Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Memory effects induce structure in social networks with activity-driven agents

A D Medus and C O Dorso

Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón 1, Ciudad Universitaria, Ciudad Autónoma de Buenos Aires (1428), Argentina
E-mail: admedus@df.uba.ar and codorso@df.uba.ar

Received 17 April 2014
Accepted for publication 13 July 2014
Published

Online at stacks.iop.org/JSTAT/00/000000
doi:10.1088/JSTAT/2014/00/000000

Abstract. Activity-driven modelling has recently been proposed as an alternative growth mechanism for time varying networks,displaying power-law degree distribution in time-aggregated representation. This approach assumes memoryless agents developing random connections with total disregard of their previous contacts. Thus, such an assumption leads to time-aggregated random networks that do not reproduce the positive degree-degree correlation and high clustering coefficient widely observed in real social networks. In this paper, we aim to study the incidence of the agents' long-term memory on the emergence of new social ties. To this end, we propose a dynamical network model assuming heterogeneous activity for agents, together with a triadic-closure step as main connectivity mechanism. We show that this simple mechanism provides some of the fundamental topological features expected for real social networks in their time-aggregated picture. We derive analytical results and perform extensive numerical simulations in regimes with and without population growth. Finally, we present an illustrative comparison with two case studies, one comprising face-to-face encounters in a closed gathering, while the other one corresponding to social friendship ties from an online social network.

Keywords: stochastic processes (theory), analysis of algorithms, growth processes, network dynamics

1742-5468/14/000000+22$33.00

JNL: JSTAT PIPS: 499229 TYPE: PAP TS: NEWGEN DATE: 2/8/2014 EDITOR: CB SPELLING: UK

Contents

1. Introduction

Social networks represent the different substrates on which we develop many aspects of our lives. Knowledge, news, rumours and diseases are transmitted through an intricate social framework usually represented by a complex network, which explains the growing interest of scientific community in such complex systems.

Many of the early social networks analyzed in the literature from about 15 years ago are, in fact, the aggregated picture of some dynamical growing process. At that time, there was almost no access to time-resolved data. Thus, researchers mainly worked on time-aggregated networks just as if they were static objects. The efforts were mostly focused on the study of topological features and on proposing a variety of growth mechanisms in order to reproduce it. Different topological characteristics of complex networks, as node degree distribution, clustering coefficient, average shortest path length, modularity and degree-degree correlation have proved to be related with their spreading properties [1–6]. In particular, many kinds of human acquaintance networks also display heterogeneous

J. Stat. Mech. (2014) 000000

degree distribution (mainly power-law), high clustering coefficient [7], strong communality [8, 9] and positive degree-degree correlation (i.e. degree assortativity) [10] as distinctive features. Concerning the network growth mechanisms, preferential attachment (**PA**) constitutes one of the most widespread heuristic mechanisms giving rise to networks with the ubiquitous power-law degree distribution (*scale-free* networks) [11–14].

The current massivity of new information technologies has enabled the availability of huge amounts of time-resolved data from virtual social networks as well as from real social contacts between individuals [15–18]. In the light of this new insight, the exploration of the microscopic dynamic of social contacts is possible nowadays. Time-varying networks have been recently raised as a dynamic variant of the original static network representation. This new approach accounts for the changing nature of many empirical networks, comprising evolving interactions by creating and, eventually, deleting edges between nodes. In this context, an alternative mechanism based on the concept of *activity rate* was recently proposed in order to explain the scale-free feature from a dynamic perspective [19]. The activity rate describes the degree of participation of a given individual in a particular social network. Participation can account for published papers in the case of scientific collaboration networks, movies and TV series filmed in actor networks, or messages shared in the Twitter microblogging network. For each time step δt, a node is activated or not with probability proportional to its activity rate. Active nodes perform random connections to other nodes chosen over the entire population. Then, while the greater the activity rate of a given node, the greater his accumulated acquaintances in a given time window. Similar concepts like *fitness* [20, 21] and *attractiveness* [22], have been introduced in previous growth models for static networks, in order to take account of intrinsic nodes' ability to acquire new social connections. These approaches can be encompassed in a general *hidden-variable* model, by which nodes are tagged with some feature that fully determines the final network topology [23].

