Memory and burstiness in dynamic networks

Ewan R. Colman\textsuperscript{e} and Danica Vukadinović Greetham
Centre for the Mathematics of Human Behaviour,
Department of Mathematics and Statistics, University of Reading, RG6 6AX, UK
(Dated: January 23, 2015)

The mathematics of interactive complex systems has a vital role to play in the interpretation of large-scale social and biological data. Technology which facilitates the collection of vast amounts of information is increasingly becoming available for both academic and commercial purposes; however, in the absence of a detailed understanding of the underlying processes, there will always be a risk of deriving the wrong conclusion from the facts. Complexity science provides numerous models of social, biological, physical and economic systems which combine large numbers of individual components to reproduce the types of behaviour observed on the systemic level. The components in such systems are usually uninteresting in isolation, but when allowed to interact with each other they produce complex non-trivial patterns which in some cases agree very well with empirical results. This poses a challenge for data scientists: Given information only about the system as a whole, with all its complex and interactive dynamics, how can one conclude anything about the individual components? To begin answering that question we need to understand, in mathematical terms, the form and extent of the biases that complexity creates.

The purpose of the present work is to provide an understanding of how one very simple mechanism, a memory effect (brought about by interaction), will bias the statistical properties of a complex system such as the the distribution of communication activity in a social network, or the distribution of brain activity of different cortical regions in a fMRI scan. We consider a hypothetical system of individual agents (nodes) and the instantaneous pairwise interactions which happen between them (edges). By aggregating all of the interactions that occur within some given time window, a network is formed whose structure can be analysed for a deeper understanding of the system.

Throughout this paper we will be comparing two possible forms of stochastic process: Markovian and non-Markovian. In the non-Markovian case the changes which occur on the network depend on events which happened at earlier times. We will refer to this as the “memory effect”. The system we analyse here has been carefully designed so that the memory is encoded in the structure of the network. As a consequence, the number of interactions a node can ‘remember’ is effectively the same as its degree, this way the process by which the non-Markovian model develops is a form of preferential attachment. We first reference the relevant background material and related work. In the sections that follow we describe our theoretical results regarding how memory, interaction, and sampling, can influence the large scale structure of a system of interacting agents.

I. RELATED WORK

The motivation for this work is the growing evidence of memory dependent, burst-like activity in complex interactive systems. Recently this has been most prominent in the study of online communication patterns\textsuperscript{2}. Evidence for burstiness is found in the distribution of inter-event times between actions; in a Poisson process, for example, which is Markovian i.e. memoryless, the inter-event times follow an exponential distribution; periodic events, such as a heartbeat, have inter-event times which generally stay close to the mean; and lastly, in systems which are generally said to be “bursty”, the inter-event times follow a power-law distribution. A formal definition has been proposed to quantify these behaviours in\textsuperscript{3}. A slightly different approach in\textsuperscript{4} identifies a “burst” as a sequence of events where each event follows the previous one within a given time interval. This definition naturally leads to the consideration of two possible types of event: those which happen spontaneously, and those which occur as reactions to previous events (as we experience in most

\textsuperscript{E}. Colman@Reading.ac.uk
The current explanations for why a sequence of events might have a power-law inter-event time distribution rely on a memory effect; in other words the probability of an event occurring at a given time is dependent on events which occurred at previous times. Models have been proposed based on queuing theory where incoming messages are replied to according to some prioritisation strategy [5,17]. By adjusting a parameter which controls the randomness of the strategy, these models have been shown to create power-law distributed inter-event times with exponents that agree with a number of real-world data-sets.

Many of the systems in which burstiness has been observed cannot be considered to have the internal mechanics of a queuing model, these include studies of the human brain [8], animal movement patterns [4] and consumer behaviour [10]. Bursts of activity closely resemble cascading events such as avalanches and mass extinctions and therefore might possibly be examples of self-organised criticality (SOC) [11], where the focus is on the emergence of scale-free distributions based on very few assumptions about the system. Another similarity that this area has with SOC is the absence of analytical results. Only in highly simplified circumstances has any progress been achieved [12–15].

The present work examines how bursty behaviour in the individual agents in an interactive systems affects the large scale macroscopic view of the system. We consider activity on a dynamic network which is closely related to several previously studied models: Preferential attachment [1,16] is a non-Markovian method by which many networks grow. In this process, nodes are added to the network sequentially in discrete time-steps and edges are created between the new node and old nodes selected randomly but with probability proportional to their degree. The rate of growth in connectivity of a node at any given time therefore depends on its entire history. Conversely in “fitness” networks the connectivity of a node accords only to an attractiveness value drawn from some probability distribution [17]. Such models are versatile in their applications as they can incorporate various topological network features such as clustering [18]. A simple way to combine fitness and preferential attachment has been achieved by defining the attractiveness of a node to be either the sum or product (or a combination of both) of its degree and its intrinsic fitness [19,21].

