ODE approximations to some Markov chain models

Perla Sousi

January 13, 2009

Contents

Abstract

Markov chains are widely used to model various dynamical systems that evolve randomly in time. Differential equations on the other hand model deterministic processes and can often be handled more easily. It is of interest to us to try to approximate a Markov chain by a differential equation, since in this way we can get more insight on the behaviour of the stochastic system.

In this essay we present the general theory and then use it to approximate three different Markov models from three different application areas.

The first one is a viral model (a biological application) where we apply the homogenization technique to prove a limit theorem.

The second one is the 3-Satisfiability problem of Computer Science. We analyze an algorithm which gives us a lower bound for the 3-satisfiability threshold. The method: we investigate a Markov chain, approximate it by a differential equation using the homogenization technique again and finally deduce facts about the Markov chain by looking at the ODE. The ODEs we obtain are the limit as $\theta \to 0$ of the ODEs given in [4].

Finally we consider an application in the area of communication networks, which is a model of the popular filesharing BitTorrent network. The basic feature of it is the segmentation of a file into chunks that peers either download or swap in order to get the whole file. We model the number of users that own a particular set of chunks as a Markov chain. We prove that it is positive recurrent and then we approximate it by a set of differential equations. Moreover, we prove, under assumptions, that the stationary process converges to the equilibrium point of the differential equation.

Supervisor: Prof J.R. Norris

In Chapter 1, I give an overview of the methods that are used in the next chapters. Citations are given where appropriate.

In Chapter 2, I present an approximation of a viral model which has been analyzed before and is cited accordingly, but the method that I am using is novel. The idea to apply the homogenization technique was suggested to me by my supervisor, but the details were implemented by me.

In Chapter 3, I analyze an algorithm for the 3-Satisfiability problem using differential equations, in a different way to the ones used before. Again the idea to apply the homogenization technique was my supervisor's and I implemented it.

Chapter 4 contains an analysis of BitTorrent, which is joint work with Professor Takis Konstantopoulos (Heriot-Watt University) and Professor George Kesidis (The Pennsylvania State University) and will appear in Proceedings of Net-Coop 2008, Springer Lecture Notes in Computer Science 5425. The detailed formulation of the Markov model as well as all the calculations for the vector field were done by me. The proof of the convergence to the differential equation was done originally by me by applying the technique from Chapter 1, but the proof that is presented here is more general and does not require Markovian assumptions and is due to Takis Konstantopoulos. In this essay I slightly changed the model and the theorems proved in Section 4.4.2 are my work and citations are given where appropriate. The simulations were generated by Youngmi Jin, who is a student of George Kesidis, and I thank them for giving them to me.

Chapter 1

Fluid limits and homogenization

1.1 General theory

In this chapter we will recall the set-up used in [1] and [2]. In the next chapters we are going to use this machinery to approximate some Markov chain models by differential equations.

Let $X = (X_t)_{t\geq 0}$ be a continuous-time Markov chain in a countable state space S with generator Q having elements $q(\xi, \xi')$ giving the jump rate from ξ to ξ' for $\xi \neq \xi'$. We assume that $q(\xi)$ (the total jump rate) is finite for all $\xi \in S$ and that X does not explode. Let $\boldsymbol{x}$ be a mapping $\boldsymbol{x} : S \to \mathbb{R}^d$. In practice, $\boldsymbol{x}$ will give the coordinates of the Markov chain that we want to approximate. Then we can write

$$\boldsymbol{x}(X_t) = \boldsymbol{x}(X_0) + M_t + \int_0^t \beta(X_s)\mathrm{d}s, \tag{1.1}$$

where β is the drift vector field

$$\beta(\xi) = \sum_{\xi' \neq \xi} q(\xi, \xi')(\boldsymbol{x}(\xi') - \boldsymbol{x}(\xi))$$

and M is a martingale given by

$$M_t = \int_0^t \int_S (\boldsymbol{x}(y) - \boldsymbol{x}(X_{s-}))(\mu - \nu)(\mathrm{d}s, \mathrm{d}y),$$

with μ and ν being the following measures:

$$\mu = \delta_{(J_n, Y_n)}, \quad \nu(\mathrm{d}s, \mathrm{d}y) = q(X_{s-}, \mathrm{d}y)\mathrm{d}s,$$

with $(J_n)_n$ being the jump times and $(Y_n)_n$ the embedded discrete time jump chain.

Consider now the following differential equation

$$\dot{x}_t = b(x_t),$$

started from $x_0 \in U$, U being an open subset of $\mathbb{R}^d$, and defined for all $t < \zeta$ (maximal solution) and where b is a Lipschitz vector field chosen to be "close" to β in a sense to be made explicit below. So x_t satisfies the equation

$$x_t = x_0 + \int_0^t b(x_s)\mathrm{d}s. \tag{1.2}$$

Our goal is to prove convergence of the Markov chain, or more precisely of the function $\boldsymbol{x}$ of the Markov chain, to the solution of the differential equation under some assumptions to be analyzed below. What we are basically going to do in what follows is to compare the equations (1.1) and (1.2). To this end, we define the following events:

$$\Omega_0 = \{|\boldsymbol{x}(X_0) - x_0| \le \delta\}, \quad \Omega_1 = \{\int_0^{T\wedge t_0} |\beta(X_t) - b(\boldsymbol{x}(X_t))|\mathrm{d}t \le \delta\} \quad \text{and}$$
$$\Omega_2 = \{\int_0^{T\wedge t_0} \alpha(X_t)\mathrm{d}t \le At_0\},$$

where α is the variance field given by

$$\alpha(\xi) = \sum_{\xi' \neq \xi} |\boldsymbol{x}(\xi') - \boldsymbol{x}(\xi)|^2 q(\xi, \xi')$$

and $T = \inf\{t \ge 0 : \boldsymbol{x}(X_t) \notin U\}$.

We assume that b is Lipschitz with Lipschitz constant K on U with respect to the Euclidean norm $|.|$ and we take t_0 such that

$$\forall \xi \in S \quad \text{and} \quad t \le t_0, \quad |\boldsymbol{x}(\xi) - x_t| \le \varepsilon \Longrightarrow \boldsymbol{x}(\xi) \in U.$$

We also take $\varepsilon > 0$, $\delta = \varepsilon e^{-Kt_0}/3$ and $A > 0$. The reason why we define α is to control the expectation of the square of the martingale term appearing in (1.1), because this martingale term, M, satisfies the following equation:

$$M_t^2 = \text{local martingale} + \int_0^t \alpha(X_s)\mathrm{d}s.$$

Now we are ready to state the main theorem which we are going to use in various examples.

Theorem 1.1.1. *Under the above assumptions, we have the following convergence of the Markov chain to the solution of the differential equation:*

$$P(\sup_{t \le t_0} |\boldsymbol{x}(X_t) - x_t| > \varepsilon) \le \frac{4At_0}{\delta^2} + P(\Omega_0^c \cup \Omega_1^c \cup \Omega_2^c). \tag{1.3}$$

From here we see that in order to have a useful estimate, A must be small compared to ε^2. In the applications to follow, we will always try to prove that the second term appearing in the right hand side of (1.3) is sufficiently small.

There are many physical models where this theorem can be applied. However, there are many models of interest where we cannot apply it to approximate a given Markov chain.

The problem arises, because some components of the Markov chain oscillate rapidly and randomly and so there is no hope for a differential equation approximation for them. We may though want to approximate the slow components, but we cannot ignore the presence of the fast components, because the drift vector field may depend on them. A way to overcome this problem is proposed in [2]. The idea is to change the coordinate process by adding a corrector term so that the drift vector field for the modified process loses its strong dependence on the fast components. We will then be able to approximate using the machinery already introduced the modified process by a differential equation. Finally, we will have to prove that the corrector remains small in a suitable sense, so that we can transfer the approximation back to the original coordinate process which is of interest to us. So, the question that arises now is how to find the corrector so as to satisfy the above requirements. In the examples we elaborate, a closed form of the corrector is found and we can get this in two different ways. Either by straightforward calculations or by using a formula for the corrector given in [2]. Here is what we need in order to apply the homogenization technique analysed in [2].

1.2 Homogenization

Suppose that for the coordinate process $\boldsymbol{x}$ the drift vector field is not only a function of $\boldsymbol{x}(\xi)$ but also a function of the other components that are the "fast" ones, so it is given by $b(\boldsymbol{x}(\xi), y)$ and $Y = y(X)$ oscillates rapidly and randomly, where $y : S \to I$. In the examples we will encounter, when we freeze the value of X to some $x \in S$, then the process Y behaves "aproximately" as a Markov chain with transition rates $G = (g(x, y, y') : y, y' \in S)$ for a fixed x. In our cases, this Markov chain is positive recurrent and so there exists a stationary distribution which for a fixed x we denote by $\pi(x, y)$. Now, we set

$$\bar{b}(x) = \sum_y b(x, y)\pi(x, y),$$

which will give us the vector field for the differential equation.
Next, we consider two Markov chains Y and Z started from y and z respectively each with the above generator and coupled so that they meet at time $T = \inf\{t \geq 0 : Y_t = Z_t\}$. Here is then the expression for the corrector:

$$\chi(x, y) = E\left[\int_0^T (b(x, Z_t) - b(x, Y_t))\mathrm{d}t\right].$$

We are free to start Z from wherever we like and we are also free to choose the coupling of the two processes that gives us the smallest corrector. In [2] there are a few assumptions imposed on b (for instance boundedness of b), which in our cases didn't hold, but still we were able to prove that the corrector remained "small" with high probability.

Here is what we do when we apply the homogenization technique:
We change the coordinate process to $\bar{\boldsymbol{x}}$ by substracting the corrector:

$$\bar{\boldsymbol{x}}(\xi) = \boldsymbol{x}(\xi) - \chi(\boldsymbol{x}(\xi), y(\xi)).$$

Next, we compute the new drift vector field $\bar{\beta}$ for the new coordinate process $\bar{\boldsymbol{x}}$ and then we apply Theorem 1.1 with coordinate process $\bar{\boldsymbol{x}}$, drift vector field $\bar{\beta}$ and vector field for the differential equation given by $\bar{b}$ defined above. Hence, we want to show that $\bar{\boldsymbol{x}}(X_t)$ can be approximated by the solution of the differential equation:

$$\dot{x}_t = \bar{b}(x_t).$$

But, our primary goal was to approximate $\boldsymbol{x}$, so we want to transfer somehow this approximation to the original process of interest to us. So, what we are going to do in the applications to come is to show that the corrector remains small in a suitable sense.

Chapter 2

Approximation to the viral model

2.1 Introduction

In this chapter we will find a differential equation approximation to the viral model considered in [1] illustrating the homogenization technique analyzed in [2]. This model has been analyzed in [1] and a differential equation approximation has been found there. The novelty here is that we are using the homogenization analysis in order to obtain the corrector and the vector field for the differential equation. Obviously we will obtain the same differential equation as in [1]. We also worked out the corrector and it turned out to be the same as in [1] again, hence the proof of the convergence of the Markov model to the solution of the differential equation will be exactly the same as the one given in [1] and we will not include it here. For the sake of completeness though, we will introduce the model here (taken from [1, p.53]):

There are three species, G, T and P which represent the genome, template and structural protein of a virus, respectively. We denote by ξ^1, ξ^2, ξ^3 the respective numbers of molecules of each type. There are six reactions, forming a process which may lead from a single virus genome to a sustained population of all three species and to the production of the virus. We write the reactions as follows:

$$G \xrightarrow{\lambda} T, \quad T \xrightarrow{R/\alpha} \varnothing, \quad T \xrightarrow{R} T + G,$$
$$T \xrightarrow{RN} T + P, \quad P \xrightarrow{R/\mu} \varnothing, \quad G + P \xrightarrow{\nu/N} \varnothing.$$

Here, $\alpha > 1$, $R \geq 1$, $N \geq 1$ and $\lambda, \mu, \nu > 0$ are given parameters that are of order 1. We are looking for an approximation to the genome process $(\xi_t^1)_{t\geq 0}$, which in this case is the "slow " component, while the processes (ξ_t^2) and (ξ_t^3) are the "fast" ones.

2.2 Finding the drift vector field $\bar{b}$

In order to apply the homogenization technique we firstly have to choose the function $b(x, y)$. In this example $X_t = ({\xi_t}^1, {\xi_t}^2, {\xi_t}^3)$, $\boldsymbol{x}(\xi) = \xi^1/R$ and $Y = y(X) = (\xi^2, \xi^3/N)$ and the obvious choice of b is

$$b(x, y) = -\lambda x + y_2 - \nu x y_3, \quad \text{where} \quad y = (y_2, y_3).$$

Given that $\xi^1/R = x$ we see that Y_t evolves as a Markov chain. The first component ξ_t^2 evolves as an $M_{\lambda x R}/M_{\frac{R}{\alpha}}/\infty$ queue and, conditional on ξ^2, ξ_t^3 evolves as a queue with time-dependent arrival rate, i.e. it will be an $M_{RN\xi_t^2}/M_{\frac{R}{\mu}}/\infty$ queue. So the Q-matrix $G_x = (g(x, y, y') : y, y' \in I)$ will be as follows:

$$g(x, y, y') = \begin{cases} \lambda x R, & \text{if} \quad y' = y + (1, 0) \\ \frac{R}{\alpha} y_2, & \text{if} \quad y' = y - (1, 0) \\ RN y_2, & \text{if} \quad y' = y + (0, \frac{1}{N}) \\ \frac{R}{\mu} N y_3, & \text{if} \quad y' = y - (0, \frac{1}{N}) \end{cases}, \text{where} \quad y = (y_2, y_3).$$

So in this case with this choice of G_x we have that $\gamma(\xi, y') = g(\boldsymbol{x}(\xi), y(\xi), y')$. (This will give later that $\Delta_2(\xi) = 0$.) It can be seen that the matrix G_x is irreducible and positive recurrent. Positive recurrence can be proved by means of a Lyapunov function. The proof follows along the same lines as in [5, p.170,171]. Since it is positive recurrent, it follows that there exists a stationary distribution $\pi(x, y)$ for each Q-matrix G_x. It seems difficult or even impossible to calculate the invariant distribution. Though in order to apply the homogenization technique we need the existence of the stationary distribution and secondly we need to compute the vector field $\bar{b}$ which will give us the differential equation $\dot{x}_t = \bar{b}(x_t)$. The vector field $\bar{b}$ is given by

$$\bar{b}(x) = \sum_{y \in I} \pi(x, y) b(x, y) = \sum_{y_2, y_3} (-\lambda x + y_2 - \nu x y_3) \pi(x, y) = -\lambda x + E_\pi(\xi_t^2) - \frac{\nu}{N} x E_\pi(\xi_t^3).$$

We then see that what we need is the expectation under the stationary distribution of the first and the second component. For ξ_t^2 we know that the stationary distribution is Poisson of parameter $\alpha \lambda x$, hence the required expectation will be equal to $\alpha \lambda x$. Now for ξ_t^3 we will proceed by finding its Laplace transform. To do so, conditional on ξ^2, we express ξ_t^3 as follows

$$\xi_t^3 = \sum_{k=1}^{\xi_0^3} \mathbf{1}(S_k \geq t) + \int_0^t \int_0^1 \mathbf{1}(u \leq e^{-\frac{R}{\mu}(t-s)}) m(\mathrm{d}s, \mathrm{d}u),$$

where S_k for $k = 1, \ldots, \xi_0^3$ are the service times of the customers that are initially in the system (at time 0) and m is a Poisson random measure on $[0, \infty) \times [0, 1]$ of intensity $\xi_s^2 \mathrm{d}s \mathrm{d}u$ and counts the number of customers that came after time 0 and are still in the system. So, these two summands are independent by the Markov property. Let's now compute the Laplace transform of the second summand.