The activity-driven model (**AD**) assumes memoryless agents that only perform uniformly random connections, properly reproducing the time dynamic of contacts and giving place to time-aggregated random scale-free networks [19]. However, the **AD** model shows a downside: it cannot replicate the high clustering coefficient and degree assortativity characteristics of many social networks in their time-aggregated picture, as shown in [24].

Beyond their particular characteristics, every social network representation is the result of an aggregation process over a given time window Δt at some instant $T \geqslant \Delta t$. Every plausible network growth process should be able to reproduce the main topological features observed in real social networks, in particular when $\Delta t = T$. The vast majority of empirical data sets suggest the need for some local connectivity mechanism promoting transitive ties between social agents [25–28]. This particular preference towards transitive ties is a sign showing that agents should have records of their previous social contacts, even after their effective social interactions have been concluded. Such behaviour is observed in real social networks and we study two examples in section 7.

In this work we introduce a generalized stochastic growth model (**GSG**) assuming a population of nodes with heterogeneous activity rates and long-term memory. Then, once an edge emerges between two nodes at a given time, it remains in the system memory from then on. This last characteristic allows to introduce the impact of the current social environment of nodes on their further social development, drawing a parallel with the former growth mechanisms for static networks.

J. Stat. Mech. (2014) 000000

GSG constitutes a dynamical network model with time-varying edges. However, here we are not focused on the dynamical features of contacts, but instead on the topological properties of time-aggregated networks. We show that some of the topological features observed in real social networks can be recovered also for dynamical networks by following the prescriptions of a strictly local connectivity mechanism. For this purpose, **GSG** model is based on a combination of random ties together with a triadic-closure (**TC**) mechanism, the latter being well known for adding structure to the network [29–33]. In contrast with the hidden-variable model, we show that activity rates do not fully determine the final network topology. Instead, it is the combined action of **AD** and **TC** mechanisms that defines this topological features in **GSG** model. Finally, we show that the original **AD** model of [19] can be recovered as a particular case of our **GSG** model.

The rest of this paper is organized as follows: In section 2, we introduce the details of the **GSG** model. In section 3, we present the analytical treatment for the degree distribution. In sections 4 and 5, we present exact and approximate analytical solutions together with extensive numerical simulations for constant and growing population, respectively. Degree-degree correlations and clustering are studied by means of numerical simulations in section 6. In section 7, we analyze two real social networks, the first corresponding to face-to-face encounters in a closed gathering under constant population and the second to a subgraph of Facebook online friendship network. In section 8, all relevant results are summarized and discussed. Finally, in the appendix we propose a brief review of the hidden-variable model.

2. The model

Following the traditional network representation, nodes and edges correspond respectively to individuals and their ties in a social context. Let $G_t(N;L)$ represent a network, or graph, composed of $N(t)$ nodes and $L(t)$ edges at time t. As in **AD** model, **GSG** assign to each node $i \in \{1,\ldots,N\}$ an activity rate a_i from a given activity pdf $F(a)$.

First, we will remind ourselves of some aspects of **AD** model. It assumes that nodes are activated for each time interval δt with probability proportional to their activity rate. From each active node m edges arise, that will be connected to other nodes (actives or not) chosen uniformly at random without any memory of its previous connections. Finally, all edges are deleted after each time interval δt. This process gives rise to a time-aggregated random network with degree distribution inherited from the corresponding activity density function $F(a)$. It is important to note that the same edge may be repeated for different time steps in **AD** model.

In the context of **GSG** we assume that activity a_i represents, in fact, the rate at which new edges emerge from node i. Thus, high activity nodes are more prone to acquire new connections. As the **AD** model, **GSG** is a model of dynamical networks with time-varying edges. Instead of multiple random connections the edges are added one by one, remaining in the system memory from then on (this point will be made clear later on). The edges are introduced with a rate given by:

$$\beta(t) = \sum_{i=1}^{N(t)} a_i. \tag{1}$$

Alternatively, each added edge l_j will be deleted with rate μ_j from a pdf $F_D(\mu)$, so that the total edges' deletion rate $\eta(t)$ results

$$\eta(t) = \sum_{j=1}^{L(t)} \mu_j. \tag{2}$$

However, every edge will remain in the memory of the involved nodes, even after having been deleted, thus influencing their further connections.