The networks mentioned so far are static in the sense that once a link is created between nodes it remains in that location forever. In many situations this is not the case and we here use the term “dynamic” to refer to networks whose edges can be removed as new ones are created [22–25]. The model introduced in [26] combines the preferential selection of nodes with an added fitness parameter (the same for every node) on a dynamic network where edges are removed so that the total number of edges remains constant. The authors focus on the problem of finding the degree distribution; what they do not mention is that the degree of each individual node in this model fluctuates with a memory-driven bursty process, with power-law distributed inter-event times between each attachment event. This observation is central to the results presented in this paper.

The model we introduce is a versatile and applicable dynamic network with varying node fitness. There is currently research activity in related areas that is of much interest: Time varying networks, in particular, have some similarities with dynamic networks. Informally speaking, these are multi-layered networks where each layer corresponds to a distinct time interval, they differ from dynamic networks because at the end of each time interval the entire network (rather than just a single edge) is removed and replaced [27,28]. In its most basic form the “action potential” (the propensity to act at any given time) of each node does not change with time. In [29] Memory effects are considered within the time varying formulation. Here the authors observe in social communication data a universal rule for the probability that an individual will continue an old correspondence rather than start a new one. Adding this constraint to the original time-varying concept gives accurate results regarding the number of contacts and the weight of correspondence with each contact.

In [30] the waiting time distribution between actions takes an arbitrary form thus the action potential of each node is dynamic. When the waiting time takes a power-law form they find that the exponent of the degree distribution depends on the exponent of the waiting time distribution. The aforementioned studies do not contradict the work presented here, these papers are complimentary, together they reinforce the movement to unify bursty dynamics and network structure.

The remainder of this paper is structured as follows: We describe a model of an evolving network where edges are removed and replaced at each time-step. Within this section two possible attachment kernels are described, the first is entirely fitness based, the second has an additional preferential attachment mechanism which can be interpreted as an increased propensity to act caused by previous interactions. Some interesting results are mentioned and we present figures which show the degree distributions in some special cases of the model. In Section [III] we show how the system is described mathematically, and derive results regarding the degree distribution for a general fitness distribution. In Section [IV] we look at some special cases. In Section [V] we highlight the advances achieved by this research, we mention briefly how our results can be applied, and suggest the important developments which require further study.
II. MODEL DESCRIPTION AND MAIN RESULTS

We consider a network formed of $N$ nodes and $E$ edges. Initially the edges are placed between pairs of randomly selected nodes. For each node, a positive continuous random variable $x \in \mathbb{R}$ is selected according to a probability density function $\rho(x)$ which has mean $\langle x \rangle$. Following the related literature we shall refer to this value as the “fitness” of the node, denoted $x_i$ for the node $i$. The dynamics of the system are described as follows: in each iteration a node $i$ is randomly selected with probability given by its attachment kernel $\Pi(i)$, a second node is selected in the same way and an edge is created between them. In the same iteration the oldest edge is removed (thus $E$, $N$ and the mean degree $\langle k \rangle = 2E/N$ remain constant throughout). The process is illustrated in Fig.1 for both of the attachment kernels considered here.

In most real-life situations the fitness of a node represents some hidden (or latent) quantity, whereas its degree represents something tangible that appears in empirical data-sets. In general, then, an important problem to address is in inferring the fitness of the node when given only its degree and other properties describing the structure of the network. In a stochastic system the closest we can get to achieving this is finding the likelihood that a node has fitness $x$ given what is known about the network structure. Bayes rule gives a neat expression for this quantity:

$$P(x|k) = \frac{\rho(x)P(k|x)}{p_k}$$  \hspace{1cm} (1)

where $p_k$ is the probability of randomly selecting a node which has degree $k$, and $P(k|x)$ is the same probability but this time conditioned on $x$. Thus global information about the network is necessary; as well as being interesting in its own right the degree distribution is integral to uncovering the hidden variables. Our analysis focuses mainly on deriving the degree distribution and the conditional degree distribution for a range of fitness functions.

We consider the two following possible attachment kernels.

A. Dynamic model without memory

The probability of attaching an edge to a node $i$ of fitness $x_i$ is

$$\Pi(i) = \frac{x_i}{\sum_j x_j} = \frac{x_i}{N \langle x \rangle}$$  \hspace{1cm} (2)

Under this condition the $x_i$ can be considered the rate of activity of $i$ and one might naively assume that the relationship between $x_i$ and the degree of $i$, $k_i$, is approximately linear (specifically $k_i \approx x_i \times \langle k \rangle / \langle x \rangle$ since this would give the correct result for the total degree of the network). In general, this is not the case; Figures 2a and (2c) show that the degree distributions and fitness distributions of networks created by this process after a large number of time-steps contain fundamental differences. If $\rho(x) = \lambda e^{-\lambda x}$ then the degree distribution depends only on the mean degree of the network and not at all on the parameter $\lambda$. In this case there are therefore infinitely many possible fitness distributions which produce the same degree distribution. If $\rho(x)$ follows a power-law with exponent $\gamma$ the $p_k$ will have a power-law tail with the same exponent $\gamma$, small values of $k$, however, become increasingly uncommon as we look at denser networks.

