi(x) = x(1-c)^x$ for values of $x \in [0 \dots 60]$ and for various values of chat probability c . It can be observed that for very small values of c the function is almost linear for the number of agents considered. For larger values of c the function is mostly concave and tends to an almost constant value. This implies that, informally speaking, any hypothetical probability distribution of the values of x would be mapped on an increasingly "shrinking" version of the distribution of $\phi(x)$. This means that $\mathbb{E}[P_A(t)](1-c)^{(\mathbb{E}[P_A(t)]-1)}$ approximates $\mathbb{E}[P_A(t) \cdot (1-c)^{(P_A(t)-1)}]$ indeed rather well.

Distributing expectation over the function the following ODE for the expected value of $P_A(t)$ is obtained:

$$\begin{aligned} \frac{d\mathbb{E}[P_A(t)]}{dt} \approx & -\mathbb{E}[P_A(t)](1-c)^{(\mathbb{E}[P_A(t)]-1)} \\ & +(\mathbb{E}[P_B(t)](1-c)^{(\mathbb{E}[P_B(t)]-1)})/2 \\ & +(\mathbb{E}[P_C(t)](1-c)^{(\mathbb{E}[P_C(t)]-1)})/2 \end{aligned}$$

This ODE is identical to the one generated by the Bio-PEPA toolset which follows the approach described by Ciocchetta and Hillston in [7]. The derivation is provided in [18]. In a similar way the ODEs for the other stochastic variables can be obtained. This explains why the results obtained in this paper for fluid flow and for stochastic simulation also in this non-linear setting correspond so closely.

7 Conclusions and further work

The modelling and analysis of crowd dynamics appears to be an active and open research topic. We have explored the application of the stochastic process algebra Bio-PEPA to model a simple but interesting non-linear case study concerning the emergent self-organisation of parties in the squares of a city. Bio-PEPA is based on a modular, high-level language providing notions of locality and context dependency. These features make Bio-PEPA also a promising candidate for the modelling of a class of possibly non-linear systems that goes beyond the bio-molecular applications it was originally designed for [3, 8, 1]. In this case study fluid flow approximation provides a computationally efficient analysis of the number of people, on average, that are present in the various squares when

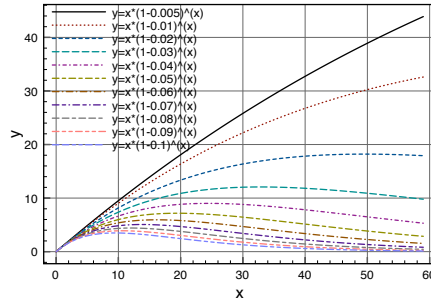


Fig. 5. Function $x(1 - c)^x$ for up to 60 agents and different chat probabilities.

time evolves. The results are shown to correspond well to those found in the literature where they were obtained by means of more elaborate and time-consuming discrete event simulation. Also, an analytical approach to explain this good correspondence for a model the rates of which cannot be expressed as functions of the average population density, has been provided. Although the simple topology addressed in this paper, chosen for reasons of validation of the approach, can be analysed analytically, more complex topologies and more realistic models, addressing issues such as the influence of the presence of friends on people's behaviour and the attractiveness of squares, may easily result too complex to be studied analytically. Fluid flow approximation with Bio-PEPA has shown to be a suitable choice in this case. Preliminary results about more complex models can be found in [18].

Future work is developing along a few main directions. We are interested in developing further linguistic abstractions to describe more precisely the dynamics of systems with a large number of mobile agents displaced in a, possibly open, physical environment. We are furthermore interested in conducting more fundamental research on the fluid flow approach and its relationship to emergent non-linear behaviour, in particular its relation to mean field analysis [2].

Acknowledgments. The authors would like to thank Stephen Gilmore, Maria Luisa Guerriero, Allan Clark and Adam Duguid (University of Edinburgh) for their support with the Bio-PEPA plug-in, Richard Hayden and Jeremy Bradley (Imperial College London) for discussions on approximations and Michael Harrison and Nigel Thomas for the case study. This research has been partially funded by the CNR project RSTL-XXL and by the Italian PRIN MIUR project PaCO. Jane Hillston is supported by the EPSRC ARF EP/c543696/01.

References

1. O. E. Akman, F. Ciocchetta, A. Degasperi & M. L. Guerriero (2009): *Modelling biological clocks with Bio-PEPA: Stochasticity and robustness for the Neurospora Crassa circadian network*. In: *Proc. of CSBM, LNCS 5688*, Springer, pp. 52–67.