In any case, here we are focused on studying the topological features of time-aggregated networks coming from **GSG** model[1]. As a consequence, all edges will remain in the time-aggregated representation as if they were static. Subsequently, the results that we show in the following sections apply to any deletion mechanism of edges.

2.1. Memory effects: triadic-closure mechanism

In **GSG** model, edges are added following a mixed connectivity mechanism. For each edge, the source node is selected proportionally to its activity rate, while the target node is chosen by the following procedure: (a) with probability q a second-neighbour of the source node is chosen in order to 'close a triangle'(**TC** mechanism), or (b) a random target node is selected with probability $(1-q)$.

Triadic-closure mechanism has been observed in many real social networks and is widely recognized as one of the most direct and natural ways to introduce transitivity (or clustering) in network growth models [25, 30–35]. It can also be understood as a way to replicate what we often experience in our social relations, namely, that usually new acquaintances are introduced to us through our current social environment. It is in this sense that we refer to long-term memory in agents, since everyone has an internal record of their previous contacts. In this work, we assume ideal agents with infinite memory.

Agents'memory may also suggest the recurrence of previous contacts giving place to weighted edges. Nevertheless, here we are focused on addressing the network development under a parsimonious approach with unweighted edges (see [36] for the relation of long-term memory with a reinforcement process). An example of the connectivity evolution for **GSG** model is shown in figure 1. The impact of **TC** mechanism is clearly evidenced in this sequence.

In the next sections, we present a detailed analytical formulation of the time-aggregated representation of the model, in order to obtain the degree distribution under two regimes: (i) with constant population ($\gamma = 0$) and (ii) with population growth ($\gamma > 0$).

3. Analytical formulation

We begin by defining the elements of **GSG** analytical formulation for time-aggregated representation on a general framework. We represent this time-aggregated picture as the final outcome of a continuous time stochastic process $\{L_t\}_{t\in\mathbb{R}_{\geqslant 0}}$ with edges' birth rate $\beta(t)$. In order to simplify the subsequent analysis, we define an embedded discrete stochastic

[1] Dynamical aspects of **GSG** model such as contacts duration and frequency (reinforcement process) will be discussed in a future paper.

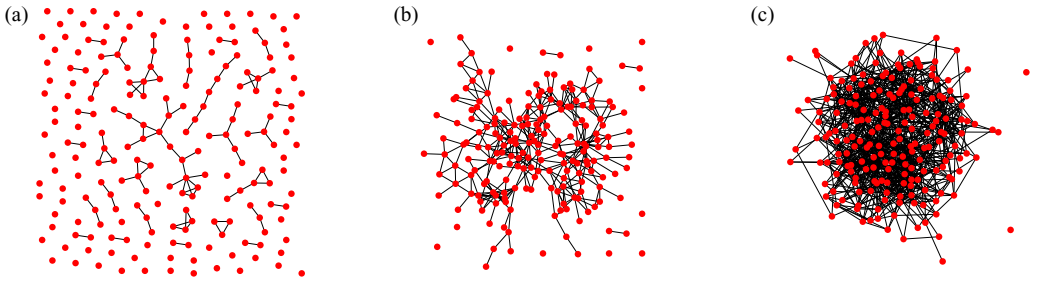

Figure 1. (a) $\ell = 100$ (b) $\ell = 400$ (c) $\ell = 1000$ An example of three network growth stages (in terms of ℓ) are shown for **GSG** model under constant population with $N = 200$ and triadic-closure probability $q = 0.7$.

process $\{L_\ell\}_{\ell\in\mathbb{N}}$ with ℓ the aggregated number of added edges. In this way, we also define the nodes' population growth rate γ in terms of the edges' population growth rate (now formally equal to 1). Hence, the evolution of the total number of edges and nodes are respectively given by $L(\ell) = L_0 + (1+\gamma)\ell$, where $\gamma\ell$ comes from those edges associated with added nodes and $N(\ell) = N_0 + \gamma\ell$, with L_0 and N_0 the initial values.