$$\begin{aligned} & E\left[\exp\left(\theta \int_0^t \int_0^1 \mathbf{1}(u \leq e^{-\frac{R}{\mu}(t-s)}) m(\mathrm{d}s, \mathrm{d}u)\right)\right] \\ = & E\left[E\left[\exp\left(\theta \int_0^t \int_0^1 \mathbf{1}(u \leq e^{-\frac{R}{\mu}(t-s)}) m(\mathrm{d}s, \mathrm{d}u)\right) \middle| \xi_s^2, s \leq t\right]\right] \end{aligned}$$

By conditioning upon $\xi_s^2, s \leq t$ we actually obtain the integral of a function with respect to a Poisson random measure of deterministic intensity, which we know how to compute by Campbell's formula. Hence,

$$E\left[\exp\left(\theta \int_0^t \int_0^1 \mathbf{1}(u \leq e^{-\frac{R}{\mu}(t-s)}) m(\mathrm{d}s, \mathrm{d}u)\right)\right] = E\left[\exp\left(-RN(1 - e^\theta) e^{-\frac{R}{\mu}t} \int_0^t e^{\frac{R}{\mu}s} \xi_s^2 \mathrm{d}s\right)\right].$$

Differentiating now with respect to θ we get

$$\frac{\mathrm{d}}{\mathrm{d}\theta}E\left[\exp\left(\theta\int_0^t\int_0^1 \mathbf{1}(u\le e^{-\frac{R}{\mu}(t-s)})m(\mathrm{d}s,\mathrm{d}u)\right)\right]\Bigg|_{\theta=0} = E\left[RNe^{-\frac{R}{\mu}t}\int_0^t e^{\frac{R}{\mu}s}\xi_s^2\mathrm{d}s\right]. \quad (2.1)$$

We want to compute the expectation under the stationary distribution, so by interchanging the integral and the expectation in the last expression using Fubini's theorem and using the fact that under stationarity ξ_t^2 is Poisson$(\alpha\lambda x)$ we obtain that (3.4.2) is equal to

$$RNe^{-\frac{R}{\mu}t}\int_0^t e^{\frac{R}{\mu}s}\alpha\lambda x\mathrm{d}s = \alpha\lambda\mu N x - \alpha\lambda\mu N x e^{-\frac{R}{\mu}t}.$$

Hence, putting things together we deduce that

$$\begin{aligned} E_\pi\left[\xi_t^3\right] &= E_\pi\left[\sum_{k=0}^{\xi_0^3}\mathbf{1}(S_k\ge t)\right] + \alpha\lambda\mu N x - \alpha\lambda\mu N x e^{-\frac{R}{\mu}t} \\ &= \alpha\lambda\mu N x - \alpha\lambda\mu N x e^{-\frac{R}{\mu}t} + E_\pi[\xi_0^3]e^{-\frac{R}{\mu}t}. \end{aligned}$$

Since we are in stationary regime, the expectation under stationarity of ξ_t^3 should not depend upon time t. Hence, in the last expression we must have that

$$\alpha\lambda\mu N x e^{-\frac{R}{\mu}t} = E_\pi[\xi_0^3]e^{-\frac{R}{\mu}t}.$$

Concluding, we get that

$$E_\pi[\xi_t^3] = \alpha\lambda\mu N x.$$

So, $\bar{b}$ will be given by

$$\bar{b}(x) = -\lambda x + \alpha\lambda x - \frac{\nu}{N}x\alpha\lambda\mu N x = \lambda(\alpha-1)x - \lambda\alpha\mu\nu x^2. \quad (2.2)$$

2.2.1 Another way of calculating $E_\pi[\xi_t^3]$

Conditional on ξ^2 we can express ξ^3 as follows:

$$\xi_t^3 = \xi_0^3 + A\left(RN\int_0^t \xi_s^2\mathrm{d}s\right) - D\left(\frac{R}{\mu}\int_0^t \xi_s^3\mathrm{d}s\right),$$

where A and D are two Poisson point processes of rate 1. So, the expectation under π of ξ_t^3 will be given by

$$\begin{aligned} E_\pi\left(\xi_t^3\right) &= E_\pi\left(\xi_0^3\right) + E_\pi\left(RN\int_0^t \xi_s^2\mathrm{d}s\right) - E_\pi\left(\frac{R}{\mu}\int_0^t \xi_s^3\mathrm{d}s\right) \\ &= E_\pi\left(\xi_0^3\right) + RN\alpha\lambda x t - \frac{R}{\mu}E_\pi\left(\xi_0^3\right)t, \end{aligned}$$

where again the interchange of integral and expectation is justified by Fubini's theorem and we substituted $E_\pi\left(\xi_s^3\right)$ in the second integral by $E_\pi\left(\xi_0^3\right)$, since we are in stationary regime. So, we will have that

$$\begin{aligned} E_\pi\left(\xi_0^3\right) &= E_\pi\left(\xi_0^3\right) + RN\alpha\lambda x t - \frac{R}{\mu}E_\pi\left(\xi_0^3\right)t, \forall t \\ &\Rightarrow E_\pi\left(\xi_0^3\right) = \alpha\lambda N\mu x \end{aligned}$$

2.3 Finding the corrector χ

Following [2] we have the following expression for the corrector χ

$$\chi(x,y) = E\left[\int_0^T (b(x,Z_t) - b(x,Y_t))\mathrm{d}t\right] = E\left[\int_0^T (\zeta_t^2 - \xi_t^2) - \frac{\nu x}{N}(\zeta_t^3 - \xi_t^3)\mathrm{d}t\right],$$

where $T = \inf\{t \geq 0 : Y_t = Z_t\}$, Y is the process described above starting from (y_2, y_3) (in what follows $y_2 = \xi_0^2$ and $y_3 = \frac{\xi_0^3}{N}$), and Z is a process constructed in the following way: we start it from $(0,0)$ and we allow the first component ζ_t^2 to evolve as an $M_{\lambda x R}/M_{\frac{R}{\alpha}}/\infty$ by taking the same arrival and departure process for it as for ξ_t^2. The first components will meet then after all the y_2 customers that were initially in queue ξ_t^2 leave.
Regarding the second components we couple them as follows: up until the first time $T_0 = \inf\{t \geq 0 : \xi_t^2 = \zeta_t^2\}$ that the first components become equal we perform a thinning. (What follows is not rigorous.) By that we mean that whenever we have an arrival in $[t, t+h]$ for ξ_t^3 we take the same arrival for ζ_t^3 with probability $\frac{\int_t^{t+h}\zeta_s^2 \mathrm{d}s}{\int_t^{t+h}\xi_s^2 \mathrm{d}s}$, otherwise with probability $1 - \frac{\int_t^{t+h}\zeta_s^2 \mathrm{d}s}{\int_t^{t+h}\xi_s^2 \mathrm{d}s}$ there is no arrival for ζ_t^3. Thus, we obtain a new Poisson random measure of intensity $RN(\xi_s^2 - \zeta_s^2)\mathrm{d}s$. So, in this way we count only the times that we had an arrival in ξ_t^3 and not in ζ_t^3. The departures of the common arrivals are then taken to be the same in both queues. So, conditional on ξ^2, ζ^2, up until time T_0 the difference $\xi_t^3 - \zeta_t^3$ can be expressed as the sum of an integral with respect to a Poisson random measure of intensity $RN(\xi_s^2 - \zeta_s^2)\mathrm{d}s\mathrm{d}u$ and the term $\sum_{k=1}^{\xi_0^3} \mathbf{1}(S_k \geq t)$, where S_k denote again the service times of the customers initially in the queue. From time T_0 and on we will have to wait until the $\xi_{T_0}^3 - \zeta_{T_0}^3$ customers leave the ξ^3 queue. ($\xi_{T_0}^3 > \zeta_{T_0}^3$ by the construction of the coupling.)
A more rigorous way proceeds as follows: Conditional on ξ^2, ζ^2, take two independent Poisson random measures, m with intensity $RN(\xi_s^2 - \zeta_s^2)\mathrm{d}s\mathrm{d}u$ and m' with intensity $RN\zeta_s^2\mathrm{d}s\mathrm{d}u$. Then their sum will be a Poisson random measure of intensity their sum, i.e. $RN\xi_s^2\mathrm{d}s\mathrm{d}u$. So, now write

$$\xi_t^3 = \int_0^t \int_0^1 \mathbf{1}\left(u \leq e^{-\frac{R}{\mu}(t-s)}\right) m(\mathrm{d}s, \mathrm{d}u) + \int_0^t \int_0^1 \mathbf{1}\left(u \leq e^{-\frac{R}{\mu}(t-s)}\right) m'(\mathrm{d}s, \mathrm{d}u) + \sum_{k=1}^{\xi_0^3} \mathbf{1}(S_k \geq t)$$

and also

$$\zeta_t^3 = \int_0^t \int_0^1 \mathbf{1}\left(u \leq e^{-\frac{R}{\mu}(t-s)}\right) m(\mathrm{d}s, \mathrm{d}u).$$

We will now have

$$E\left[\int_0^T (\zeta_t^2 - \xi_t^2)\mathrm{d}t\right] = E\left[\int_0^{T_0} (\zeta_t^2 - \xi_t^2)\mathrm{d}t\right] = E\left[-\int_0^{T_0} \sum_{k=1}^{\xi_0^2} \mathbf{1}(S_k \geq t)\mathrm{d}t\right]$$
$$= -\sum_{k=1}^{\xi_0^2} E\left[\int_0^{T_0} \mathbf{1}(S_k \geq t)\mathrm{d}t\right] = -\sum_{k=1}^{\xi_0^2} E\left[S_k\right] = -\xi_0^2 \frac{\alpha}{R},$$

since $T_0 = \max(S_k, k = 1, \ldots, \xi_0^3)$. Also,

$$E\left[\int_0^T (\xi_t^3 - \zeta_t^3)\mathrm{d}t\right] = E\left[\int_0^{T_0}\int_0^t\int_0^1 \mathbf{1}(u \le e^{-\frac{R}{\mu}(t-s)})m(\mathrm{d}s,\mathrm{d}u)\mathrm{d}t\right]$$
$$+\sum_{k=1}^{\xi_0^3} E\left[\int_0^T \mathbf{1}(S_k \ge t)\mathrm{d}t\right] + E\left[\int_{T_0}^T (\xi_{T_0}^3 - \xi_0^3 - \zeta_{T_0}^3)\mathrm{d}t\right], \qquad (2.3)$$

where m is as described above, i.e. a Poisson random measure of intensity $RN(\xi_s^2 - \zeta_s^2)\mathrm{d}s\mathrm{d}u$. Let's now work out the first term on the right hand side of (2.3).

$$E\left[\int_0^{T_0}\int_0^t\int_0^1 \mathbf{1}(u \le e^{-\frac{R}{\mu}(t-s)})m(\mathrm{d}s,\mathrm{d}u)\mathrm{d}t\right]$$
$$= E\left[E\left[\int_0^{T_0}\int_0^t\int_0^1 \mathbf{1}(u \le e^{-\frac{R}{\mu}(t-s)})m(\mathrm{d}s,\mathrm{d}u)\mathrm{d}t\middle|T_0, (\xi_s^2, \zeta_s^2, s \le T_0)\right]\right]$$
$$= E\left[\int_0^{T_0} E\left[\int_0^t\int_0^1 \mathbf{1}(u \le e^{-\frac{R}{\mu}(t-s)})m(\mathrm{d}s,\mathrm{d}u)\middle|T_0, (\xi_s^2, \zeta_s^2, s \le T_0)\right]\mathrm{d}t\right]$$

Using Campbell's formula again, we get that this last term is equal to

$$E\left[\int_0^{T_0} RNe^{-\frac{R}{\mu}t}\int_0^t e^{\frac{R}{\mu}s}(\xi_s^2 - \zeta_s^2)\mathrm{d}s\mathrm{d}t\right] = E\left[\int_0^{T_0} RNe^{-\frac{R}{\mu}t}\int_0^t e^{\frac{R}{\mu}s}\sum_{k=1}^{\xi_0^2}\mathbf{1}(S_k \ge s)\mathrm{d}s\mathrm{d}t\right]$$
$$= \sum_{k=1}^{\xi_0^2} E\left[\int_0^{T_0} RNe^{-\frac{R}{\mu}t}\int_0^{t\wedge S_k} e^{\frac{R}{\mu}s}\mathrm{d}s\mathrm{d}t\right] = \sum_{k=1}^{\xi_0^2} E\left[\int_0^{T_0} RNe^{-\frac{R}{\mu}t}\left(\frac{e^{\frac{R}{\mu}(t\wedge S_k)} - 1}{\frac{R}{\mu}}\right)\mathrm{d}t\right]$$
$$= \sum_{k=1}^{\xi_0^2} E\left[\int_0^{S_k} \mu Ne^{-\frac{R}{\mu}t}\left(e^{\frac{R}{\mu}t} - 1\right)\mathrm{d}t + \int_{S_k}^{T_0} \mu Ne^{-\frac{R}{\mu}t}\left(e^{\frac{R}{\mu}S_k} - 1\right)\mathrm{d}t\right]$$
$$= \sum_{k=1}^{\xi_0^2} E\left[\mu NS_k + \frac{\mu^2 N}{R}e^{-\frac{R}{\mu}S_k} - \frac{\mu^2 N}{R} - \frac{\mu^2 N}{R}e^{-\frac{R}{\mu}(T_0 - S_k)} + \frac{\mu^2 N}{R} + \frac{\mu^2 N}{R}(e^{-\frac{R}{\mu}T_0} - e^{-\frac{R}{\mu}S_k})\right]$$
$$= \mu N\frac{\alpha}{R}\xi_0^2 - \sum_{k=1}^{\xi_0^2}\frac{\mu^2 N}{R}E\left[e^{-\frac{R}{\mu}(T_0 - S_k)}\right] + \xi_0^2\frac{\mu^2 N}{R}E\left[e^{-\frac{R}{\mu}T_0}\right]$$

The second term of (2.3) gives

$$\sum_{k=1}^{\xi_0^3} E\left[\int_0^T \mathbf{1}(S_k \ge t)\mathrm{d}t\right] = \frac{\mu}{R}\xi_0^3,$$

by the same reasoning as before (for ξ_t^2).

The third term of (2.3) is given by $\frac{\mu}{R}E[\xi_{T_0}^3 - \xi_0^3 - \zeta_{T_0}^3]$, again by the same reasoning (we are waiting for the $\xi_{T_0}^3 - \xi_0^3 - \zeta_{T_0}^3$ to leave the ξ^3 queue).