FIG. 1. Three iterations showing the evolution of the network starting from a random initial configuration. The number of stripes inside each node corresponds to the fitness, the node with the least stripes has fitness $x = 0.5$, the others have $x = 1$, $x = 1.5$, $x = 2$, $x = 2.5$ and $x = 3$. The number of dashes in each edge corresponds to its age with the oldest having the most dashes. In each iteration the oldest edge is removed and a new edge is added between nodes selected either with probability proportional to their fitness (shown in (a)) or with probability proportional to the sum of their fitness and their degree (shown in (b)).
FIG. 2. (Colour online) The degree distributions for both types of attachment kernel and two different forms of $\rho(x)$. In each plot the (unfilled) markers show the results of a single numerical simulation of the model on a network of $2 \times 10^3$ nodes; the smooth lines show the corresponding analytical results. In (a) and (c) the filled markers show the distribution of node fitnesses, the shape of the marker indicates which fitness distribution the degree data comes from. In (a) and (b) the number of edges is $E = 4 \times 10^3$, the fitness distribution is $\rho(x) = \lambda e^{-\lambda x}$ which is a special case of the gamma distribution [Eq.(21) with $\alpha = 1$ and $\beta = 1/\lambda$], and the degree distribution is plotted for different values of $\lambda$. We see that in this particular case the parameter $\lambda$ does not affect the result in the memoryless case, however, when memory effects are introduced it does and as $\lambda$ increases the degree distribution approaches a power-law [see Eq.(28)]. In (c) and (d) the fitness distribution is given by Eq.(29) with $x_{\text{min}} = 1$ and $\gamma = 2.5$. The effect of changing the edge density is apparent here, the effect of including memory is seen in the smallest values of $k$.

B. Dynamic model with memory

The probability of attaching an edge to a node $i$ of fitness $x_i$ is

$$\Pi(i) = \frac{k_i + x_i}{\sum_j (k_j + x_j)} = \frac{k_i + x_i}{N\langle k \rangle + \langle x \rangle}$$

where $k_i$ is the degree of $i$. Memory in this system is encoded in the edges, as the current degree influences the creation of future edges. Because the edges in this system are dynamic, in that the oldest one is removed with each iteration while new ones are added, each node effectively has a memory which extends backward in time to the age of the oldest node (which happens to be $E$ time-steps). The degree of each node varies and subsequently the attractiveness $\Pi(i)$ also fluctuates. Through this memory based mechanism it is possible, by choosing a fitness distribution which ensures that $\langle x \rangle \ll \langle k \rangle$, to create burst like patterns of behaviour in the activity of $i$. This is seen in the scale-free distribution of inter-event times Fig.4 (here an “event” is the creation of an edge adjacent to $i$). Results for the degree distribution are plotted in Figs.(2b) and (2d). We find that in the case where the fitness follows a gamma distribution, $p_k$ ap-
proaches a power-law with exponent $-1$ as the mean fitness $\langle x \rangle \to 0$. If $\rho(x)$ follows a power-law with exponent $\gamma$ then $p_k$ will have a power-law tail with the same exponent $\gamma$, the effect of introducing memory is seen mostly in the lower values of $k$ which, in contrast to the memoryless case, remain relatively frequent even in dense networks. Given the degree distribution from a system whose behavior meets the description of the model, our analysis suggests a method to infer the hidden fitness distribution numerically by assuming it takes the form of a step function, reducing the problem to an optimisation problem given by Eq. (45).

III. DERIVATION OF RESULTS

A. Dynamic model without memory

For each positive integer $k$ we want to know the number of nodes $n_k$ that have degree $k$ as a function of the fitness distribution $\rho(x)$, as well as the parameters $N$ and $E$. This quantity is the expectation of the degree distribution; the mean of the ensemble of networks generated in this way. Letting $t$ be the number of iterations and $n_k(x,t)$ be the expectation of the number of nodes of degree $k$ with fitness $x$ at time $t$, we can write down the rate of change

$$\frac{\partial n_k(x,t)}{\partial t} = \frac{2x}{N(x)} [n_{k-1}(x,t) - n_k(x,t)]$$