Let $P_{k|a}(\ell, \ell_0)$ be the probability that a node introduced at step ℓ_0 with activity rate a has degree k for a subsequent algorithm step $\ell > \ell_0$. Now we can define $\bar{N}_{k|a}(\ell)$, the mean number of nodes with degree k within those with activity rate a, as

$$\bar{N}_{k|a}(\ell) = \sum_{\ell_0=1}^{\ell} P_{k|a}(\ell, \ell_0). \tag{3}$$

Thus, $\bar{N}_{k|a}(\ell)$ is equivalent to the propagator for hidden-variable model [23]. The evolution of $\bar{N}_{k|a}(\ell)$ can be described from a continuum approach through a system of coupled rate equations [12, 37] as follows:

$$\begin{aligned}\frac{\mathrm{d}\bar{N}_{k|a}}{\mathrm{d}\ell} = q\left(\Theta(k-1,a,\ell)\bar{N}_{k-1|a} - \Theta(k,a,\ell)\bar{N}_{k|a}\right) + \frac{1-q}{N(\ell)}\left(\frac{a}{\langle a\rangle}+1\right)\left(\bar{N}_{k-1|a} - \bar{N}_{k|a}\right) \\ + \frac{\gamma}{N(\ell)}\left(\bar{N}_{k-1|a} - \bar{N}_{k|a}\right) + \gamma\delta_{k1},\end{aligned} \tag{4}$$

where the first term in the right-hand-side of (4) is associated with the **TC** mechanism contribution to $\bar{N}_{k|a}(\ell)$ (with $\Theta(k,a,\ell)$ the **TC** kernel), while the second one corresponds to random edges contribution. This last term can be decomposed into the contribution of the source node chosen with probability proportional to its activity,

$$\frac{1-q}{N(\ell)}\left(\frac{a}{\langle a\rangle}\right)\left(\bar{N}_{k-1|a} - \bar{N}_{k|a}\right) \tag{5}$$

which is added to the contribution of the random connected target node, given by

$$\frac{1-q}{N(\ell)}\left(\bar{N}_{k-1|a} - \bar{N}_{k|a}\right). \tag{6}$$

Continuing with our description of equation (4), the third term in its right-hand-side is the contribution of the remaining end of the edge added together with every a-activity new node (also tied uniformly at random). Finally, the last term $\gamma\delta_{k1}$ comes from the

initial degree ($k = 1$) of each a-activity new node itself. The **TC** kernel $\Theta(k, a, \ell)$ in (4) is defined as:

$$\Theta(k, a, \ell) = \frac{a}{\langle a \rangle N(\ell)} + \frac{k}{2L(\ell)}. \tag{7}$$

Every **TC** edge is tied to a source node chosen at random with probability proportional to its activity rate a,leading to the first term in right-hand-side of (7), whereas the last one represents the preferential attachment term arising from **TC** target [31, 32]. After regrouping terms, equation (4) can be rewritten as

$$\frac{\mathrm{d}\bar{N}_{k|a}}{\mathrm{d}\ell} = \bar{N}_{k-1|a}\Phi_{\gamma,q}(k-1, a, \ell) - \bar{N}_{k|a}\Phi_{\gamma,q}(k, a, \ell) + \gamma\delta_{k1} \tag{8}$$

being $\Phi_{\gamma,q}(k, a, \ell)$ the generalized connectivity kernel given by

$$\Phi_{\gamma,q}(k, a, \ell) = \frac{1}{N(\ell)}\left(\frac{a}{\langle a \rangle} + 1 - q + \gamma\right) + \frac{k}{2L(\ell)}q. \tag{9}$$

Finally, the resulting expression for the mean population of nodes with degree k ($\bar{N}_k(\ell)$) comprising all possible activity rates with pdf $F(a)$ is given by

$$\bar{N}_k(\ell) = \int_\Omega F(a)\bar{N}_{k|a}(\ell)\mathrm{d}a \tag{10}$$

being Ω the domain of $F(a)$. Now we can formally define the degree distribution function for a given quantity of aggregated edges ℓ as $P_\ell(k) = \bar{N}_k(\ell)/N(\ell)$.