What we have to do last is to compute $E[\xi_{T_0}^3 - \xi_0^3 - \zeta_{T_0}^3]$. But, as discussed before this can be written in the form $E\left[\int_0^{T_0}\int_0^1 \mathbf{1}(u \le e^{-\frac{R}{\mu}(T_0 - s)})m(\mathrm{d}s,\mathrm{d}u)\right]$, where conditional on ξ^2, ζ^2,

m is a Poisson random measure of intensity $RN(\xi_s^2 - \zeta_s^2)\mathrm{d}s\mathrm{d}u$. Again by the same trick of conditioning upon $(T_0, (\xi_s^2, \zeta_s^2, s \le T_0))$ we obtain by applying Campbell's formula once more that

$$E\left[\int_0^{T_0}\int_0^1 \mathbf{1}(u \le e^{-\frac{R}{\mu}(T_0-s)})m(\mathrm{d}s,\mathrm{d}u)\right] = E\left[RNe^{-\frac{R}{\mu}T_0}\int_0^{T_0} e^{\frac{R}{\mu}s}(\xi_s^2 - \zeta_s^2)\mathrm{d}s\right]$$

$$= E\left[RNe^{-\frac{R}{\mu}T_0}\int_0^{T_0} e^{\frac{R}{\mu}s}\sum_{k=1}^{\xi_0^2}\mathbf{1}(S_k \ge s)\mathrm{d}s\right] = \sum_{k=1}^{\xi_0^2} RNE\left[e^{-\frac{R}{\mu}T_0}\left(\frac{e^{\frac{R}{\mu}S_k} - 1}{\frac{R}{\mu}}\right)\right]$$

$$= \sum_{k=1}^{\xi_0^2}\left(\mu NE\left[e^{-\frac{R}{\mu}(T_0-S_k)}\right] - \mu NE\left[e^{-\frac{R}{\mu}T_0}\right]\right) = \sum_{k=1}^{\xi_0^2}\mu NE\left[e^{-\frac{R}{\mu}(T_0-S_k)}\right] - \xi_0^2\mu NE\left[e^{-\frac{R}{\mu}T_0}\right].$$

So, the third term of (2.3) will be given by

$$\sum_{k=1}^{\xi_0^2}\frac{\mu^2 N}{R}E\left[e^{-\frac{R}{\mu}(T_0-S_k)}\right] - \xi_0^2\frac{\mu^2 N}{R}E\left[e^{-\frac{R}{\mu}T_0}\right].$$

Putting now things together, we conclude that

$$E\left[\int_0^T (\xi_t^3 - \zeta_t^3)\mathrm{d}t\right] = \mu N\frac{\alpha}{R}\xi_0^2 + \frac{\mu}{R}\xi_0^3.$$

Hence, the corrector will be given by

$$\chi(x,y) = \frac{1}{R}\Big(- \alpha y_2 + \alpha\mu\nu x y_2 + \mu\nu x y_3\Big),$$

which has exactly the same form as the corrector given in [1, p.15].

Chapter 3

Random 3-Satisfiability problem

3.1 Introduction

In this chapter we will analyze the UCWM (unit clause with majority) algorithm for the random 3-satisfiability problem using the homogenization technique. First, though, we will introduce the model and the questions we are trying to answer.

We start with N Boolean variables, say $x_1, \ldots, x_N$, which take values in $\{0, 1\}$, 0 meaning false and 1 meaning true. A k-clause is a set of k literals (literal=a variable or the complement of it) joined with $\vee$. For instance the expression $x \vee y \vee z$ is a 3-clause.
Now we choose 3 variables out of the N uniformly at random without replacement (there are $\binom{N}{3}$ ways of choosing them) and for each of those we toss a fair coin and so with probability $\frac{1}{2}$ we take the complement of a variable and with probability $\frac{1}{2}$ we put the variable as is in the 3-clause. In this way we have obtained a 3-clause and then we proceed in the same way to construct rN 3-clauses, where r is a positive constant. We proceed independently over different clauses and finally we have rN clauses which we join with $\wedge$. We denote the formula we have obtained by $F(N, rN)$ and it is said to be in conjunctive normal form.

The question that arises now is whether there exists a truth assignment to the variables such that the formula is true. This problem has attracted the interest of many scientists and in particular the following conjecture has gained significant popularity since it was first put forward.

> The **satisfiability threshold conjecture** states that for every $k \geq 2$, there exists a constant r_k such that for all $\varepsilon > 0$,
>
> $$\lim_{N\to\infty} P(F(N,(r_k-\varepsilon)N) \text{ is satisfiable}) = 1 \text{ and } \lim_{N\to\infty} P(F(N,(r_k+\varepsilon)N) \text{ is satisfiable}) = 0.$$

For $k = 2$, this constant has been proven to be equal to 1. For $k \geq 3$ much less is known and not even the existence of the constant has been established. A big step towards proving the above conjecture was made by Friedgut with a theorem showing that there exists a sharp threshold around some critical sequence of values.

Theorem 3.1.1 (Friedgut). *For every $k \geq 2$, there exists a sequence $r_k(N)$ such that for all $\varepsilon > 0$,*

$$\lim_{N\to\infty} P(F(N,(r_k(N)-\varepsilon)N) \text{ is satisfiable}) = 1 \text{ and } \lim_{N\to\infty} P(F(N,(r_k(N)+\varepsilon)N) \text{ is satisfiable}) = 0$$

We are going to use the following corollary of the above theorem

Corollary 3.1.1. *If r is such that $\underline{\lim}_{N\to\infty} P(F(N, rN)$ is satisfiable$) > 0$, then for any $\varepsilon > 0$,*

$$\lim_{N\to\infty} P(F(N,(r-\varepsilon)N) \text{ is satisfiable}) = 1$$

All upper bounds for the constant r_3 have been proved via probabilistic counting arguments, while the lower bounds are algorithmic, i.e., in each case a particular algorithm is shown to satisfy $F(N, rN)$ w.h.p. for r below a certain value r^*. Note that if we show positive probability of success for $r < r^*$ it suffices to deduce $r_3 \geq r^*$, in view of Corollary 3.1.1.

In this section we are going to analyze the unit clause with majority algorithm (UCWM) to derive a lower bound for the constant r_3. This algorithm has been analyzed using an ODE approximation to a Markov chain in [4]. In [4] though, the author had to modify the algorithm a bit, because the Markov chain corresponding to the original one had a fast component which couldn't be approximated and a classical theorem due to Wormald couldn't be applied. The novelty here is that with the homogenization technique we succeeded in analyzing the original algorithm, hence obtaining a slightly better bound. Basically, the differential equations we obtained were the limit as $\theta \to 0$ of the differential equations in [4].

3.2 UCWM algorithm

Here we have the UCWM algorithm:

<u>Unit clause with majority</u>

1. If there are any 1-clauses, then pick a 1-clause uniformly at random and satisfy it
2. Otherwise,
   (a) Pick an unset variable x uniformly at random
   (b) If x appears positively in at least half the remaining 3-clauses, then set x = True
   (c) Otherwise, set x = False

Let $(C_i(n), i = 1, 2, 3)$ be the number of i clauses at step n of the execution of the algorithm. We are going to embed that in continuous time, so as to apply the fluid limit result. Let ν_t be a Poisson process of rate N, which stops when $\nu_t = (1-\varepsilon)N$. Consider now the process

$$X(t) = (\nu_t, C_1(\nu_t), C_2(\nu_t), C_3(\nu_t)),$$

which stops at the time $T_0 = \inf\{t \geq 0 : \nu_t = (1-\varepsilon)N\}$.

An explanation why $X(t)$ is a Markov chain and a justification of the form of the transition probabilities for the discrete chain are given in [4]. Here we will only write down the transition probabilities which are given as follows: $(\Delta C_i(n) = C_i(n+1) - C_i(n))$

- If $C_1(n) = 0$, then $\Delta C_2(n) = Y - Z$ and $\Delta C_3(n) = -X$, where $Y \sim Y_1 \wedge Y_2$, where Y_1 is the number of occurences of the selected literal u in 3-clauses and Y_2 is the number of occurences of $\bar{u}$ in 3-clauses, so $Y_1 \sim \text{Bin}\left(C_3(n), \frac{3}{2(N-n)}\right)$, $Y_2 \sim \text{Bin}\left(C_3(n), \frac{3}{2(N-n)}\right)$ and $Y_1 + Y_2 \sim \text{Bin}\left(C_3(n), \frac{3}{N-n}\right)$. Also $X \sim \text{Bin}\left(C_3(n), \frac{3}{N-n}\right)$.
- If $C_1(n) \neq 0$, then $\Delta C_2(n) = Y' - Z$ and $\Delta C_3(n) = -X$, where again as before $X \sim \text{Bin}\left(C_3(n), \frac{3}{N-n}\right)$, $Z \sim \text{Bin}\left(C_2(n), \frac{2}{N-n}\right)$ and $Y' \sim \text{Bin}\left(C_3(n), \frac{3}{2(N-n)}\right)$.

In the following sections we are going to approximate the number of 2 and 3-clauses by a differential equation. For the sake of completeness, we will now explain briefly how the differential equations will help us derive a lower bound for the constant r_3. The theorems stated below have been taken from [4] but here we present the proofs in more detail.

The next theorem states that if the density of the 2-clauses stays bounded away from 1 during the execution of the algorithm, then the formula is satisfiable (i.e. true) with probability tending to 1 as $N \to \infty$. Here we have an intuitive explanation of this result: from the above transition probabilities we see that the rate at which 1-clauses are generated at step n of the algorithm is given by $\frac{C_2(n)}{N-n}$. When a unit clause exists then at the next step we satisfy it. So, we can think of it as a queue with one server, where at each time step one customer is served and leaves the system and the arrival rate is bounded by 1. This queue is then stable and we prove this result for this particular process in Theorem 3.4.3.

Theorem 3.2.1. *If there exists $\delta > 0$ such that*

$$\frac{C_2(n)}{N-n} < 1 - \delta, \forall n = 0, \ldots, (1-\varepsilon)N, \text{ then } r^* \geq r.$$

Proof. We note that if at step n there are no unit clauses, then the probability of a 0-clause being generated is 0. Conditional though on $C_1(n) = a$, then

$$P(\text{ no 0-clause after step n }) = \left(1 - \frac{1}{2(N-n)}\right)^{a-1}.$$

Now we recall that we stop the algorithm after $(1-\varepsilon)N$ steps, so the probability that there are no 0-clauses at the end of the algorithm is given by:

$$P(C_0((1-\varepsilon)N) = 0) = E[P(C_0((1-\varepsilon)N) = 0 | C_1(n), n = 0, \ldots, (1-\varepsilon)N)]$$
$$= E\left[\left(1 - \frac{1}{2N}\right)^{C_1(0)-1} \cdots \left(1 - \frac{1}{2(N-(1-\varepsilon)N)}\right)^{C_1((1-\varepsilon)N)-1}\right] >$$
$$> E\left[\left(1 - \frac{1}{2(N-(1-\varepsilon)N)}\right)^{\sum_{n=0}^{(1-\varepsilon)N} C_1(n) - (1-\varepsilon)N}\right] = E\left[\left(1 - \frac{1}{2N\varepsilon}\right)^{\sum_{n=0}^{(1-\varepsilon)N} C_1(n) - (1-\varepsilon)N}\right].$$

From Theorem 3.4.3 we see that we can bound this last sum $\sum_{n=0}^{(1-\varepsilon)N} C_1(n) < MN$, so

$$P(C_0((1-\varepsilon)N) = 0) > \left(1 - \frac{1}{2\varepsilon N}\right)^{(M-(1-\varepsilon))N} \to \exp\left(-\frac{M}{2\varepsilon}\right), \text{ as } N \to \infty.$$

Hence, we deduce that with positive probability as $N \to \infty$ there are no 0-clauses, so in view of Corollary 3.1.1 the formula is satisfiable w.h.p. □

If we succeed thus in approximating the number of 2-clauses during the execution of the algorithm by a differential equation, then by looking at the ODE, we can find the values of r such that the solution $c_2(t)$ divided by $1-t$ remains bounded below 1. The reason for introducing $(1-\varepsilon)N$ is clear now, since otherwise the denominator above would become infinite.

3.3 The approximation procedure

We are going to approximate the number of 2 and 3-clauses by applying the homogenization result. The number of 3-clauses is approximated by the same differential equation as in [4], since its dynamics do not depend on the fast variable $C_1(t)$. For the number of 2-clauses though, we are going to use the homogenization technique, because the drift vector field depends heavily on the value of the number of 1-clauses, which cannot be approximated by a differential equation.

We will thus approximate $\boldsymbol{x}(X(t)) = N^{-1}(\nu_t, C_2(\nu_t), C_3(\nu_t))$, which is a Markov chain started from the state $(0,0,r)$ and we will run it up to the stopping time $T = \inf\{t \geq 0 : \frac{C_2(\nu_t)}{N-\nu_t} \geq 1-\delta\} \wedge T_0$. The reason for introducing this stopping time follows from the discussion in the previous section. When we obtain the differential equation though, we will find the values of r such that the density of the 2-clauses is always bounded below 1, so this will imply that this stopping time is basically equal to T_0.

We now have all we need to start applying the homogenization result.

The drift vector field for the process $\boldsymbol{x}(X(t)) = N^{-1}(\nu_t, C_2(\nu_t), C_3(\nu_t))$ is given by

$$\beta^1(\xi) = 1,$$

$$\beta^2(\xi) = \mathbf{1}(\xi^1 = 0)\left(-\xi^2\frac{2}{N-\xi^0} + E\left[\mathrm{Bin}\left(\xi^3, \frac{3}{2(N-\xi^0)}\right) \wedge \mathrm{Bin}\left(\xi^3, \frac{3}{2(N-\xi^0)}\right)\right]\right)$$
$$+\mathbf{1}(\xi^1 \neq 0)\left(-\xi^2\frac{2}{N-\xi^0} + \xi^3\frac{3}{2(N-\xi^0)}\right) = -\frac{2\xi^2}{N-\xi^0} + \mathbf{1}(\xi^1 = 0)M(\lambda) + \mathbf{1}(\xi^1 \neq 0)\lambda,$$

$$\beta^3(\xi) = -\xi^3\frac{3}{N-\xi^0},$$

where $\xi = (\xi^0, \xi^1, \xi^2, \xi^3) \in \{(y^1, y^2, y^3, y^4) \in \mathbb{N}^4 : y^1 \leq (1-\varepsilon)N, \frac{y^2}{N-y^0} < 1\}$ and the two Binomial random variables appearing under the expectation sign are the variables Y_1 and Y_2 defined earlier. Also, $\lambda = \frac{3\xi^3}{2(N-\xi^0)}$ and $M(\lambda) = E\left[\mathrm{Bin}\left(\xi^3, \frac{3}{2(N-\xi^0)}\right) \wedge \mathrm{Bin}\left(\xi^3, \frac{3}{2(N-\xi^0)}\right)\right]$. As $N \to \infty$ the pair (Y_1, Y_2) converges in distribution to a pair of two independent Poisson random variables, because as we discussed above ξ^3/N converges to the solution of the differential equation, which is $c_3(t) = r(1-t)^3$, so the parameter of each will be equal to $\frac{3c_3(t)}{2(1-t)}$. We give here the proof of the convergence in distribution stated above:

Proof.

$$E[e^{\theta_1 Y_1+\theta_2 Y_2}] = \sum_{k+l=0}^{n} \binom{n}{k+l}\binom{k+l}{k} p^{k+l}(1-p)^{n-(k+l)}\left(\frac{1}{2}\right)^{k+l} e^{\theta_1 k}e^{\theta_2 l}$$

$$= \sum_{k=0}^{n}\sum_{l=0}^{n-k} \frac{n!}{k!l!(n-(k+l))!}\left(\frac{p}{2}\right)^{k+l}(1-p)^{n-(k+l)}e^{\theta_1 k}e^{\theta_2 l} = \sum_{k=0}^{n} e^{\theta_1 k}\left(\frac{p}{2}\right)^{k}\binom{n}{k}\left(e^{\theta_2}\frac{p}{2}+1-p\right)^{n-k}$$

$$= \left(e^{\theta_1}\frac{p}{2}+e^{\theta_2}\frac{p}{2}+1-p\right)^{n} \to e^{\frac{\lambda}{2}(e^{\theta_1}-1)}e^{\frac{\lambda}{2}(e^{\theta_2}-1)}, \text{ as } n\to\infty,$$

where $Y_1, Y_2 \sim \mathrm{Bin}(n,p)$ with $np \to \lambda$ as $n \to \infty$.