$$+ \frac{1}{E} [(k+1)n_{k+1}(x,t) - kn_k(x,t)].$$

The first two terms on the right hand side account for the creation and destruction (respectively) of nodes of degree $k$ which occurs when an edge is attached to a node of degree $k-1$ (creation) or to a node of degree $k$ (destruction). The last two terms on the right hand side account for the creation and destruction of nodes of degree $k$ which occurs when the oldest edge is removed from a node of degree $k+1$ (creation) or removed from a node of degree $k$ (destruction). We have made the assumption here that the degree of the node is not correlated with its age. After a large number of iterations the system will be in equilibrium, $n_k(x,t) = n_k(x)$, and the left hand side will be equal to zero. Using a similar method to that found in [26] we solve Eq. (4) by introducing

$$H(k,x) = \frac{2x}{N(x)} n_k(x)$$

and

$$G(k,x) = \frac{1}{E} k n_k(x).$$

Eq. (4) now becomes

$$G(k+1) - G(k) = H(k) - H(k-1).$$

By summing Eqs. (6) over all $k \geq 1$ we find that $G(0,x) = H(1,x)$ and consequently $G(k,x) = H(k-1,x)$ for all $k \geq 1$, solving this leads to

$$n_k(x) = \left( \frac{2Ex}{N(x)} \right)^k \frac{1}{k!} n_0(x).$$

To find $n_0$ we consider $N(x)$, the expected number of nodes of fitness $x$,

$$N(x) = \sum_{k=0}^{\infty} (lx)^k \frac{1}{k!} n_0(x) = n_0(x) e^{lx}$$

where $l = 2E/N(x)$. The conditional probability $P(k|x) = n_k(x)/N(x)$ is found by combining Eq. (7) and Eq. (8) to get

$$P(k|x) = \frac{1}{k!} (lx)^k e^{-lx}.$$
To finally reveal the fraction of nodes in the entire network of degree \( k \), \( p_k = n_k/N \), we need to solve the integral

\[
p_k = \int_0^\infty \rho(x)P(k|x)dx = \frac{k^k}{k!} \int_0^\infty \rho(x)x^k e^{-x}dx. \tag{10}
\]

This is as far as a the general solution can be taken but the solutions for two special forms of \( \rho(x) \) are presented in Section IV.

### B. Dynamic model with memory

The rule determining whether a node is active at any given time can be divided into two constituent mechanisms: one is regarded as a reaction to one or more previous interactions, it is memory dependent and is responsible for bursts of activity. The other is the fitness of the node which encompasses all the other reasons why a node may become active at any given time. Modifying the model (Eq. (4)) for the new kernel Eq. (3) we now have

\[
\frac{\partial n_k(x, t)}{\partial t} = \frac{1}{N(k + (x))}[(k + x - 1)n_{k-1}(x, t) - (k + x)n_k(x, t)] + \frac{1}{E}[n_{k+1}(x, t) - kn_k(x, t)]. \tag{11}
\]

As before, we set the left hand side to zero to get a difference equation

\[
k_{n_k}(x) - (k + 1)n_{k-1}(x) = m[(k + 1)n_{k-1}(x) - kn_k(x) + x n_{k-1}(x) - x n_k(x)], \tag{12}
\]

where \( m = \langle k \rangle / (\langle k \rangle + \langle x \rangle) \). To solve this we introduce the generating function,

\[
g(z, x) = \sum_{k=0}^\infty n_k(x)z^k \tag{13}
\]

by multiplying Eq. (12) by \( z^k \) and summing over all \( k \geq 0 \) we arrive at

\[
(z - 1)(1 - m z) \frac{\partial g(z, x)}{\partial z} - mx(z - 1)g(z, x) = 0 \tag{14}
\]

which has the solution \( g(z, x) = [C(1 - m z)]^{-x} \) (a general description of this method is described in the appendix of [23]). We find \( C \) by substituting \( g(1, x) = N(x) \) into the solution and get

\[
g(z, x) = N(x) \left( \frac{1 - m}{1 - m z} \right)^x. \tag{15}
\]

The coefficient of \( z^k \) in the expansion of the right hand side is \( n_k(x) \), dividing this by \( N(x) \) then gives the following conditional probability which contrasts with Eq. (9)

\[
P(k|x) = \left( \frac{x + k - 1}{k} \right)(1 - m)^x m^k \tag{16}
\]

As \( \langle x \rangle \to \infty \), \( P(k|x) \) tends towards the Poisson distribution with the same mean we had in Eq. (9). This is expected since in this limit the attachment kernel for any given node will be dominated by its fitness. Let \( p_k \) be the fraction of nodes with degree \( k \) and is the integral of the product of \( \rho(x) \) and the right hand side of Eq. (16) over all possible values of \( x \)

\[
p_k = \frac{m^k}{k!} \int_0^\infty x(x + 1)...(x + k - 1)(1 - m)^x \rho(x)dx. \tag{17}
\]

We can simplify the integral by multiplying out all the the brackets which contain \( x \), this gives

\[
p_k = \frac{m^k}{k!} \sum_{n=0}^k c(k, n) \int_0^\infty x^n(1 - m)^x \rho(x)dx. \tag{18}
\]