GSG allows a broad flexibility in both activity distribution and population growth regimes. In the next sections, we will solve (8) under constant population ($\gamma = 0$) and population growth ($\gamma > 0$) regimes, in order to bring out a detailed analysis of $P_\ell(k)$ in these cases. On the other hand, we will focus on two paradigmatic cases for activity pdf: (i) constant activity ($F(a) = \delta(a - a_0)$) as the most frequent assumption and (ii) power-law activity pdf ($F(a) \propto a^{-\xi}$) recently found in some real social networks [38].

4. Constant population

In the particular case of constant population, network growth takes place only through the addition of new edges between existing nodes. Thus, the coupled system of ordinary differential equations governing the evolution of $\bar{N}_{k|a}(\ell)$ can be obtained by substituting $\gamma = 0$ in (8) that, after replacing $\Phi_{0,q}(k, a, \ell)$, reads

$$\frac{\mathrm{d}\bar{N}_{k|a}}{\mathrm{d}\ell} = \bar{N}_{k-1|a}\left[\frac{1}{N_0}\left(\frac{a}{\langle a \rangle} + 1 - q\right) + \frac{(k-1)q}{2L(\ell)}\right] - \bar{N}_{k|a}\left[\frac{1}{N_0}\left(\frac{a}{\langle a \rangle} + 1 - q\right) + \frac{kq}{2L(\ell)}\right] \tag{11}$$

where $N(\ell) = N_0\ \forall \ell \geqslant 0$ and $L(\ell) = L_0 + \ell$. There is a natural constraint imposed to k in (11), that is $k \leqslant (N_0 - 1)$. Moreover, the asymptotic solution to (11) adopt the trivial form $\bar{N}_{k|a} = N_0\delta_{k,(N_0-1)}\ \forall a \geqslant 0$. However, here we are interested only in the non-trivial transient solution.

Equation (11) can be solved in general by means of an iterative scheme as follows:

$$\bar{N}_{k|a}(\ell) = \left(\bar{N}_{k|a}(0) + \int_0^\ell \frac{\bar{N}_{k-1|a}(\ell')}{\Pi_{k|a}(\ell')}\Phi_{0,q}(k-1, a, \ell')\,\mathrm{d}\ell'\right)\Pi_{k|a}(\ell) \tag{12}$$

where $\Pi_{k|a}(\ell) = A\,(L_0+\ell)^{-kq/2}\exp\left(-\left(a/\langle a\rangle + 1 - q\right)\ell/N_0\right)$ is solution of

$$\frac{\mathrm{d}\Pi_{k|a}(\ell)}{\mathrm{d}\ell} = -\Pi_{k|a}(\ell)\Phi_{0,q}(k,a,\ell). \tag{13}$$

Closed form solutions to (11) are only reached in some particular cases. For instance, it can be easily shown that if $q = 0$, the solution to (11) is

$$\bar{N}_{k|a}(\ell) = N_0 \times \mathrm{Pois}\left(k;\lambda = \ell(a+\langle a\rangle)/(\langle a\rangle N_0)\right), \tag{14}$$

where $\mathrm{Pois}(k;\lambda) = (\lambda^k/k!)\exp(-\lambda)$ is the Poisson distribution with mean λ. Replacing (14) in (10), we can recover the following asymptotic form for the degree distribution

$$P_\ell(k) = \frac{\bar{N}_k(\ell)}{N_0} \sim \frac{1}{\ell}F\left(\frac{N_0}{\ell}k - 1\right), \tag{15}$$

in accordance with previous results for **AD** model (see [24] for a detailed derivation of (15) for **AD** model). Nevertheless, we will acquire some insight about exact behaviour of $\bar{N}_{k|a}$ by analyzing approximate solutions to (11) under extreme conditions. To this end, we assume the condition

$$\textbf{a.}\quad \frac{1}{N_0}\left(\frac{a}{\langle a\rangle} + 1 - q\right) \gg q\frac{k}{2(L_0+\ell)}. \tag{16}$$