2. M. Benaïm & J.-Y. Le Boudec (2008): *A Class of Mean Field Interaction Models for Computer and Communication Systems*. *Performance Evaluation* 65(11-12), pp. 823–838.
3. A. Bracciali, J. Hillston, D. Latella & M. Massink (2010): *Reconciling Population and Agent Models for Crowd Dynamics*. In: *Proc. of 3rd LAM'10*, to appear.
4. M. Calder, S. Gilmore & J. Hillston (2005): *Automatically deriving ODEs from process algebra models of signalling pathways*. In: G. Plotkin, (Ed.), *Proc. of CMSB 2005*, pp. 204–215.
5. F. Ciocchetta, A. Duguid, S. Gilmore, M. L. Guerriero & Hillston J. (2009): *The Bio-PEPA Tool Suite*. In: *Proc. of the 6th Int. Conf. on Quantitative Evaluation of SysTems (QEST 2009)*, pp. 309–310.
6. F. Ciocchetta & M. L. Guerriero (2009): *Modelling Biological Compartments in Bio-PEPA*. *ENTCS* 227, pp. 77–95.
7. F. Ciocchetta & J. Hillston (2008): *Bio-PEPA: An Extension of the Process Algebra PEPA for Biochemical Networks*. *ENTCS* 194(3), pp. 103–117.
8. F. Ciocchetta & J. Hillston (2009): *Bio-PEPA: A framework for the modelling and analysis of biological systems*. *TCS* 410(33-34), pp. 3065–3084.
9. J. R. Dormand & P. J. Prince (1980): *A family of embedded Runge-Kutta formulae*. *Journal of Computational and Applied Mathematics* 6(1), pp. 19–26.
10. D. T. Gillespie (1977): *Exact Stochastic Simulation of Coupled Chemical Reactions*. *The Journal of Physical Chemistry* 81(25), pp. 2340–2361.
11. R. A. Hayden & J. T. Bradley (2010): *A fluid analysis framework for a Markovian process algebra*. *TCS* 411(22-24), pp. 2260–2297.
12. J. Hillston (2005): *Fluid flow approximation of PEPA models*. In: *Proc. of QEST'05*, IEEE Computer Society, pp. 33–43.
13. J. L. W. V. Jensen (1906): *Sur les fonctions convexes et les inégalités entre les valeurs moyennes*. *Acta Mathematica* 30(1), pp. 175–193.
14. T. G. Kurtz (1970): *Solutions of Ordinary Differential Equations as Limits of Pure Markov Processes*. *Journal of Applied Probability* 7(1), pp. 49–58.
15. M. Kwiatkowska, G. Norman & D. Parker (2009): *PRISM: Probabilistic Model Checking for Performance and Reliability Analysis*. *ACM SIGMETRICS Performance Evaluation Review*.
16. Love-parade. *Stampede at German Love Parade festival kills 19*. <http://www.bbc.co.uk/news/world-europe-10751899>, accessed on 10/08/2010.
17. M. Massink, D. Latella, A. Bracciali & M. Harrison (2010): *A Scalable Fluid Flow Process Algebraic Approach to Emergency Egress Analysis*. In: *Proc. of SEFM 2010*, IEEE, pp. 169–180.
18. M. Massink, D. Latella, A. Bracciali & J. Hillston (2010): *A Combined Process Algebraic, Agent and Fluid Flow Approach to Emergent Crowd Behaviour*. Technical Report 2010-TR-025, CNR-ISTI.
19. M. Massink, D. Latella, A. Bracciali & J. Hillston (2010). *Modelling Crowd Dynamics in Bio-PEPA – Extended Abstract*. Participant Proc. of the 9th Workshop on Process Algebra and Stochastically Timed Activities (PASTA 2010).
20. J. E. Rowe & R. Gomez (2003): *El Botellón: Modeling the Movement of Crowds in a City*. *Complex Systems* 14, pp. 363–370.
21. G. Santos & B. E. Aguirre (2005): *A critical review of emergency evacuation simulation models*. In: *Proc. of the NIST Workshop on Building Occupant Movement during Fire Emergencies, June 10-11, 2004*, NIST/BFRL Publications Online, Gaithersburg, MD, USA, pp. 27–52.
22. G. K. Still (2000). *Crowd Dynamics*. Ph.D. Thesis, University of Warwick.