Here \( c(k, n) \) denotes the unsigned Stirling numbers of the first kind (the number of permutations of \( k \) symbols that have exactly \( n \) cycles [21]), since an explicit expression for these is not known, Eq. (18) is only useful at small values of \( k \). For large \( k \) we examine the generating function

\[
G(z) = \sum_{k=0}^\infty p_k z^k. \tag{19}
\]

It follows from Eq. (15) that

\[
G(z) = \int_0^\infty \rho(x) \left( \frac{1 - m}{1 - m z} \right)^x dx. \tag{20}
\]

When the fitness parameter is the same for all nodes, \( x_i = \alpha \), the model reduces to that studied in [26]. Substituting \( \rho(x) = \delta(x - \alpha) \) into Eq. (20) yields the expected result.

### IV. EXAMPLES

#### A. Gamma distribution

We examine in detail the possible scenario where the fitness of the population follows the gamma distribution

\[
\rho(x; \alpha, \beta) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \tag{21}
\]

which generalises a number of distributions that have applications in social sciences including \( \chi^2 \) and the exponential distribution. In general it has the appearance of an asymmetric bell curve and we consider it entirely...
likely that a system might exist where the fitness values are clustered around the mean in this way.

**Dynamic model without memory:** We solve Eq. (10) to find the degree distribution. The integral becomes

\[ p_k = \frac{k^k}{k! \beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{k+\alpha-1} e^{-x(t+1/\beta)} dx. \]  

(22)

By applying the change of variables \( y = (t+1/\beta)x \) the integral becomes the product of a gamma function and some other factors. We arrive at

\[ p_k = \frac{k^k \Gamma(k+\alpha)}{\beta^\alpha (t+1/\beta)^{k+\alpha} \Gamma(\alpha) k!}. \]  

(23)

For large values of \( k \) this solution becomes a gamma distribution \( p_k \sim \rho(k; \alpha, \log(1+1/\beta)) \).

**Dynamic model with memory:** We substitute Eq. (21) into Eq. (18) and applying the change of variables; \( y = x[(1/\beta)-\log(1-m)] \), we can again take a gamma function out as a factor, leaving

\[ p_k = \frac{m^k}{k! \beta^\alpha \Gamma(\alpha)} \sum_{n=0}^k c(k,n) \Gamma(n+\alpha) \left[ \log \left( \frac{1}{1-m} \right) + \frac{1}{\beta} \right]^{-(n+\alpha)}. \]  

(24)

Substituting Eq. (21) into Eq. (20) and solving the integral we arrive at

\[ G(z) = \left[ 1 + \log \left( \frac{1-mz}{1-m} \right) \right]^{\beta}. \]  

(25)

As \( z \to 1 \), the logarithm in the above expression approaches 0 making the approximation \( \log(X) \approx X-1 \) appropriate to use. For \( z \approx 1 \) we have

\[ G(z) \approx \left( \frac{1-m}{1-mz} \right)^{\alpha \beta}. \]  

(26)

which can be expanded to recover the power series. Equating the coefficients of the expansion with those of Eq. (19) we find

\[ p_k \approx (1-m)^{\alpha \beta} \left( -\frac{\alpha \beta}{k} \right) (-m)^k \]  

(27)

For large \( k \) this is

\[ p_k \approx \frac{(1-m)^{\alpha \beta}}{\Gamma(\alpha \beta)} m^k k^{\alpha \beta-1} = c \rho \left( k; \alpha \beta, \frac{1}{\log(1/m)} \right) \]  

(28)

where \( c = [(1-m)/\log(1/m)]^{\alpha \beta} \) is a normalising constant. As the mean \( \langle x \rangle = \alpha \beta \) tends towards 0 the distribution tends towards a power-law with exponent -1. This represents a scenario where the majority of actions are in fact reactions to previous events.

**B. Power-law distribution**

Suppose fitness is distributed according to the following power-law

\[ \rho(x; \min, \gamma) = \begin{cases} \frac{\gamma-1}{\min} \left( \frac{x}{\min} \right)^{-\gamma} & \text{if } x \geq \min \\ 0 & \text{if } x < \min \end{cases} \]  

(29)

which has the mean

\[ \langle x \rangle = \frac{\gamma-1}{\gamma-2} \min. \]  

(30)

**Dynamic model without memory:** To find the degree distribution we substitute Eq. (20) into Eq. (10), giving

\[ p_k = \frac{(\gamma-1)k^k}{x_{\min}^{\gamma-1} k!} \int_{x_{\min}}^\infty x^{k-\gamma} e^{-lx} dx. \]  

(31)

Using the substitution \( y = lx \), the integral can be expressed as an incomplete gamma function (see (21)) defined as \( \Gamma(u, v) = \int_u^\infty e^{-v} dv \) for real numbers \( u \) and \( v \). We can also simplify the solution by combining the parameters using

\[ A = lx_{\min} = \frac{2E(\gamma-2)}{N(\gamma-1)}. \]  