In order to get rid of this expectation, which doesn't have a closed form expression, we will do exactly the same as in [4], i.e. we will use instead the following bound

$$B_q(\lambda) = \lambda + \sum_{j=0}^{q}\sum_{k=0}^{q} \frac{e^{-2\lambda}\lambda^{j+k}}{j!k!}\left(\min\{j,k\} - \frac{j+k}{2}\right) \geq M(\lambda),$$

where $M(\lambda)$ is the expectation of the minimum of two Poisson random variables with the same parameter and λ stands for $\frac{3\xi^3}{2(N-\xi^0)}$. In [4] they modify the algorithm a bit, so as to be able to substitute $M(\lambda)$ by $B_q(\lambda)$. With probability $p = \frac{\lambda - B_q(\lambda)}{\lambda - M(\lambda)}$ they set the literal u "by majority" and with probability $1-p$ "at random".

As we noted earlier, the drift vector field for the number of 2-clauses depends significantly on the number of 1-clauses. This is where the homogenization technique is going to be used. What we want to do is to compensate the coordinate process $\boldsymbol{x}^2$ so that it takes account of this effect. We seek to find a new coordinate function $\boldsymbol{x}^2$ on the state space of the form

$$\boldsymbol{x}^2(\xi) = \frac{\xi^2}{N} + \chi(\xi),$$

where χ is a small correction so that the drift vector $\beta^2(\xi)$ has the form

$$\beta^2(\xi) = b(\boldsymbol{x}(\xi)) + \frac{\Delta(\xi)}{N},$$

where again $\frac{\Delta(\xi)}{N}$ is small for N large. From now on, we are going to use the technique from [2] and the same notation again.

3.3.1 Finding the drift vector field $\bar{b}$

As we said before, $X_t = (\nu_t, C_1(\nu_t), C_2(\nu_t), C_3(\nu_t))$ and $\boldsymbol{x}(\xi) = N^{-1}(\nu_t, C_2(\nu_t), C_3(\nu_t))$. Moreover, $Y = y(X)$ is the number of 1-clauses, so $Y_t = C_1(t)$. The obvious choice for b is

$$b(x,y) = -2\frac{x^2}{1-x^1} + \mathbf{1}(y \neq 0)\frac{3x^3}{2(1-x^1)} + \mathbf{1}(y=0)M(\lambda),$$

where $\lambda = \frac{3x^3}{2(1-x^1)}$.
Given that $\boldsymbol{x}(\xi) = x$, Y_t can be approximated by a Markov chain, whose Q- matrix $G_x = (g(x, y, y'), y, y' \in \mathbb{N})$ is given below:

$$g(x, y, y') = \begin{cases} Ne^{-\lambda}\frac{\lambda^j}{j!}, & \text{if} \quad y = 0, y' = j \\ Ne^{-\lambda}\frac{\lambda^j}{j!}, & \text{if} \quad y > 0, y' = y - 1 + j, \end{cases}$$

where $\lambda = \frac{x^2}{1-x^1} < 1$. So, this Markov chain basically jumps at the points of a Poisson process of rate N and either it jumps to a Poisson random variable if it is currently at 0, or it jumps to a Poisson -1, if not at 0. Obviously, this Q-matrix is irreducible on the positive integers. We now want to find the stationary distribution π of this Markov chain, and in particular we only need $\pi(0)$ in order to compute the vector field $\bar{b}$.

$$E_0(T_0) = 1 + \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}E_k(T_0),$$

where $E_i(T_0)$ is the expected time to hit 0 starting from i ($T_0 = \inf\{t > 0 : Y_t = 0\}$). It is easy to see that $E_k(T_0) = kE_1(T_0)$, and we also see that

$$E_1(T_0) = 1 + \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}E_k(T_0),$$

which gives that $E_0(T_0) = E_1(T_0)$, hence we deduce that

$$E_0(T_0) = 1 + \lambda E_0(T_0),$$

and since $\lambda < 1$, the minimal nonnegative solution to the above equation is

$$E_0(T_0) = \frac{1}{1-\lambda},$$

which yields that $\pi(0) = 1 - \lambda$.
So, the vector field $\bar{b}$ is

$$\bar{b}(x) = -2\frac{x^2}{1-x^1} + \frac{x^2}{1-x^1}\frac{3x^3}{2(1-x^1)} + \left(1 - \frac{x^2}{1-x^1}\right) M\left(\frac{3x^3}{2(1-x^1)}\right)$$

3.3.2 Finding the corrector χ

Following [2] we have the following expression for the corrector χ

$$\chi(x, y) = E\left[\int_0^T (b(x, Z_t) - b(x, Y_t))dt\right],$$

where $T = \inf\{t \geq 0 : Y_t = Z_t\}$, Y is the process described above starting from y (we shall think of it as having y customers at time 0) and Z is a coupled process started from 0 and having the same arrival process as Y. By positive recurrence, we have that $E[T] < \infty$. Now the corrector will be equal to

$$\chi(x, y) = E\left[\int_0^T \Big(\lambda(\mathbf{1}(Y_t = 0) - \mathbf{1}(Z_t = 0)) + M(\lambda)(\mathbf{1}(Z_t = 0) - \mathbf{1}(Y_t = 0))\Big)dt\right].$$

Let T_i be the points of the underlying Poisson process. In the intervals $[T_i, T_{i+1}]$ when $Z_t = 0$, Y_t will not be 0, unless the initial y customers all leave. So, in these intervals only Z_t will be 0, and there will be y such intervals in total. In the intervals $[T_i, T_{i+1}]$ when Z_t is not 0, nor is Y_t. So, the corrector will be given by

$$\chi(x, y) = (M(\lambda) - \lambda)\frac{y}{N},$$

because $E[T_{i+1} - T_i] = \frac{1}{N}$, ($T_{i+1} - T_i$ being the interarrival times of the Poisson process of rate N).

3.3.3 Calculations

What remains now to be proven is that the corrector is small and that $\bar{\beta}^2$ is close to $\bar{b}$ in an appropriate sense. Basically, $\bar{\beta}^2(\xi) - \bar{b}(\bar{\boldsymbol{x}}(\xi))$ must be small, where

$$\bar{\boldsymbol{x}}(\xi) = \boldsymbol{x}(\xi) - \chi(\boldsymbol{x}(\xi), y(\xi)) = \frac{\xi^2}{N} - \frac{\xi^1}{N}(M(\lambda) - \lambda)$$

and $\bar{\beta}^2$ is the drift vector field for the modified coordinate function $\bar{\boldsymbol{x}}$.

After some straightforward calculations, $\bar{\beta}^2$ is given by:

$$\begin{aligned}
\bar{\beta}^2(\xi) = \sum_{\xi' \neq \xi} \left(\bar{\boldsymbol{x}}(\xi') - \bar{\boldsymbol{x}}(\xi)\right) q(\xi, \xi') = \beta^2(\xi) + \frac{3\xi^2\xi^3}{2(N - \xi^0)(N - \xi^0 - 1)} \\
-\frac{1}{(N - \xi^0 - 1)}\frac{9\xi^2\xi^3}{2(N - \xi^0)^2} - \mathbf{1}(\xi^1 \neq 0)\frac{3\xi^1\xi^3}{2(N - \xi^0)} + \mathbf{1}(\xi^1 \neq 0)\frac{3\xi^1\xi^3}{2(N - \xi^0 - 1)} \\
-\mathbf{1}(\xi^1 \neq 0)\frac{1}{2(N - \xi^0 - 1)}\frac{9\xi^1\xi^3}{(N - \xi^0)} - \mathbf{1}(\xi^1 \neq 0)\frac{1}{2(N - \xi^0 - 1)}\left(3\xi^3 - \frac{9\xi^3}{(N - \xi^0)}\right)\left(1 + \frac{\xi^1 - 1}{N - \xi^0}\right) \\
-\frac{\xi^2}{N - \xi^0}E\left[M\left(\frac{3(\xi^3 + x^3)}{2(N - \xi^0 - 1)}\right)\right] - \mathbf{1}(\xi^1 \neq 0)\xi^1\left(E\left[M\left(\frac{3(\xi^3 + x^3)}{2(N - \xi^0 - 1)}\right)\right] - M\left(\frac{3\xi^3}{2(N - \xi^0)}\right)\right) \\
+\mathbf{1}(\xi^1 \neq 0)E\left[M\left(\frac{3(\xi^3 + x^3)}{2(N - \xi^0 - 1)}\right)\right]\left(1 + \frac{\xi^1 - 1}{N - \xi^0}\right).
\end{aligned}$$

Also recall that

$$\beta^2(\xi) = -2\frac{\xi^2}{N - \xi^0} + \mathbf{1}(\xi^1 = 0)M(\lambda) + \mathbf{1}(\xi^1 \neq 0)\lambda.$$

The vector field $\bar{b}(\bar{\boldsymbol{x}}(\xi))$ will be given by

$$\bar{b}^2(\bar{\boldsymbol{x}}(\xi)) = -2\frac{\xi^2 + \xi^1(M(\lambda) - \lambda)}{N - \xi^0} + \frac{\xi^2 + \xi^1(M(\lambda) - \lambda)}{N - \xi^0}\frac{3\xi^3}{2(N - \xi^0)} + \left(1 - \frac{\xi^2 + \xi^1(M(\lambda) - \lambda)}{N - \xi^0}\right)M(\lambda),$$

where $\lambda = \frac{3\xi^3}{2(N - \xi^0)}$.

3.4 The convergence

To prove convergence, we will check that the conditions of Theorem 1.3 are all satisfied for the coordinate process $\bar{\boldsymbol{x}}^2$ and then we will prove that the corrector is small in an appropriate sense, so as to be able to transfer the approximation to the initial coordinate process $\boldsymbol{x}^2$. We choose $t_0 = 1 - \varepsilon$ and $x_0^2 = 0$. We also have that $\frac{\xi^2}{N} \leq r$ and $\frac{\xi^3}{N} \leq r$ at all times. Finally we have that

$$P\left(\int_0^{T\wedge t_0} |\bar{\beta}^2(X_t) - \bar{b}^2(\bar{\boldsymbol{x}}(X_t))|dt \leq \delta\right) \to 1, \quad \text{as} \quad N \to \infty,$$

and

$$P\left(\chi(\boldsymbol{x}(X_t), Y_t) > \delta\right) \to 0, \quad \text{as} \quad N \to \infty$$

because of the Theorems 3.4.1 and 3.4.2 below. We can also compute the variance field α and by the same theorems we can deduce that $P(\Omega_2) \to 1$, as $N \to \infty$.

Theorem 3.4.1. *For the number of 1-clauses we have the following:* $\forall \gamma > 0$,

$$P\left(\sup_{t=1,\ldots,N} C_1(t) \geq \gamma N\right) \to 0, \quad \textit{as} \quad N \to \infty$$

Proof. $C_1(t)$ will jump at most $(1-\varepsilon)N$ times, so that's why the supremum above is taken over discrete time steps. Also, $C_1(t)$ can be coupled, so that it remains always below a Markov chain with the following transition rates:

$$q(y, y') = \begin{cases} Ne^{-\lambda}\frac{\lambda^j}{j!}, & \text{if} \quad y = 0, y' = j \\ Ne^{-\lambda}\frac{\lambda^j}{j!}, & \text{if} \quad y > 0, y' = y - 1 + j, \end{cases}$$

where $\lambda = 1 - \delta < 1$.

Let $t_1, t_2, \ldots, t_N$ be the successive times that $C_1(t)$ hits 0. Then the size of $C_1(t)$ in the interval (t_i, t_{i+1}) cannot exceed $(t_{i+1} - t_i)$, because it goes down by at most 1 at each step.

Also, these lengths are i.i.d. and

$$\begin{aligned} P(t_{i+1} - t_i \geq k) = P(t_1 \geq k) &= P(\forall l \leq k-1, \sum_{i=1}^{l} X_i \geq l) \\ \leq P(X_1 + \ldots + X_{k-1} \geq k-1) &\leq \exp\left(k-1\right)(1 - \lambda + \log(\lambda)), \end{aligned}$$

by the Markov inequality and optimizing, where $X_i \sim P(\lambda)$ and $1 - \lambda + \log(\lambda) < 0$.

$$P\left(\sup_{t=1,\ldots,N} C_1(t) \geq \gamma N\right) \leq P\Big(\sup C_1(t) \geq \gamma N \text{ in one of the intervals} \quad (t_i, t_{i+1})\Big) \leq Ne^{-bN},$$

for some $b > 0$. ∎

Theorem 3.4.2. *$M(\lambda) = E[P_1(\lambda) \wedge P_2(\lambda)]$, where $P_1(\lambda)$ and $P_2(\lambda)$ are two independent Poisson random variables, satisfies the following: $\forall \delta > 0$*

$$M(\lambda + \delta) - M(\lambda) \leq 2\delta.$$

Proof. Let Y_1, Y_2 be two independent Poisson random variables each with parameter λ. Let X_1, X_2 be two independent Poisson random variables each with parameter $\lambda + \delta$. Then we can construct them all in such a way that $X_1 = Y_1 + Z_1$ and $X_2 = Y_2 + Z_2$, where Z_1, Z_2 are independent Poisson with parameter δ each. So, we have

$$\begin{aligned} M(\lambda + \delta) - M(\lambda) = E[X_1 \wedge X_2 - Y_1 \wedge Y_2] = E[(Y_1 + Z_1) \wedge (Y_2 + Z_2) - Y_1 \wedge Y_2] \\ \leq E[(Y_1 + Z_1 + Z_2) \wedge (Y_2 + Z_1 + Z_2) - Y_1 \wedge Y_2] = E[Y_1 \wedge Y_2 + Z_1 + Z_2 - Y_1 \wedge Y_2] \\ = E[Z_1 + Z_2] = 2\delta. \end{aligned}$$

□

Theorem 3.4.3. *There exists a constant M such that*

$$P\left(\sum_{t=1}^{N} C_1(t) \leq MN\right) \to 1, \quad \textit{as} \quad N \to \infty$$

Proof. Following the same logic as before,

$$P\left(\sum_{t=1}^{N} C_1(t) \geq MN\right) \leq P\left(\sum_{t=1}^{N} (t_{i+1} - t_i)^2 \geq MN\right).$$

Now, write $S_N = \sum_{t=1}^{N} (t_{i+1} - t_i)^2$, so it is the sum of N i.i.d. random variables each with mean $E[t_1^2] < \infty$. By the strong law of large numbers

$$\frac{S_N}{N} \to E[t_1^2], \text{ as } N \to \infty.$$

So

$$P\left(\sum_{t=1}^{N} (t_{i+1} - t_i)^2 \geq MN\right) = P\left(\frac{S_N}{N} - E[t_1^2] \geq M - E[t_1^2]\right) \to 0, \text{ when } M > E[t_1^2].$$

□

So, we have proved the following theorem:

Theorem 3.4.4. *The number of 2-clauses can be approximated by the following differential equation:*

$$\begin{aligned} \frac{\mathrm{d}c_2(x)}{\mathrm{d}x} &= \frac{c_2(x)}{1-x}\left(\frac{3c_3(x)}{2(1-x)} - \frac{2c_2(x)}{1-x}\right) + \left(1 - \frac{c_2(x)}{1-x}\right)\left(B_q\left(\frac{3c_3(x)}{2(1-x)}\right) - \frac{2c_2(x)}{1-x}\right) \\ &= \frac{c_2(x)}{1-x}\left(\frac{3r(1-x)^2}{2} - \frac{2c_2(x)}{1-x}\right) + \left(1 - \frac{c_2(x)}{1-x}\right)\left(B_q\left(\frac{3r(1-x)^2}{2}\right) - \frac{2c_2(x)}{1-x}\right), \end{aligned}$$

which is the differential equation given in [4, p.174] with $\theta = 0$.