By virtue of condition **a.**, the approximate solution to (11) results

$$\bar{N}_{k|a}(\ell) \sim \frac{1}{k!}\left(\frac{a/\langle a\rangle + 1 - q}{N_0}\,\ell\right)^k \mathrm{e}^{-\frac{a/\langle a\rangle+1-q}{N_0}\ell} \tag{17}$$

i.e. a Poisson distribution with mean $\lambda = \ell(a/\langle a\rangle + 1 - q)/N_0$.

On the other hand, the opposite condition to (16) corresponds to assume

$$\textbf{b.}\quad \frac{1}{N_0}\left(\frac{a}{\langle a\rangle} + 1 - q\right) \ll q\frac{k}{2(L_0+\ell)}, \tag{18}$$

yielding another approximate extreme solution to (11) satisfying

$$\bar{N}_{k|a}(\ell) \sim (L_0+\ell)^{-q/2}\left[1 - \left(\frac{L_0+\ell}{L_0}\right)^{-q/2}\right]^{k-1} \approx (L_0+\ell)^{-q/2}\,\mathrm{e}^{-(k-1)\left(\frac{L_0+\ell}{L_0}\right)^{-q/2}}, \tag{19}$$

showing a clear exponential decay with independence of the activity rate a. We shall see that another activity-independent solution is obtained again for large-k values under population growth regime.

Finally, $\bar{N}_k(\ell)$ is obtained by performing the integral of (10) between $\bar{N}_{k|a}(\ell)$ for constant population and the activity pdf $F(a)$.

We perform numerical simulations of **GSG** in time-aggregated representation for constant population with $N = 10^5$ nodes,together with a numerical integration of equations (11) and (10), in order to obtain the corresponding degree distribution $P_\ell(k) = \bar{N}_k(\ell)/N(\ell)$. We analyze two particular activity regimes: (i) constant activity $F(a) = \delta(a - a_0)$ and (ii) power-law activity pdf $F(a) \propto a^{-1.5}$. We show the very good agreement between simulations and theoretical predictions for constant population in figure 2. For constant and homogeneous activity, figure 2(a) shows the expected behaviour for $P(k)$, i.e. Poisson-like for small-k with a marked exponential decay when condition (18) is satisfied (see inset in figure 2(a)). Under power-law activity pdf, $P(k)$ is dominated by $F(a)$ power-law decay for mid-range k-values as shown in figure 2(b), while the limit cases are similar to those of the previous scenario.