(32)

and we get

\[ p_k = \frac{(\gamma-1)A^{\gamma-1}}{k!} \Gamma(k-\gamma+1, A). \]  

(33)

Notice that all choices of \( x_{\min} \) yield the same result. This is not unexpected: the scale-invariance of the power law distribution means that generating a random fitness \( x_i \) using Eq. (20) is equivalent to generating \( \xi \) from \( \rho(\xi, 1, \gamma) \) and taking \( x_i = \xi x_{\min} \) as the fitness value. Substituting this fitness value into Eq. (2) we see that \( x_{\min} \) is no longer present. We show in the Appendix that if \( \gamma \) is a positive integer then we also have

\[ p_k = \frac{(\gamma-1)A^{\gamma-1} e^{-A}}{k(k-1) \ldots (k-\gamma+1)} \sum_{s=0}^{k-\gamma} A^s. \]  

(34)

It is now easy to see that the degree distribution has a power-law tail (see Fig. (2)).

**Dynamic model with memory:** First we substitute Eq. (20) into Eq. (18). We introduce \( L = -\log(1-m) \), then, by applying a change of variables \( y = Lx \), we can factorise out an incomplete gamma function. This gives the following exact solution for the degree distribution

\[ p_k = \frac{m^k(k-1)^{-\gamma}}{k! x_{\min}^{\gamma-1}} \sum_{n=0}^k c(k,n) L^{n-1} \Gamma(n-\gamma+1, Lx_{\min}). \]  

(35)
For large values of $k$ we solve Eq.(20) to find
\[
G(z) = (\gamma - 1)\Gamma[1 - \gamma, \Phi(x_{\text{min}}, z)[\Phi(x_{\text{min}}, z)]^{-1} \tag{36}
\]
where
\[
\Phi(x_{\text{min}}, z) = x_{\text{min}} \log \left( \frac{1 - mz}{1 - m} \right). \tag{37}
\]
Using the approximation $\log(X) \approx X - 1$ as $z$ approaches 1 we find
\[
G(z) \approx (\gamma - 1)^{-1}\Gamma(1 - \gamma)|x_{\text{min}}A|^{-1}(1 - z)^{-1} \tag{38}
\]
where $A$ is given by Eq.(32). For non-integer values of $\gamma$ this can be expanded and the coefficients of the expansion can be equated with Eq.(19). We see that
\[
p_k \approx \frac{(\gamma - 1)A^{-1}\Gamma(k - \gamma + 1)}{\Gamma(k + 1)}. \tag{39}
\]
The parameter $x_{\text{min}}$, which was absent in Eq.(33), now controls the overall effect of fitness in proportion to memory. The $k$ dependence exists in the form of the ratio of two gamma functions so asymptotically $p_k \sim k^{-\gamma}$. It is worth remarking that the power-law exponent in the fitness distribution is the same exponent found in the degree distribution and is not affected by the choice of the other parameters $N, E$ or $x_{\text{min}}$ as can be seen in Fig.(2d).

### C. Step function distribution

For practical purposes it is useful to have a general method of inferring a fitness distribution from a degree distribution. We suggest one such approach here and focus exclusively on the case where memory effects are present.

By assuming the fitness distribution has the form of a step function (otherwise know as a staircase function) we can minimise the error between the theoretical prediction and the observed data by adjusting the height of each step (or stair). Suppose we have a vector of parameters $\mathbf{a} = [a_0, a_1, \ldots, a_J]$, we then define the distribution as
\[
\rho(x, \mathbf{a}, \delta) = \begin{cases} 
 a_0 & \text{for } 0 < x \leq \delta \\
 a_1 & \text{for } \delta < x \leq 2\delta \\
 \vdots & \\
 a_j & \text{for } j\delta < x \leq (j + 1)\delta \\
 \vdots & \\
 a_J & \text{for } J\delta < x \leq (J + 1)\delta 
\end{cases} \tag{40}
\]
where $\sum_{i=0}^{J} a_j = [\delta J]^{-1}$. The mean fitness is
\[
\langle x \rangle = \mathbf{a} \mathbf{v} \tag{41}
\]
where $\mathbf{v} = ([\delta/2][1, 3, \ldots, 2J + 1]$. Substituting $\rho$ into Eq.(20) we get
\[
G(z) = \sum_{i=0}^{J} a_i \int_{i\delta}^{(i+1)\delta} \left( \frac{1 - m}{1 - mz} \right)^x \, dx
= \frac{[(1 - m)/(1 - mz)]^\delta - 1}{\log[(1 - m)/(1 - mz)]^\delta} \sum_{i=0}^{J} a_i \left( \frac{1 - m}{1 - mz} \right)^{\delta_i}. \tag{42}
\]