Finally, we are going to describe briefly how we can terminate the algorithm, because as we recall, we stopped after $(1-\varepsilon)N$ steps, which means at time $1-\varepsilon$ of the differential equation. We can then see that at that time the total number of clauses $C_2((1-\varepsilon)N)+C_3((1-\varepsilon)N)$ (there are no 1-clauses with positive probability) is bounded by $\frac{3}{4}\varepsilon N$, so by deleting one literal from each 3-clause we can reduce the problem to a 2-satisfiability problem, for which we know that $r_2 = 1$, hence we deduce that this last formula is satisfied w.h.p.

Chapter 4

BitTorrent-like peer-to-peer file-sharing networks

4.1 Introduction

In this chapter, we propose a stochastic model for a file-sharing peer-to-peer network which resembles the popular BitTorrent system: large files are split into chunks and a peer can download or swap from another peer only one chunk at a time. We will prove that the fluid limits of a scaled Markov model of this system are of the coagulation form, special cases of which are well-known epidemiological (SIR) models. Peer-to-peer (p2p) activity continues to represent a very significant fraction of overall Internet traffic, 44% by one recent account [8]. Under BitTorrent, peers join "swarms" (or "torrents") where each swarm corresponds to a specific data object (file). The process of finding the peers in a given swarm to connect to is typically facilitated through a centralised "tracker". Recently, a trackerless BitTorrent client has been introduced that uses distributed hashing for query resolution.

For file sharing, a peer typically uploads pieces ("chunks") of the file to other peers in the swarm while downloading his/her missing chunks from them. This chunk swapping constitutes a transaction-by-transaction incentive for peers to cooperate (i.e., trading rather than simply download) to disseminate data objects. Large files may be segmented into several hundred chunks, all of which the peers of the corresponding swarm must collect and in the process disseminate their own chunks before they can reconstitute the desired file and possibly leave the file's swarm.

In this chapter, we motivate a deterministic epidemiological model of file dissemination for peer-to-peer file-sharing networks that employ BitTorrent-like incentives, a generalisation of that given in [9]. Our model is different from those explored in [11, 16, 12] for BitTorrent, and we compute different quantities of interest. Though our model is significantly simpler than that of prior work, it is derived directly from an intuitive transaction-by-transaction Markov process modelling file-dissemination of the p2p network and its numerical solutions clearly demonstrate the effectiveness of the aforementioned incentives.

4.2 The stochastic model

We fix a set F (a file) which is partitioned into n (on the order of hundreds) pieces called chunks. Consider a large networked "swarm" of nodes called peers. Each peer possesses a certain (possibly empty) subset A of F. As time goes by, this peer interacts with other peers, the goal being to enlarge his set A until, eventually, the peer manages to collect all n chunks of F. The interaction between peers can either be a download or a swap; in both cases, chunks are being copied from peer to peer and are assumed never lost. Here we make the extra assumption that peers are impatient, which means that they can leave the network even withouth having obtained all the chunks of the file. In ..... a peer will stay in the network as long as he does not possess all chunks and after collecting everything, sooner or later a peer departs or switches off. By splitting the desired file into many chunks we give incentives to the peers to remain active in the swarm for long time during which other peers will take advantage of their possessions.

4.2.1 Possible interactions

We here describe how two peers, labelled A, B, interact. The following types of interactions are possible:

1. **Download:** Peer A downloads a chunk i from B. This is possible only if A is a strict subset of B. If $i \in B$ then, after the downloading A becomes $A' = A \cup \{i\}$ and but B remains B since it gains nothing from A. Denote this interaction as:

$$\boxed{(A \leftarrow B) \rightsquigarrow (A', B)}$$

   The symbol on the left is supposed to show the type of interaction and the labels before it, while the symbol on the right shows the labels after the interaction.

2. **Swap:** Peer A swaps with peer B. In other words, A gets a chunk j from B and B gets a chunk i from A. It is required that j is not an element of A and i not an element of B. We denote this interaction by

$$\boxed{(A \leftrightarrows B) \rightsquigarrow (A', B')}$$

   where $A' = A \cup \{j\}$, $B' = B \cup \{i\}$. We thus need $A \setminus B \neq \varnothing$ and $B \setminus A \neq \varnothing$.

3. **Full swap:** This is a special case of a swap that makes the two peers become identical after the interaction. For this to happen we need $|A \setminus B| = |B \setminus A| = 1$. After the interaction both peers attain the same labels: $A' = B' = A + (B \setminus A) = B + (A \setminus B)$. Thus, a full swap is denoted by:

$$\boxed{(A \leftrightarrows B) \rightsquigarrow (A', A')}$$

4.2.2 Notation

The set of all combinations of n chunks, which partition F, is denoted by $\mathcal{P}(F)$, where $|\mathcal{P}(F)| = 2^n$ and the empty set is included. We write $A \subset B$ (respectively, $A \subsetneq B$) when A is a subset (respectively, strict subset) of B. We (unconventionally) write

$$A \sqsubset A' \text{ when } A \subset A' \text{ and } |A' - A| = 1.$$

If $A \cap B = \varnothing$, we use $A + B$ instead of $A \cup B$; if $B = \{b\}$ is a singleton, we often write $A + b$ instead of $A + \{b\}$. If $A \subset B$ we use $B - A$ instead of $B \setminus A$. We say that

$$A \text{ relates to } B \text{ (and write } A \sim B \text{) when } A \subset B \text{ or } B \subset A;$$

if this is not the case, we write $A \not\sim B$. Note that $A \not\sim B$ if and only if two peers labelled A, B can swap chunks. The space of functions (vectors) from $\mathcal{P}(F)$ into $\mathbb{Z}_+$ is denoted by $\mathbb{Z}_+^{\mathcal{P}(F)}$. The stochastic model will take values in this space. The deterministic model will evolve in $\mathbb{R}_+^{\mathcal{P}(F)}$. We let $e_A \in \mathbb{Z}_+^{\mathcal{P}(F)}$ be the vector with coordinates

$$e_A^B := \mathbf{1}(A = B), \quad B \in \mathcal{P}(F).$$

For $x \in \mathbb{Z}_+^{\mathcal{P}(F)}$ or $\mathbb{R}_+^{\mathcal{P}(F)}$ we let $|x| := \sum_{A \in \mathcal{P}(F)} |x^A|$. If $\mathcal{A} \subset \mathcal{P}(F)$ then the $\mathcal{A}$-face $\mathbb{R}_+^{\mathcal{A}}$ of $\mathbb{R}_+^{\mathcal{P}(F)}$ is defined by $\mathbb{R}_+^{\mathcal{A}} := \{x \in \mathbb{R}_+^{\mathcal{P}(F)} : \sum_{A \in \mathcal{A}} x^A = 0, \prod_{B \notin \mathcal{A}} x^B > 0\}$.

4.2.3 Defining the rates of individual interactions

We follow the logic of stochastic modelling of chemical reactions or epidemics and assume that the chance of a particular interaction occurring in a short interval of time is proportional to the number of ways of selecting the peers needed for this interaction [13]. Accordingly, the interaction rates *must* be given by the formulae described below.

Consider first finding the rate of a download $A \leftarrow B$, where $A \subsetneq B$, when the state of the system is $x \in \mathbb{Z}_+^{\mathcal{P}(F)}$. There are x^A peers labelled A and x^B labelled B. We can choose them in $x^A x^B$ ways. Thus the rate of a download $A \leftarrow B$ that results into A getting *some* chunk from B should be proportional to $x^A x^B$. However, we are interested in the rate of the *specific* interaction $(A \leftarrow B) \rightsquigarrow (A', B)$, that turns A into a specific set A' differing from A by one single chunk ($A \sqsubset A'$); there are $|B - A|$ chunks that A can download from B; the chance that picking one of them is $1/|B - A|$. Thus we have:

$$(DR) \quad \begin{cases} \textit{the rate of the download } (A \leftarrow B) \rightsquigarrow (A', B) \textit{ equals } \beta x^A \dfrac{x^B}{|B - A|}, \\ \textit{as long as } A \sqsubset A' \subset B, \end{cases}$$

where $\beta > 0$.

Consider next a swap $A \leftrightarrows B$ and assume the state is x. Picking two peers labelled A and B (provided that $A \not\sim B$) from the population is done in $x^A x^B$ ways. Thus the rate of a swap $A \leftrightarrows B$ is proportional to $x^A x^B$. So if we *fix* two chunks $i \in A \setminus B, j \in B \setminus A$ and specify that $A' = A + j, B' = B + i$, then the chance of picking i from $A \setminus B$ and j from $B \setminus A$ is $1/|A \setminus B||B \setminus A|$. Thus,

$$(SR) \quad \begin{cases} \textit{the rate of the swap } (A \leftrightarrows B) \rightsquigarrow (A', B') \textit{ equals } \gamma \dfrac{x^A x^B}{|A \setminus B||B \setminus A|}, \\ \textit{a long as } A \sqsubset A', \quad B \sqsubset B', \quad A' - A \subset B, \quad B' - B \subset A, \end{cases}$$

where $\gamma > 0$.

4.2.4 Deriving the Markov chain rates

Having defined the rates of each individual interaction we can easily define rates $q(x, y)$ of a Markov chain in continuous time and state space $\mathbb{Z}_+^{\mathcal{P}(F)}$ as follows.

Define functions $\lambda_{A,A'}, \mu_{A,B} : \mathbb{R}^{\mathcal{P}(F)} \to \mathbb{R}$ by:

$$\lambda_{A,A'}(x) := \left[\beta x^A \sum_{C: C \supset A'} \frac{x^C}{|C - A|} \right] \mathbf{1}(A \sqsubset A') \tag{4.1a}$$

$$\mu_{A,B}(x) := \gamma \frac{x^A x^B}{|A \setminus B||B \setminus A|} \mathbf{1}(A \not\sim B). \tag{4.1b}$$

Consider also constants $\delta^A \geq 0$ and $\alpha^A \geq 0$ for $A \in \mathcal{P}(F)$, i.e., $\alpha \in \mathbb{R}_+^{\mathcal{P}(F)}$. The transition rates of the Markov chain are given by:

$$q(x, y) := \begin{cases} \lambda_{A,A'}(x), & \text{if } y = x - e_A + e_{A'} \\ \mu_{A,B}(x), & \text{if } \begin{cases} y = x - e_A - e_B + e_{A'} + e_{B'} \\ A \sqsubset A', B \sqsubset B', A' - A \subset B, B' - B \subset A, \end{cases} \\ \alpha^A & \text{if } y = x + e_A \\ \delta^A x^A & \text{if } y = x - e_A \\ 0, & \text{for any other value of } y \neq x, \end{cases} \tag{4.2}$$

where x ranges in $\mathbb{Z}_+^{\mathcal{P}(F)}$.

A little justification of the first two cases is needed: that $q(x, x - e_A - e_B + e_{A'} + e_{B'}) = \mu_{A,B}(x)$ is straightforward. It corresponds to a swap, which is only possible when $A \sqsubset A', B \sqsubset B', A' - A \subset B, B' - B \subset A$. The swap rate was defined by (SR). To see that $q(x, x - e_A + e_{A'}) = \lambda_{A,A'}(x)$ we observe that a peer labelled A can change its label to $A' \sqsupset A$ by downloading a chunk from some set C that contains A', so we sum the rates (DR) over all these possible individual interactions to obtain the first line in (4.2). We can think of having Poisson process of arrivals of new peers at rate $|\alpha|$, and that each arriving peer is labelled A with probability $\alpha^A/|\alpha|$. Peers can depart, even withouth having acquired the whole file, and it takes an exponentially distributed amount of time (with mean $1/\delta^A$) for a peer labelled A to depart. Thus, $q(x, x - e_A) = \delta x^A$. We shall let Q denote the generator of the chain, i.e. $\mathsf{Q}f(x) = \sum_y (f(y) - f(x))q(x, y)$, when f is an appropriate functional of the state space.

Definition 1 (`BITTORRENT`$[x_0, n, \alpha, \beta, \gamma, \delta]$). Given $x_0 \in \mathbb{Z}_+^{\mathcal{P}(F)}$ (initial configuration), $n = |F| \in \mathbb{N}$ (number of chunks), $\alpha \in \mathbb{R}_+^{\mathcal{P}(F)}$ (arrival rates), $\beta > 0$ (download rate), $\gamma \geq 0$ (swap rate), $\delta \geq 0$ (departure rate) we let `BITTORRENT`$[x_0, n, \alpha, \beta, \gamma, \delta]$ be a Markov chain $(X_t, t \geq 0)$ with transition rates (4.2) and $X_0 = x_0$. We say that the chain (network) is open if $\alpha^A > 0$ for at least one A and $\delta > 0$; it is closed if $\alpha^A = 0$ for all A; it is conservative if it is closed and $\delta = 0$; it is dissipative if it is closed and $\delta > 0$.

In a conservative network, we have $q(x, y) = 0$ if $|y| \neq |x|$ and so $|X_t| = |X_0|$ for all $t \geq 0$. Here, the actual state space is the simplex

$$\{x \in \mathbb{Z}_+^{\mathcal{P}(F)} : |x| = N\},$$

where $N = |X_0|$. It is easy to see that the state e_F is reachable from any other state, but all rates out of e_F are zero. Hence a conservative network has e_F as a single absorbing state.

In a dissipative network, we have $|X_t| \leq |X_0|$ for all $t \geq 0$. Here the state space is

$$\{x \in \mathbb{Z}_+^{\mathcal{P}(F)} : |x| \leq N\},$$

where $N = |X_0|$. It can be seen that a dissipative network has many absorbing points.

In an open network, there are no absorbing points. On the other hand, one may wonder if certain components can escape to infinity. This is not the case:

Lemma 4.2.1. *If $\alpha^F > 0$ then the open* `BITTORRENT`$[x, n, \beta, \gamma, \alpha, \delta]$ *is positive recurrent Markov chain.*

Proof. If $\alpha^F > 0$, $\delta > 0$ the Markov chain is irreducible. The remainder of the proof is based on the construction of a simple Lyapunov function:

$$V(x) := |x|,$$

where $|x|$ is the sum of all the components. For this Lyapunov function there exists a bounded set of states $K = \{x^A, A \subset F \text{ s.t. } \sum_A \delta^A x^A < \sum_A \alpha^A\}$ such that

$$\sup_{x \notin K} (\mathsf{Q}V)(x) < 0,$$

because

$$\sup_{x \notin K} (\mathsf{Q}V)(x) = \sum_A \alpha^A - \sum_A \delta^A x^A < 0, \text{ for } x \notin K,$$

since the sum of the components changes only if we have an arrival or a departure. □

4.2.5 Example: $n = 1$

Let us take the special case where the file consists of a single chunk ($n = 1$). The state here is $x = (x^\varnothing, x^1 := x^F)$. The rates are:

$$\begin{aligned}
q\big((x^\varnothing, x^1), (x^\varnothing + 1, x^1)\big) &= \alpha^\varnothing \\
q\big((x^\varnothing, x^1), (x^\varnothing, x^1 + 1)\big) &= \alpha^1 \\
q\big((x^\varnothing, x^1), (x^\varnothing - 1, x^1 + 1)\big) &= \beta x^\varnothing x^1 \\
q\big((x^\varnothing, x^1), (x^\varnothing, x^1 - 1)\big) &= \delta x^1 \\
q\big((x^\varnothing, x^1), (x^\varnothing - 1, x^1)\big) &= \delta^\varnothing x^\varnothing.
\end{aligned} \tag{4.3}$$

If $\alpha^{\varnothing} = \alpha^1 = \delta^{\varnothing} = 0$, this is the stochastic version of the classical (closed) Kermack-McKendrick (or susceptible-infective-removed (SIR)) model for a simple epidemic process [14]. Its absorbing points are states of the form $(x^{\varnothing}, 0)$. In epidemiological terminology, x^1 is the number of infected individuals, whereas $x^{\varnothing}$ is the number of susceptible ones. Contrary to the epidemiological interpretation, infection *is* desirable, for infection is tantamount to downloading the file.