We can generate (randomly or systematically) a vector of values $\mathbf{z} = [z_0, z_1, \ldots, z_J]$ at which the generating function can be evaluated. The empirical data is the degree distribution $p$ whose generating function as given by Eq.(19), fitness parameters $\mathbf{a}$, and the generating function as given by Eq.(42) are all connected by the following expression:
\[
Zp = Wa. \tag{43}
\]
Here $Z$ is a $I \times K$ matrix whose $(i, k)$th entry is $z_{i,k} = z_{i-1}^{-k}$ and $W$ is a $I \times J$ matrix whose $(i, j)$th entry is given by
\[
w_{i,j} = \frac{[(1 - m)/(1 - mz_{i-1})]^\delta - 1}{\log[(1 - m)/(1 - mz_{i-1})]^\delta} \left( \frac{1 - m}{1 - mz_{i-1}} \right)^{\delta(j-1)}. \tag{44}
\]
While Eq.(43) appears to be a simple linear algebra problem, it is complicated by the fact that $m$ depends on $\langle x \rangle$, which is only known after a choice of $\mathbf{a}$ has been made, therefore $W$ is a function of both $\mathbf{a}$ and $\delta$. This does however provide a neat way to formally present the problem: We choose $J < K$ to prevent over-fitting (i.e. having more parameters than datapoints) and solve
\[
\rho(x) = \rho(x, \mathbf{a}, \delta)
\]
such that
\[
\|Zp - Wa\| = \min_{\mathbf{a}, \delta} \|Zp - Wa\|. \tag{45}
\]

### V. DISCUSSION

This analysis has potential to be useful in many applications. Suppose we have a system where data arrives in the form of a list of interactions between a finite number of agents, this model provides a framework for interpreting such data. A sample of say, $n$ interactions, can be thought of as a network with $n$ edges, all of which are placed according to some hidden fitness variable which the present model makes no assumptions about. (It should also be noted that the assumption in our model that interactions are pairwise can easily be generalised so that any given number of nodes may be active at each time-step.) We have shown the impact of the edge density (i.e. the size of the sample) on the degree distribution and the effect of bursty, memory driven, behaviour.
We also suggested a method to recover the hidden fitness distribution from the data. We note that the variability with edge density is very similar to the problem of time varying networks discussed in [27] although the authors here focus on the issue of not counting multiple edges more than once (something the present analysis ignores) and are not concerned with the aggregate network after a long time when it reaches a high density. The results of this paper have shown that the effect at high densities is profound and can be significantly altered by the addition of memory.

The motivation for this work was the potential applicability to two specific areas of data analysis: Online social interactions (e.g. Twitter) and the data received from fMRI scans, in particular when the cortical regions of the brain are considered as nodes and activity may transmit from one region to another (see for example [32, 33]). Both systems are known to exhibit bursty activity. In the brain are considered as nodes and activity may transmit from one region to another (see for example [32, 33]). Both systems are known to exhibit bursty activity. In the brain are considered as nodes and activity may transmit from one region to another (see for example [32, 33]). Both systems are known to exhibit bursty activity. In the brain are considered as nodes and activity may transmit from one region to another (see for example [32, 33]). Both systems are known to exhibit bursty activity. In the brain are considered as nodes and activity may transmit from one region to another (see for example [32, 33]). Both systems are known to exhibit bursty activity. In the brain are considered as nodes and activity may transmit from one region to another (see for example [32, 33]). Both systems are known to exhibit bursty activity. In the brain are considered as nodes and activity may transmit from one region to another (see for example [32, 33]). Both systems are known to exhibit bursty activity. In the brain are considered as nodes and activity may transmit from one region to another (see for example [32, 33]). Both systems are known to exhibit bursty activity. In the brain are considered as nodes and activity may transmit from one region to another (see for example [32, 33]). Both systems are known to exhibit bursty activity. In the brain are considered as nodes and activity may transmit from one region to another (see for example [32, 33]).

Our last remark is a mention of the burst pattern result observed in the activity of nodes (Fig. 3). In many studies of burstiness (such as [30]) the power-law inter-event time distribution is included as an a priori assumption in the description of the model. We have shown that this pattern can emerge from a simpler, lower-level process, suggesting that there could be a universal reason why these patterns are observed so frequently in complex systems. The relationship between this result and the well studied SOC models needs to be established in order to move towards an analytical understanding of both phenomena, hopefully broadening this model to a wide range of universality classes thus improving its potential applicability.

**ACKNOWLEDGMENTS**

This work is funded by the RCUK Digital Economy programme via EPSRC Grant EP/G065802/1 ‘The Horizon Hub’.

**Appendix: Solution to Eq. (31) for integer values of γ**

We can express the part inside the integral as a derivative

\[ x^{k-\gamma}e^{-lx} = (-1)^{k-\gamma} \frac{d^{k-\gamma}e^{-xy}}{dy^{k-\gamma}} \bigg|_{y=l} \]  

(A.1)

before performing the integration with respect to \( x \). Since

\[ \int_{x_{\text{min}}}^{\infty} e^{-xy} dx = \frac{\exp(-x_{\text{min}}y)}{y} \]  

(A.2)

and

\[ \frac{d^n}{dy^n} \left( \frac{\exp(-x_{\text{min}}y)}{y} \right) = (-1)^n \exp(-x_{\text{min}}y) \sum_{s=0}^{n} \frac{n!}{s!} x_{\text{min}}^s y^{s-n-1} \]

for \( n \in \mathbb{N} \), for integer values of \( \gamma \) we arrive at

\[ p_k = \frac{(\gamma - 1)(lx_{\text{min}})^{\gamma - 1} \exp(-lx_{\text{min}})}{k(k-1) \ldots (k-\gamma+1)} \sum_{s=0}^{k-\gamma} \frac{(lx_{\text{min}})^s}{s!}. \]

which, using Eq. (32) simplifies to

\[ p_k = \frac{(\gamma - 1)A^{\gamma - 1}e^{-A}}{k(k-1) \ldots (k-\gamma+1)} \sum_{s=0}^{k-\gamma} \frac{A^s}{s!}. \]  

(A.5)

**References**

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. 74, 47 (2002).
[2] S. Gaito, M. Zignani, G. P. Rossi, A. Sala, X. Zhao, H. Zheng, and B. Y. Zhao, in Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research (ACM, 2012) pp. 1–8.
[3] K.-I. Goh and A.-L. Barabási, EPL (Europhysics Letters) 81, 48002 (2008).
[4] M. Karsai, K. Kaski, A.-L. Barabási, and J. Kertész, Scientific reports 2 (2012).
[5] A.-L. Barabasi, Nature 435, 207 (2005).
[6] B. Gonçalves, N. Perra, and A. Vespignani, PloS one 6, e22656 (2011).
[7] T. Maillart, D. Sornette, S. Frei, T. Duebendorfer, and A. Saichev, Phys. Rev. E 83, 056101 (2011).
[8] J. M. Beggs and D. Plenz, The Journal of neuroscience 23, 11167 (2003).
[9] A. Reynolds, Physica A: Statistical Mechanics and its Applications 390, 245 (2011).
[10] F. Deschatres and D. Sornette, Physical Review E 72, 016112 (2005).
[11] P. Bak, *How nature works* (Oxford university press Oxford, 1997).

[12] A. Vázquez, J. G. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabási, Physical Review E 73, 036127 (2006).

[13] S. Vajna, B. Tóth, and J. Kertész, New Journal of Physics 15, 103023 (2013).

[14] Z. Jiang, Y. Zhang, H. Wang, and P. Li, Journal of Statistical Mechanics: Theory and Experiment 2013, P02006 (2013).

[15] A. Vázquez, Physica A: Statistical Mechanics and its Applications 373, 747 (2007).

[16] D. D. S. Price, Journal of the American Society for Information Science 27, 292 (1976).

[17] G. Caldarelli, A. Capocci, P. De Los Rios, and M. Muñoz, Phys. Rev. Lett. 89, 258702 (2002).

[18] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Physical review letters 89, 258702 (2002).

[19] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Physical Review Letters 85, 4633 (2000).

[20] G. Bianconi and A.-L. Barabási, EPL (Europhysics Letters) 54, 436 (2001).

[21] G. Ergün and G. Rodgers, Physica A: Statistical Mechanics and its Applications 303, 261 (2002).

[22] C. Moore, G. Ghoshal, and M. E. J. Newman, Phys. Rev. E 74, 036121 (2006).

[23] E. Ben-Naim and P. Krapivsky, Journal of Physics A: Mathematical and Theoretical 40, 8607 (2007).

[24] T. Evans, The European Physical Journal B 56, 65 (2007).

[25] E. R. Colman and G. J. Rodgers, Physica A: Statistical Mechanics and its Applications 416, 80 (2014).

[26] Y.-B. Xie, T. Zhou, and B.-H. Wang, Physica A: Statistical Mechanics and its Applications 387, 1683 (2008).

[27] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, Scientific reports 2 (2012).

[28] K. Hoppe and G. Rodgers, Physical Review E 88, 042804 (2013).

[29] M. Karsai, N. Perra, and A. Vespignani, Scientific reports 4 (2014).

[30] A. Moinet, M. Starnini, and R. Pastor-Satorras, arXiv preprint arXiv:1412.0587 (2014).

[31] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1975).

[32] P. Grindrod, Z. V. Stoyanov, G. M. Smith, and J. D. Saddy, Journal of Complex Networks , cnt015 (2013).

[33] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. Hellyer, and F. Vaccarino, Journal of The Royal Society Interface 11, 20140873 (2014).