4.3 Macroscopic description: fluid limit

Analysing the Markov chain in its original form is complicated. We thus resort to a first-order approximation by an ordinary differential equation (ODE).

Let $v(x)$ be the vector field on $\mathbb{R}_+^{\mathcal{P}(F)}$ with components $v^A(x)$ defined by

$$v^A(x) = \alpha^A - x^A\big(\beta\varphi_d^A(x) + \gamma\varphi_s^A(x)\big) + \beta \sum_{B:A\subset B} \frac{\psi_d^A(x)x^B}{1+|B\setminus A|} + \gamma \sum_{B:A\not\subset B} \frac{\psi_s^{A,B}(x)x^B}{1+|B\setminus A|} - \delta^A x^A, \quad (4.4)$$

where

$$\varphi_d^A(x) := \sum_{B\supset A} x^B, \quad \varphi_s^A(x) := \sum_{B\not\supset A} x^B$$
$$\psi_d^A(x) := \sum_{a\in A} x^{A-a}, \quad \psi_s^{A,B}(x) := \sum_{a\in A\cap B} x^{A-a} \quad (4.5)$$

Consider the differential equation

$$\dot{x} = v(x) \text{ with initial condition } x_0. \quad (4.6)$$

Consider the sequence of stochastic models $\texttt{BITTORRENT[}X_{N,0}, n, N\alpha, \frac{\beta}{N}, \frac{\gamma}{N}, \delta\texttt{]}$ for $N \in \mathbb{N}$ and let $X_{N,t}$ be the corresponding jump Markov chain.

Theorem 4.3.1. *There is a unique smooth (analytic) solution to (4.6), denoted by x_t for $t \geq 0$. Also, if there is an $x_0 \in \mathbb{R}_+^{\mathcal{P}(F)}$ such that $X_{N,0}/N \to x_0$ as $N \to \infty$, then for any $T, \varepsilon > 0$,*

$$\lim_{N\to\infty} P\big(\sup_{0\leq t\leq T} |N^{-1}X_{N,t} - x_t| > \varepsilon\big) = 0.$$

Proof. Let $\mathcal{N}$ be the set of vectors $-e_F$, e_A, $-e_A + e_{A'}$, $-e_A - e_B + e_{A'} + e_{B'}$, where $A, B \in \mathcal{P}(F)$ and $A \sqsubset A'$, $B \sqsubset B'$. From (4.2), we have that $q(x,y) = 0$ if $y - x \notin \mathcal{N}$. Introduce, for each $\zeta \in \mathcal{N}$, a unit rate Poisson process Φ_ζ on the real line, and assume that these Poisson processes are independent. Consider the Markov chain (X_t) for the $\texttt{BITTORRENT[}X_0, n, \alpha, \beta, \gamma, \delta\texttt{]}$. Its rates are of the form

$$q(x, x+\zeta) = Q_\zeta(x), \quad \zeta \in \mathcal{N}, \quad (4.7)$$

where $Q_\zeta(x)$ is a polynomial in 2^n variables of degree 2, and which can be read directly from (4.2); its coefficients depend on the parameters α, β, γ, δ. We can represent [13, ?] (X_t) as:

$$X_t = X_0 + \sum_{\zeta \in \mathcal{N}} \zeta \Phi_\zeta\big(\int_0^t Q_\zeta(X_s) ds \big).$$

Consider now the Markov chain $\frac{1}{N} X_{N,t}$ corresponding to to `BITTORRENT[`$X_{N,0}, n, (N\alpha^A), \beta/N, \gamma/N, \delta$`]`. The transition rates for $\frac{1}{N} X_{N,t}$ are

$$q(x/N, (x+\zeta)/N) = N Q_\zeta(x/N), \quad x \in \mathbb{Z}_+^{\mathcal{P}(F)}, \quad \zeta \in \mathcal{N},$$

and 0, otherwise. Here, $Q_\zeta(x)$ is the polynomial defined through (4.7) and (4.2) and we now assume that its variables range over the reals. Therefore, $\frac{1}{N} X_{N,t}$ can be represented as

$$\frac{1}{N} X_{N,t} = \frac{1}{N} X_{N,0} + \sum_{\zeta \in \mathcal{N}} \zeta \frac{1}{N} \Phi_\zeta \Big(N \int_0^t Q_\zeta(\frac{1}{N} X_{N,s}) ds \Big).$$

Define x_t by the (deterministic) integral equation

$$x_t = x_0 + \sum_{\zeta \in \mathcal{N}} \zeta \int_0^t Q_\zeta(x_s) ds \tag{4.8}$$

and assume that it is unique for all $t \geq 0$. Fix a time horizon $T > 0$ and let

$$\begin{aligned} B &:= \max_{t \leq T} |x_t|, \\ M_\zeta &:= \max_{|x| \leq B} |Q_\zeta(x)| \\ L_\zeta &:= \sup_{\substack{|x|,|y| \leq B \\ x \neq y}} \frac{|Q_\zeta(x) - Q_\zeta(y)|}{|x-y|} \\ \tau_N &:= \inf\{t > 0 : \ |X_{N,t}| > NB\}. \end{aligned}$$

We then have:

$$\begin{aligned} \Delta_{N,t} := \frac{X_{N,t}}{N} - x_t = \frac{X_{N,0}}{N} - x_0 + \sum_{\zeta \in \mathcal{N}} \zeta \Big[\frac{1}{N} \Phi_\zeta \Big(N \int_0^t Q_\zeta(X_{N,s}/N) ds \Big) - \int_0^t Q_\zeta(X_{N,s}/N) ds \Big] \\ + \sum_{\zeta \in \mathcal{N}} \zeta \int_0^t \big(Q_\zeta(X_{N,s}/N) - Q_\zeta(x_s) \big) ds \end{aligned}$$

Suppose that $t \leq T \wedge \tau_N$. Then, for all $s \leq t$,

$$|Q_\zeta(X_{N,s}/N) - Q_\zeta(x_s)| \leq L_\zeta |\Delta_{N,s}|.$$

So, if we let

$$\mathcal{E}_{N,t} := \frac{X_{N,0}}{N} - x_0 + \sum_{\zeta \in \mathcal{N}} \zeta \frac{1}{N} \Big[\Phi_\zeta \Big(N \int_0^t Q_\zeta(X_{N,s}/N) ds \Big) - N \int_0^t Q_\zeta(X_{N,s}/N) ds \Big],$$

we have, by the Gronwall-Bellman lemma, that

$$|\Delta_{N,t}| \leq |\mathcal{E}_{N,t}| \exp\big(t \sum_{\zeta\in\mathcal{N}} |\zeta| L_\zeta\big), \quad \text{if } t \leq T \wedge \tau_N.$$

Let

$$\Phi^*_\zeta(t) := \sup_{s\leq t} |\Phi_\zeta(s) - s|.$$

We recall that, as $N \to \infty$,

$$\frac{1}{N}\Phi^*_\zeta(Nt) \to 0 \ \text{ a.s.} \tag{4.9}$$

If $s \leq t \leq \tau_N$, we have $X_{N,s}/N \leq B$ (definition of τ_N) and so $Q_\zeta(X_{N,s}/N) \leq M_\zeta$, implying that

$$\sup_{t\leq T\wedge\tau_N} |\mathcal{E}_{N,t}| \leq \big|\frac{X_{N,0}}{N} - x_0\big| + \sum_{\zeta\in\mathcal{N}} |\zeta| \frac{1}{N}\Phi^*_\zeta(NM_\zeta T)$$

which converges to zero, a.s., due to (4.9). Since

$$\sup_{t\leq T\wedge\tau_N} |\Delta_{N,t}| \leq \sup_{t\leq T\wedge\tau_N} |\mathcal{E}_{N,t}| \exp\big(T \sum_{\zeta\in\mathcal{N}} |\zeta| L_\zeta\big),$$

we have

$$\sup_{t\leq T\wedge\tau_N} |\Delta_{N,t}| \to 0, \quad \text{a.s.}$$

Now observe that

$$\begin{aligned} P(\tau_N \leq T) &\leq P(\sup_{t\leq T\wedge\tau_N} |X_{N,t}| > NB) \\ &\leq P(\sup_{t\leq T\wedge\tau_N} |\Delta_{N,t}| + \sup_{t\leq T\wedge\tau_N} |x_t| > B) \to 0. \end{aligned}$$

So we have $\sup_{t\leq T} |\Delta_{N,t}| \to 0$ a.s.

To show that x_t, defined via (4.8), satisfies the ODE $\dot{x} = v(x)$ with v given by (4.4) is a matter of straightforward (but tedious) algebra, see Section 4.5.

Uniqueness and analyticity of the solution of the ODE is immediate from the form of the vector field (its components are polynomials of degree 2 and hence locally Lipschitzian).

To show that the trajectories do not explode, we consider the function

$$V(x) := \sum_A x^A.$$

It is a matter of algebra to check that

$$\langle \nabla V(x), v(x)\rangle = \sum_A v^A(x) = \sum_A \alpha^A - \sum_A \delta^A x^A$$

which (since $\delta^A > 0$, for some $A \subset F$) is negative and bounded away from zero for x outside a bounded set of $\mathbb{R}_+^{\mathcal{P}(F)}$ containing the origin. We then apply the Lyapunov criterion for ODEs to conclude that x_t is defined for all $t \geq 0$ and this justifies the fact that we could choose an arbitrary time horizon T earlier in the proof. $\square$

Comment: The quantities defined in (4.5), have physical meanings as follows:

$$\varphi_d^A(x) := \sum_{B \supset A} x^B = \text{no. of peers from which an } A\text{-peer can download},$$
$$\varphi_s^A(x) := \sum_{B \not\sim A} x^B = \text{no. of peers an } A\text{-peer can swap with},$$
$$\psi_d^A(x) := \sum_{a \in A} x^{A-a} = \text{no. of peers which can assume label } A \text{ after a download},$$
$$\psi_s^{A,B}(x) := \sum_{a \in A \cap B} x^{A-a} = \text{no. of peers which can assume label } A \text{ after a } B\text{-peer swap}.$$

It is helpful to keep these in mind because they aid in writing down the various parts of $v(x)$, again, see Section 4.5.

4.4 Performance analysis in presence of BitTorrent incentives

4.4.1 An example

We address the following question: When is it advantageous to split a file into chunks? In other words, assuming we fix certain system parameters (e.g., arrival rates), will peers acquire the file faster if the file is split into chunks? We attempt here to answer the question in a simple case only by using the deterministic approximation. Let λ be the total peer arrival rate. Let β be the download rate. Assume that only $\varnothing$ peers arrive exogenously. In the absence of BitTorrent incentives, we have the single-chunk case

$$\dot{x}^{\varnothing} = \lambda - \beta x^{\varnothing} x^1$$
$$\dot{x}^1 = \beta x^{\varnothing} x^1 - \delta x^1.$$

The globally attracting stable equilibrium is given by

$$x^* = (\delta/\beta,\ \lambda/\delta).$$

Consider splitting into $n = 2$ chunks. Let $\widetilde{x}$ be the state of the system. Suppose that the new parameters are $\widetilde{\lambda} = \lambda$, $\widetilde{\delta} = \delta$, $\widetilde{\beta}$, $\widetilde{\gamma}$. Then

$$\dot{\widetilde{x}}^{\varnothing} = \lambda - \widetilde{\beta}\widetilde{x}^{\varnothing}(\widetilde{x}^1 + \widetilde{x}^2 + \widetilde{x}^{12})$$
$$\dot{\widetilde{x}}^1 = -\widetilde{x}^1(\widetilde{\beta}\widetilde{x}^{12} + \widetilde{\gamma}\widetilde{x}^2) + \widetilde{\beta}\widetilde{x}^{\varnothing}(\widetilde{x}^1 + \tfrac{1}{2}\widetilde{x}^{12})$$
$$\dot{\widetilde{x}}^2 = -\widetilde{x}^2(\widetilde{\beta}\widetilde{x}^{12} + \widetilde{\gamma}\widetilde{x}^1) + \widetilde{\beta}\widetilde{x}^{\varnothing}(\widetilde{x}^2 + \tfrac{1}{2}\widetilde{x}^{12})$$
$$\dot{\widetilde{x}}^{12} = \widetilde{\beta}(\widetilde{x}^1 + \widetilde{x}^2)\widetilde{x}^{12} + 2\widetilde{\gamma}\widetilde{x}^1\widetilde{x}^2 - \delta\widetilde{x}^{12}.$$

The new equilibrium is easily found to be

$$\widetilde{x}^* = \left(\frac{\delta}{\widetilde{\beta}}\Big(\frac{\delta}{\lambda}u + 1\Big)^{-1},\ \frac{u}{2},\ \frac{u}{2},\ \frac{\lambda}{\delta}\right),$$

where u is the positive number which solves

$$q(u) = 0,$$

and

$$q(u) := u^2 + \frac{2\widetilde{\beta}\lambda}{\widetilde{\gamma}\delta}u - \frac{2\lambda}{\widetilde{\gamma}}. \tag{4.10}$$

To see this, set the vector field equal to zero and solve for $\widetilde{x}^A$, $A \in \{\varnothing, 1, 2, 12 := F\}$. The simplest way to obtain the solution is by first adding the equations; this gives

$$\lambda - \delta\widetilde{x}^{12} = 0,$$

whence $\widetilde{x}^{12} = \lambda/\delta$. Then add the middle two equations after setting $u = \widetilde{x}^1 + \widetilde{x}^2$:

$$-\widetilde{\beta}u\widetilde{x}^{12} - 2\widetilde{\gamma}\widetilde{x}^1\widetilde{x}^2 + \widetilde{\beta}x^{\varnothing}(u + \widetilde{x}^{12}) = 0.$$

Replace $\widetilde{x}^{12}$ by λ/δ and observe that, due to symmetry, $\widetilde{x}^1 = \widetilde{x}^2 = u/2$. This gives the quadratic equation $q(u) = 0$ with q defined by (4.10). Finally, the first equation becomes

$$\lambda - \widetilde{\beta}\widetilde{x}^{\varnothing}(u + \lambda/\delta) = 0,$$

which is solved for $\widetilde{x}^{\varnothing}$ giving:

$$\widetilde{x}^{*\varnothing} = \frac{\lambda}{\widetilde{\beta}}\left(u + \frac{\lambda}{\delta}\right)^{-1} = \frac{\delta}{\widetilde{\beta}}\left(\frac{\delta}{\lambda}u + 1\right)^{-1} < \frac{\delta}{\widetilde{\beta}}.$$

Thus:

Corollary 4.4.1. *If* $\widetilde{\beta} \geq \beta$ *then*

$$\widetilde{x}^{*\varnothing} < x^{*\varnothing}.$$

So, by introducing splitting into 2 chunks, we have fewer peers who have no parts of the file at all. Using Little's theorem (see below), this can be translated into smaller waiting time from the time a peer arrives until he gets his first chunk.

Suppose now we are interested in determining how long it will take for a newly arrived peer to acquire the full file. On the average, a peer spends time equal to $\lambda^{-1}|x^*|$ before it exits the system. During last part of his sojourn interval (which is a random variable with mean $1/\delta$), the peer possess the full file. It thus takes on the average $\lambda^{-1}|x^*| - \delta^{-1}$ for a peer to acquire the full file. Since we assume that $\widetilde{\lambda} = \lambda$, $\widetilde{\delta} = \delta$, it suffices to show that

$$|x^*| > |\widetilde{x}^*|.$$

But

$$\begin{aligned}
|x^*| - |\widetilde{x}^*| &= \left[\frac{\delta}{\beta} + \frac{\lambda}{\delta}\right] - \left[\frac{\delta}{\widetilde{\beta}}\left(\frac{\delta}{\lambda}u + 1\right)^{-1} + u + \frac{\lambda}{\delta}\right] \\
&= \frac{\delta}{\beta} - \frac{\delta}{\widetilde{\beta}}\left(\frac{\delta}{\lambda}u + 1\right)^{-1} - u - \frac{\lambda}{\delta} \\
&= \left(\frac{\delta}{\lambda}u + 1\right)\left[\left(\frac{\delta}{\beta} - \frac{\delta}{\widetilde{\beta}}\right) + \left(\frac{\delta^2}{\beta\lambda} - 1\right)u - \frac{\delta}{\lambda}u^2\right]
\end{aligned}$$

recalling $u > 0$ solves $q(u) = 0$. So, $|x^*| - |\widetilde{x}^*| > 0$ if and only if

$$0 > \widetilde{q}(u) := u^2 - (\frac{\delta}{\beta} - \frac{\lambda}{\delta})u - (\frac{\lambda}{\beta} - \frac{\lambda}{\widetilde{\beta}}). \tag{4.11}$$

Define $\widetilde{u}$ as the unique positive number which satisfies

$$\widetilde{q}(\widetilde{u}) = 0.$$

Corollary 4.4.2. *If $\widetilde{\beta} \geq \beta$, a necessary and sufficient condition for $|x^*| > |\widetilde{x}^*|$ is $u < \widetilde{u}$.*

4.4.2 Some theorems

To justify the use of deterministic approximation for estimating performance measures, and, specifically, the use of mean values, we need to show that as $N \to \infty$, we can approximate stationary averages in the original stochastic network by equilibria of the resulting ODE.

From now on we will assume that $\delta^A > 0$, for all $A \subset F$, which means that peers can depart at any instant of time, without having downloaded the whole file. So, in this case the sum of the components $\sum_A X^A$ can be bounded by an $M/M/\infty$ queue with arrival rate equal to the sum of the arrival rates for each A, i.e. $\alpha = \sum_A \alpha^A$, and departure rate equal to the minimum of all the $\delta^A, A \subset F$, which we will denote by δ.

It is easy to show that the convergence to the ODE limit can be traslated into convergence of the means, using a uniform integrability argument (the process can be bounded by an $M/M/\infty$ queue as stated above). Namely,

$$\frac{1}{N} E X_{N,t} \xrightarrow[N\to\infty]{} x_t \xrightarrow[t\to\infty]{} x^*,$$

where the second limit concerns the behaviour of the ODE alone, which we conjecture that exists and that is asymptotically stable, because of the form of the differential equation. On the other hand, if we fix N and look at the asymptotic behaviour of the process $\frac{1}{N} X_{N,t}$ as $t \to \infty$, we have

$$\frac{1}{N} E X_{N,t} \xrightarrow[t\to\infty]{} \frac{1}{N} E \widetilde{X}_N,$$

where the law of $\widetilde{X}_N$ is the stationary distribution of the chain $(\frac{1}{N} X_{N,t})_{t\geq 0}$, which exists since the chain is positive recurrent as proved in Lemma 4.2.1. Now we are going to prove that we can also interchange the two limits above, namely that $\frac{1}{N} E\widetilde{X}_N \to x^*$, as $N \to \infty$. We will prove that in the same way as in [15], but for the sake of completeness we will analyze the technique here too.
We will follow three steps in the proof:

1. Let ν^N be a sequence of initial distributions that converges weakly to a probability measure ν, then we will prove that $\mathcal{L}\left(\frac{1}{N} X_{N,\cdot}, \nu^N\right) \to \mathcal{L}\left(x_\cdot, \nu\right)$, as $N \to \infty$.

2. If ν is any probability measure on $\mathbb{R}^{2^n}$, then $\mathcal{L}\left(x_t, \nu\right) \to \delta_{x^*}$ as $t \to \infty$, where δ_{x^*} is the Dirac measure on x^*.

3. The sequence $\left(\frac{1}{N}\widetilde{X}_N, N \geq 1\right)$ is tight.

The proof of the interchange of the limits now goes as follows: Since the sequence $\left(\frac{1}{N}\widetilde{X}_N, N \geq 1\right)$ is tight, it has a convergent subsequence, say $\mathcal{L}\left(\frac{1}{N_k}\widetilde{X}_{N_k}, k \geq 1\right)$ converging weakly to a measure ν. Then, because of 1, $\mathcal{L}\left(\frac{1}{N_k}X_{N_k,t}, \mathcal{L}\left(\frac{1}{N_k}\widetilde{X}_{N_k}\right)\right) \to \mathcal{L}\left(x_t, \nu\right)$ as $k \to \infty$. But, step 2 above gives that $\mathcal{L}\left(x_t, \nu\right) \to \delta_{x^*}$ as $t \to \infty$. Also, we have that $\mathcal{L}\left(\frac{1}{N_k}X_{N_k,t}, \mathcal{L}\left(\frac{1}{N_k}\widetilde{X}_{N_k}\right)\right)$ is a stationary Markov process, so by taking the limit we get that $\mathcal{L}\left(x_t, \nu\right)$ is also stationary, hence we deduce that $\nu = \delta_{x^*}$. So, we have proved that any convergent subsequence of $\mathcal{L}\left(\frac{1}{N}\widetilde{X}_N, N \geq 1\right)$ converges to δ_{x^*}, so the whole sequence converges to that measure.

Theorem 4.4.1. *Let ν^N be a sequence of probability measures converging weakly to a measure ν. Then*

$$\mathcal{L}\left(\left(\frac{1}{N}X_{N,t}\right)_t, \nu^N\right) \to \mathcal{L}\left((x_t)_t, \nu\right), \text{ as } N \to \infty.$$

Proof. Let f be a function from $D([0,T], \mathbb{R}^{2^n})$, which is continuous and bounded. It suffices to prove that

$$E\left|f\left(\left(\frac{1}{N}X_{N,t}^{X_{N,0}}\right)_{t\leq T}\right) - f((x_t^{\nu})_{t\leq T})\right| \to 0, \text{ as } N \to \infty,$$

where $\frac{1}{N}X_{N,t}^{X_{N,0}}$ means that the Markov chain is started from the initial state $X_{N,0}$ distributed as ν^N and equivalently for the differential equation that the initial condition is distributed according to measure ν. Now we have

$$\begin{aligned}
&E\left|f\left(\left(\frac{1}{N}X_{N,t}^{X_{N,0}}\right)_{t\leq T}\right) - f((x_t^{\nu})_{t\leq T})\right| \\
&= E\left|f\left(\left(\frac{1}{N}X_{N,t}^{X_{N,0}}\right)_{t\leq T}\right) - f((x_t^{X_{N,0}})_{t\leq T}) + f((x_t^{X_{N,0}})_{t\leq T}) - f((x_t^{\nu})_{t\leq T})\right| \\
&\leq E\left|f\left(\left(\frac{1}{N}X_{N,t}^{X_{N,0}}\right)_{t\leq T}\right) - f((x_t^{X_{N,0}})_{t\leq T})\right| + E\left|f((x_t^{X_{N,0}})_{t\leq T}) - f((x_t^{\nu})_{t\leq T})\right| \qquad (4.12)
\end{aligned}$$

Since the sequence of measures ν^N converges weakly, it follows that it is tight, hence there exists a compact set A such that $\nu^N(A^c) < \varepsilon$ for all N and also $\nu(A^c) < \varepsilon$. The first term of (4.12) can be written as

$$\begin{aligned}
&E\left|f\left(\left(\frac{1}{N}X_{N,t}^{X_{N,0}}\right)_{t\leq T}\right) - f((x_t^{X_{N,0}})_{t\leq T})\right| \\
&= E\left|f\left(\left(\frac{1}{N}X_{N,t}^{X_{N,0}}\right)_{t\leq T}\right) - f((x_t^{X_{N,0}})_{t\leq T})\right| \mathbf{1}(X_{N,0} \in A) \\
&+ E\left|f\left(\left(\frac{1}{N}X_{N,t}^{X_{N,0}}\right)_{t\leq T}\right) - f((x_t^{X_{N,0}})_{t\leq T})\right| \mathbf{1}(X_{N,0} \notin A) \qquad (4.13)
\end{aligned}$$

Now, the second summand of (4.13) is bounded from above by $M\nu^N(A^c)$, where $M/2$ is a bound for the function f, which can be made as small as we like by choosing A suitably. The first summand of (4.13) can be written as follows:

$$\begin{aligned} E&\left| f\left(\left(\frac{1}{N}X_{N,t}^{X_{N,0}}\right)_{t\le T}\right) - f((x_t^{X_{N,0}})_{t\le T})\right| \mathbf{1}(X_{N,0}\in A) \\ &= \int_A E\left| f\left(\left(\frac{1}{N}X_{N,t}^{x}\right)_{t\le T}\right) - f((x_t^{x})_{t\le T})\right| \nu^N(\mathrm{d}x), \end{aligned} \tag{4.14}$$

where the quantity under the integral sign can be made as small as we like for N large enough because of Theorem 4.3.1 and then the integral gives $\delta\nu^N(A)$, but since A is a compact set, it follows that $\overline{\lim}\,\nu^N(A) \le \nu(A)$.

So it remains to be proved that the second term of (4.12), $E\left|f((x_t^{X_{N,0}})_{t\le T}) - f((x_t^{\nu})_{t\le T})\right|$, can also be made sufficiently small for N large enough. To prove that, we will firstly prove that the sequence $(x_t^{\nu^N})_{t\le T}$ is tight and then that the only possible limit is $(x_t^{\nu})_{t\le T}$. The tightness part follows easily from the continuity of the solution to the differential equation, irrespective of the initial condition.

For the second step, let $a_1,\dots,a_k \in \mathbb{R}^{2^n}$ and $0\le t_1,\dots,t_k\le T$. Then we will prove that

$$E\left[\exp\left(i\sum_{l=1}^{k}\langle a_l, x_{t_l}^{\nu^N}\rangle\right)\right] \to E\left[\exp\left(i\sum_{l=1}^{k}\langle a_l, x_{t_l}^{\nu}\rangle\right)\right], \text{ as } N\to\infty. \tag{4.15}$$

We are using again the same set A as above and we obtain:

$$\begin{aligned} &E\left[\exp\left(i\sum_{l=1}^{k}\langle a_l, x_{t_l}^{\nu^N}\rangle\right)\right] \\ &= E\left[\exp\left(i\sum_{l=1}^{k}\langle a_l, x_{t_l}^{\nu^N}\rangle\right)\mathbf{1}(X_{N,0}\in A)\right] + E\left[\exp\left(i\sum_{l=1}^{k}\langle a_l, x_{t_l}^{\nu^N}\rangle\right)\mathbf{1}(X_{N,0}\notin A)\right]. \end{aligned}$$

The second term again is bounded by $\nu^N(A^c)$ which is small and the first term converges to

$$E\left[\exp\left(i\sum_{l=1}^{k}\langle a_l, x_{t_l}^{\nu}\rangle\right)\mathbf{1}(x_0\in A)\right],$$

because of the weak convergence.

So, we have proved that for any finite k and for all t_l

$$\mathcal{L}(x_{t_1}^{\nu^N},\dots,x_{t_k}^{\nu^N}) \to \mathcal{L}(x_{t_1}^{\nu},\dots,x_{t_k}^{\nu}),$$

hence

$$E\left|f((x_t^{X_{N,0}})_{t\le T}) - f((x_t^{\nu})_{t\le T})\right| \to 0, \text{ as } N\to\infty.$$

□

Theorem 4.4.2. *If ν is any probability measure on $\mathbb{R}^{2^n}$, then $\mathcal{L}(x_t, \nu) \to \delta_{x^*}$ as $t \to \infty$, where δ_{x^*} is the Dirac measure on x^*.*

Proof. Let f be a continuous and bounded function. Then we have

$$E_\nu[f(x_t)] = \int_{\mathbb{R}^{2^n}} E_x[f(x_t)]\nu(\mathrm{d}x).$$

But $E_x[f(x_t)] \to f(x^*)$, since we have conjectured that the point x^* is asymptotically stable, so by dominated convergence, we get that

$$E_\nu[f(x_t)] \to f(x^*).$$

□

Theorem 4.4.3. *The sequence $\left(\frac{1}{N}\widetilde{X}_N, N \geq 1\right)$ is tight.*

Proof. We have

$$P\left(\left|\left|\frac{1}{N}\widetilde{X}_N\right|\right| > K\right) = \lim_{t\to\infty} P\left(\left|\left|\frac{1}{N}X_{N,t}\right|\right| > K\right),$$

where $||.||$ is the sum of the components. As we have already observed the sum of the components can be bounded by an $M/M/\infty$ queue with arrival rate α and departure rate δ, and now the proof of this theorem follows in the same way as Theorem 6.3 of [15]. □

Thus, we have proved the three steps, hence obtaining that the law of the stationary process converges weakly to the Dirac measure at x^*. Since the stationary process converges in distribution to a constant, we deduce that it converges to that constant in probability too. Now we can use a uniform integrability argument to show that also the expectation converges.

Now we are going to explain the use of $|x^*|$ as a measure of the sojourn time in the system of a peer, by first using the approximation outlined above and then appealing to Little's law. This is as follows.

Consider an open $\texttt{BITTORRENT}[X_0, n, \alpha, \beta, \gamma, \delta]$, i.e. $|\alpha| > 0$, $\delta > 0$. We know that the Markov chain (X_t) is positive recurrent and has thus a unique stationary distribution. It makes sense to assess the performance of the network by looking at steady-state performance measures, such as the mean time it takes for an $\varnothing$-peer to become an F-peer (a seed). Consider then the process $(\widetilde{X}_t, t \in \mathbb{R})$ defined to be a stationary Markov process with time index $\mathbb{R}$ and transition rates as those of (X_t). The law of the process $(\widetilde{X}_t, t \in \mathbb{R})$ is unique. Let $T_k^A, k \in \mathbb{Z}$ be the times at which A-peers arrive (and, say, $T_0^A \leq 0 < T_1^A$, by convention). These are the points of a stationary Poisson process in $\mathbb{R}$ with rate α^A. Let W_k^A be the sojourn time (the time it takes to acquire the whole file) in the system of a peer arriving at time T_k^A. So the time W_k^A is the sum of the times it takes for the peer to become a seed plus the time that the peer hangs out in the system after becoming a seed (the latter is an exponential time with mean $1/\delta$). Clearly then, for all $t \in \mathbb{R}$,

$$\sum_{B\supset A} \widetilde{X}_t^B = \sum_{k\in\mathbb{Z}} \mathbf{1}(T_k^A \leq t < T_k^A + W_k^A).$$

Using Campbell's formula, we obtain

$$\sum_{B \supset A} E\widetilde{X}_0^B = \alpha^A E^A W_0^A, \tag{4.16}$$

where E^A is expectation with respect to P^A–the Palm probability of P with respect to the point process $(T_k^A, k \in \mathbb{Z})$.

In particular, with $A = \varnothing$, and $\lambda = \alpha^\varnothing$, we have that

$$E^\varnothing W_0^\varnothing = \frac{1}{\lambda} E|\widetilde{X}_0|,$$

which can be read as: the mean sojourn time of a $\varnothing$-peer is, in steady state, equal to the mean number of peers in the system divided by the rate of arrivals of $\varnothing$-peers. If N is a parameter of the process as in Theorem 4.3.1 then, λ being proportional to N, we have that the right hand side converges to something that is proportional to $|x^*|$, as required.

4.5 Drift calculation

We consider the set of vectors

$$\begin{aligned}\mathcal{N} = \{-e_F\} &\cup \{e_A : A \subset F\} \cup \{-e_A + e_{A'} : A \subset A' \subset F\} \\ &\cup \{-e_A - e_B + e_{A'} + e_{B'} : A \subset A' \subset F, B \subset B' \subset F, A' - A \subset B, B' - B \subset A\}.\end{aligned}$$

For each $\zeta \in \mathcal{N}$ we define a polynomial $Q_\zeta(x)$, by comparing (4.7) and (4.2):

$$\begin{aligned} Q_{e_A}(x) &:= \alpha^A \\ Q_{-e_F}(x) &:= \delta x^F \\ Q_{-e_A+e_{A'}}(x) &:= \lambda_{A,A'}(x) \\ Q_{-e_A-e_B+e_{A'}+e_{B'}}(x) &:= \mu_{A,B}(x)\delta_{A,A',B,B'}, \end{aligned} \tag{4.17}$$

where $\lambda_{A,A'}(x)$, $\mu_{A,B}(x)$ are given by (4.1a), (4.1b), respectively, and

$$\delta_{A,A',B,B'} := \mathbf{1}(A \sqsubset A', A' - A \subset B, B \sqsubset B', B' - B \subset A).$$

The variable x ranges in $\mathbb{Z}_+^{\mathcal{P}(F)}$ or in $\mathbb{R}_+^{\mathcal{P}(F)}$. The algebra is the same in both cases. Define the drift vector field by $\sum_y (y - x) q(x, y)$. Comparing (4.17) and (4.2) we have

$$\sum_y (y - x) q(x, y) = \sum_{\zeta \in \mathcal{N}} \zeta Q_\zeta(x).$$

The latter sum appears in (4.8), in the course of the proof of Theorem 4.3.1. We shall verify that

$$u(x) := \sum_{\zeta \in \mathcal{N}} \zeta Q_\zeta(x) = v(x),$$

where $v(x)$ is defined by (4.4).

Consider the terms in the summation $u(x) = \sum_{\zeta\in\mathcal{N}} \zeta Q_\zeta(x)$ involving $\zeta = -e_A - e_B + e_{A'} + e_{B'}$. Notice that swapping A with B or A' with B' will not change the value of $x - e_A, x - e_B + e_{A'} + e_B$, so we need to make sure to take into account this change only once in the summation. If we *simultaneously* swap A with B *and* A' with B' then neither $x - e_A, x - e_B + e_{A'} + e_B$ nor the value of $Q_{-e_A-e_B+e_{A'}+e_{B'}}(x) = \mu_{A,B}(x)\delta_{A,A',B,B'}$ will change because, clearly,

$$\mu_{A,B}(x)\delta_{A,A',B,B'} = \mu_{B,A}(x)\delta_{B,B',A,A'},$$

as readily follows from (4.1a) and (4.1b). We now see that to swap A with B *without* swapping A' with B' is impossible (unless $A' = B'$). Indeed, it is an easy exercise that

$$\delta_{A,B,A',B'} = \delta_{B,A,A',B'} \;\Rightarrow\; A' = B'.$$

Taking into account this, we write

$$u(x) = \alpha^A e_A - \delta x^F e_F + \sum_{A,A'} (-e_A + e_{A'})\lambda_{A,A'}(x) + \frac{1}{2}\sum_{A,B,A',B'} (-e_A - e_B + e_{A'} + e_{B'})\mu_{A,B}(x)\delta_{A,B,A',B'}, \quad (4.18)$$

where the 1/2 appears because each term must be counted exactly once. The variables A, A', B, B' in both summations are free to move over $\mathcal{P}_n(F)$ (but notice that restrictions have effectively been pushed in the definitions of $\lambda_{A,A'}$, $\mu_{A,B}$, and $\delta_{A,B,A',B'}$).

Since

$$\sum_{A,B,A',B'} e_A \mu_{A,B}(x)\delta_{A,B,A',B'} = \sum_{A,B,A',B'} e_B \mu_{A,B}(x)\delta_{A,B,A',B'},$$
$$\sum_{A,B,A',B'} e_{A'} \mu_{A,B}(x)\delta_{A,B,A',B'} = \sum_{A,B,A',B'} e_{B'} \mu_{A,B}(x)\delta_{A,B,A',B'},$$

we have

$$v(x) = \alpha^A e_A - \delta x^F e_F - \sum_{A,A'} e_A \lambda_{A,A'}(x) + \sum_{A,A'} e_{A'}\lambda_{A,A'}(x) - \sum_{A,B,A',B'} e_A \mu_{A,B}(x)\delta_{A,B,A',B'} + \sum_{A,B,A',B'} e_{A'}\mu_{A,B}(x)\delta_{A,B,A',B'}.$$

Call the four sums appearing in this display as $u_{\rm I}(x), u_{\rm II}(x), u_{\rm III}(x), u_{\rm IV}(x)$, in this order. We use the definitions (4.1a), (4.1b) of $\lambda_{A,A'}$, $\mu_{A,B}$ and find the components of the vectors $u_{\rm I}, \ldots, u_{\rm IV}$ by hitting each one with a unit vector e_G, i.e. by taking the inner products $v_{\rm I}^G = \langle e_G, u_{\rm I}\rangle, \ldots, v_{\rm IV}^G = \langle e_G, u_{\rm IV}\rangle$. We have:

$$u_{\rm I}^G(x) = -\sum_{A'} \lambda_{G,A'}(x) = -\sum_{A'} \beta x^G \sum_{B:B\supset A'} \frac{x^B}{|B-G|}\mathbf{1}(G \sqsubset A') = -\beta x^G \sum_B \frac{x^B}{|B-G|} \sum_{A'} \mathbf{1}(G \sqsubset A' \subset B) = -\beta x^G \sum_{B\supset G} x^B, \quad (4.19)$$

where, in deriving the last equality we just observed that the number of sets A' that contain one more element than G and are contained in B is equal to $|B - G|$, as long as $G \subset B$:

$$\sum_{A'} \mathbf{1}(G \sqsubset A' \subset B) = |B - G| \; \mathbf{1}(G \subset B).$$

Next,

$$\begin{aligned} u_{\mathrm{II}}^G(x) = \sum_A \lambda_{A,G}(x) &= \sum_A \beta x^A \sum_{B \supset G} \frac{x^B}{|B - A|} \; \mathbf{1}(A \sqsubset G) \\ &= \beta \sum_{B \supset G} x^B \sum_A \frac{x^A}{|B - A|} \; \mathbf{1}(A \sqsubset G) \end{aligned}$$

Notice that, in the last summation, G contains exactly one more element than A and is strictly contained in B, so $|B - A| = |B - G| + 1$. Hence

$$u_{\mathrm{II}}^G(x) = \beta \sum_{B:B \supset G} \frac{x^B}{|B - G| + 1} \sum_A x^A \mathbf{1}(A \sqsubset G) = \beta \sum_{B:B \supset G} \frac{x^B}{|B - G| + 1} \sum_{g \in G} x^{G-g}. \quad (4.20)$$

For $u_{\mathrm{III}}(x)$, we have:

$$\begin{aligned} u_{\mathrm{III}}^G(x) &= - \sum_{B,A',B'} \mu_{G,B}(x) \delta_{G,A',B,B'} \\ &= -\gamma \sum_B \frac{x^G x^B}{|G \setminus B||B \setminus G|} \cdot \sum_{A'} \mathbf{1}(G \sqsubset A', A' - G \subset B) \cdot \sum_{B'} \mathbf{1}(B \sqsubset B', B' - B \subset G) \\ &= -\gamma \sum_B \frac{x^G x^B}{|G \setminus B||B \setminus G|} \cdot |B \setminus G| \; \mathbf{1}(B \setminus G \neq \varnothing) \cdot |G \setminus B| \; \mathbf{1}(G \setminus B \neq \varnothing) \\ &= -\gamma x^G \sum_{B \not\sim G} x^B. \quad (4.21) \end{aligned}$$

As for the last term, we have:

$$\begin{aligned} u_{\mathrm{IV}}^G(x) &= \sum_{A,B,B'} \mu_{A,B}(x) \delta_{A,G,B,B'} \\ &= \gamma \sum_B \sum_A \frac{x^A x^B}{|A \setminus B||B \setminus A|} \mathbf{1}(A \sqsubset G, G - A \subset B) \sum_{B'} \mathbf{1}(B \sqsubset B', B' - B \subset A) \\ &= \gamma \sum_B \sum_A \frac{x^A x^B}{|B \setminus A|} \mathbf{1}(A \sqsubset G, G - A \subset B) \; \mathbf{1}(A \setminus B \neq \varnothing) \\ &= \gamma \sum_B \frac{x^B}{|B \setminus G| + 1} \sum_A x^A \; \mathbf{1}(A \sqsubset G, G - A \subset B) \; \mathbf{1}(G \not\subset B) \\ &= \gamma \sum_B \frac{x^B}{|B \setminus G| + 1} \sum_{g \in G \cap B} x^{G-g} \; \mathbf{1}(G \not\subset B) \quad (4.22) \end{aligned}$$

Adding (4.19) and (4.21) we obtain the first part of (4.4), while (4.21) and (4.22) give the second part.

4.6 Results of a simulation study

Here we present some numerical simulations that were generated by Youngmi Jin. We take $n = 2$, $\beta = 0.04$ and $N = 1000$ users. $S(t)$ represents the number of $\varnothing$-users, $I_1(t)$ the number of users possessing chunk 1, $I_2(t)$ those possessing chunk 2 and $I_F(t)$ the users owning the whole file. We also assume no arrivals and also that only peers with the whole file depart at rate δ. We start $I_F(0) = 1$ and $S(0) = N - 1$.

In Figure 1, we assume no swapping and only downloading takes place. By varying the departure rate we observe how the curve I_F behaves. In Figure 2, both swapping and downloading take place. Comparing Figures 1 and 2 we see that peers in the bidirectional case become seeds slightly sooner.

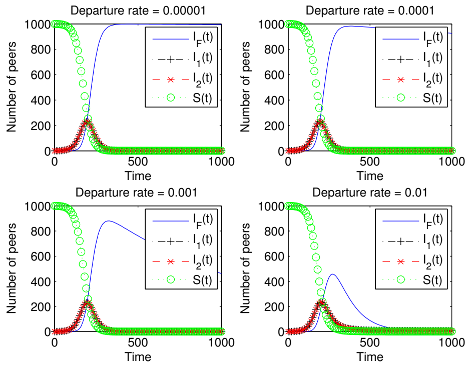

Fig 1. Evolution of peers in unidirectional transaction case

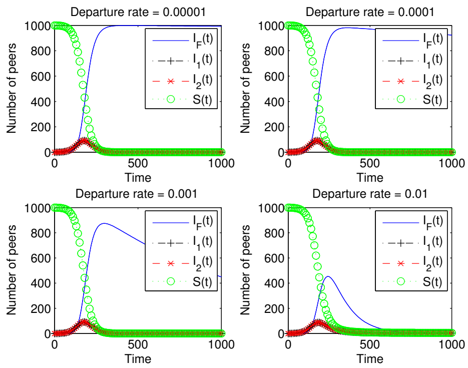

Fig 2. Evolution of peers in bidirectional transaction case

In Figure 3, we compare the evolutions of I_1 or I_2 when we vary the departure rate δ. We see, as expected, that I_1 (or I_2) with $\delta = 0.01$ decreases slower than with $\delta = 0.00001$. Since there are no arrivals, this implies that a user becomes a seed quicker.

In Figure 4, S_0 denotes the number of susceptible in the classical Kermack-McKendrick (SIR) model. Here we see that $S(t)$ hits 0 faster when we segment the file in 2 chunks rather than having the whole file unsegmented. We also see that changing the swap rate does not change things too much, which is something we observed in Figures 1 and 2.

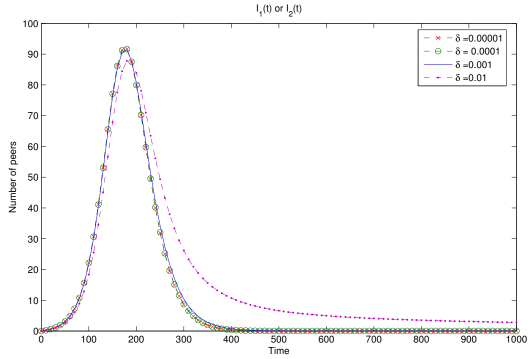

Fig 3. Evolution of peers in bidirectional transaction case

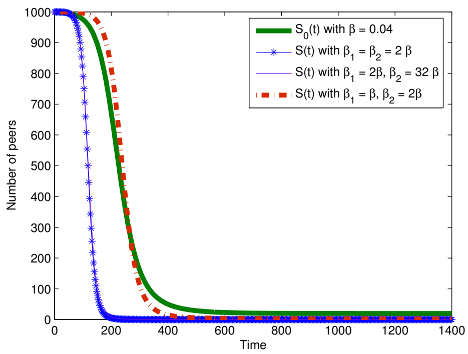

Fig 4. Comparison of $s_0(t)$ and $s(t)$

Bibliography

[1] R.W.R. Darling and J.R. Norris (2008). Differential equation approximations for Markov chains. *Probability Surveys* **5**, 37-79.

[2] M.J. Luczak and J.R. Norris (2008). Homogenization of fast variables in the fluid limit for Markov chains.

[3] J.R. Norris. *Markov chains.* Cambridge University Press

[4] D. Achlioptas (2001). Lower bounds for random 3-SAT via differential equations.*Theoretical Computer Science* **265**, 159-185

[5] Brémaud, Markov Chains

[6] D. Achlioptas and G.B. Sorkin(2000). Optimal myopic algorithms for random 3-SAT

[7] D. Achlioptas(2000). Setting 2 variables at a time yields a new lower bound for random 3-SAT

[8] DC Info (Distributed Systems Newsletter). Vol. XXII, Issue 8, June 30, 2008; Available at: `http://www.dcia.info/news/#Newsletters`

[9] G. Kesidis, Y. Jin, B. Mortazavi, and T. Konstantopoulos. An epidemiological model for file-sharing with BitTorrent-like incentives. In *Proc. IEEE GLOBECOM*, Nov. 2006.

[10] G. Kesidis, M. Vojnovic, I. Hamadeh, Y. Jin, S. Jiwasurat. A Model of the Spread of Randomly Scanning Internet Worms that Saturate Access Links. *ACM TOMACS*, May 2008.

[11] L. Massoulié and M. Vojnovic. Coupon replication systems. In *Proc. ACM SIGMETRICS*, Banff, Alberta, Canada, 2005.

[12] D. Qiu and R. Srikant. Modeling and performance analysis of BitTorrent-like peer-to-peer networks. In *Proc. ACM SIGCOMM*, Portland, Oregon, 2004.

[13] T. Kurtz. *Approximation of Population Processes.* SIAM, Philadelphia, 1981.

[14] D.J. Daley and J. Gani. *Epidemic Modeling, an Introduction.* Cambridge University Press, 1999.

[15] L. Coutin, I. Decreusefond, J.S. Dhersin A Markov model for the spread of hepatitis C virus Available at: `arXiv:0809.1824v2 [math.PR]`

[16] X. Yang and G. de Veciana. Service capacity of peer to peer networks. In *Proc. IEEE INFOCOM*, San Francisco, 2004.