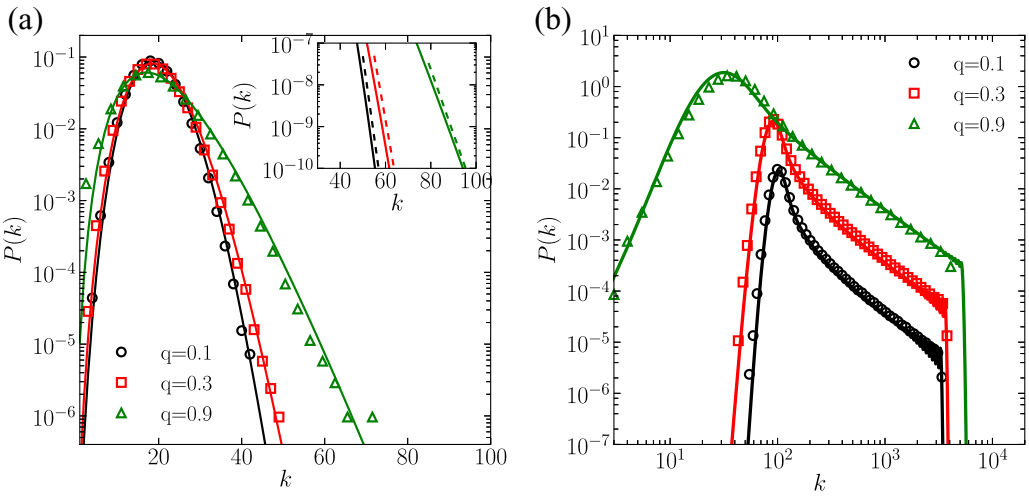

Figure 2. Degree distribution function $P_\ell(k) = \bar{N}_k(\ell)/N_0$ under constant population regime. (a) Activity pdf $F(a) = \delta(a - a_0)$ on time-aggregated networks with $N = 10^5$ and $\langle k\rangle = 20$ and (b) Activity pdf $F(a) \propto a^{-1.5}$ on networks with $N = 10^5$ and $\langle k\rangle = 200$. In both cases, symbols correspond to averages over 100 numerical simulations for: $q = 0.1$ (black circles), $q = 0.3$ (red squares) and $q = 0.9$ (green triangles). For the sake of clarity, the plots for different q values have been shifted in all cases. Solid lines correspond to numerical solutions to (12) subsequently integrated in (10), in order to obtain $P_\ell(k) = \bar{N}_k(\ell)/N(\ell)$. (Inset in ($a$)) Enlarged detail of high-k values behaviour in order to compare the exponential decay constant for exact solutions (solid lines) with those corresponding to the approximate formulation of (19) (dashed lines).

5. Population growth

Now we study the population growth regime considering $\gamma > 0$ in equation (8). This case is very relevant because growth constitutes one of the fundamental assumptions to obtain scale-free networks from preferential attachment mechanism. We will show here that this feature is also present for **GSG** mechanisms in time-aggregated representation under population growth.

Regrouping terms in (8), we can rewrite it now for $\gamma > 0$ as

$$\frac{\mathrm{d}\bar{N}_{k|a}}{\mathrm{d}\ell} = \bar{N}_{k-1|a}\left[\frac{1}{N(\ell)}\left(\frac{a}{\langle a\rangle} + 1 - q + \gamma\right) + \frac{(k-1)q}{2L(\ell)}\right] - \bar{N}_{k|a}\left[\frac{1}{N(\ell)}\left(\frac{a}{\langle a\rangle} + 1 - q + \gamma\right) + \frac{kq}{2L(\ell)}\right] + \gamma\delta_{k1} \tag{20}$$

where now $N(\ell) = N_0 + \gamma\ell$ and $L(\ell) = L_0 + (1+\gamma)\ell$. Unlike the previous case, now is possible to obtain non-trivial asymptotic solutions to (20). These kinds of solutions have the general form $\bar{N}_{k|a}(\ell) = n_{k|a}\ell$, with $n_{k|a}$ an unknown function of degree k and activity rate a [12]. Then, solving equation (20) for $n_{k|a}$ under asymptotic condition ($\ell \to \infty$), we obtain

$$n_{k|a} = n_{1|a}\prod_{\mathtt{j}=1}^{k-1}\frac{2(\gamma+1)(a/\langle a\rangle + 1 - q + \gamma) + q\gamma\,\mathtt{j}}{2(\gamma+1)\,(a/\langle a\rangle + 1 - q + 2\gamma) + q\gamma\,(\mathtt{j}+1)} \tag{21}$$

where $n_{1|a}$ is the solution for $k = 1$ given by

$$n_{1|a} = \frac{2\gamma^2(\gamma+1)}{2(\gamma+1)\left(a/\langle a\rangle + 1 - q + 2\gamma\right) + q\gamma}. \tag{22}$$

Beyond its rigorous expression, $n_{k|a}$ takes very simple forms under approximate scenarios. The idea behind these approximations is to obtain a simplified framework where solutions have a more evident meaning than under its exact form.

The first approximate scenario corresponds to neglect preferential attachment terms in (20), assuming that

$$\textbf{a.} \quad \left(\frac{a}{\langle a\rangle} + 1 - q + \gamma\right)\langle k\rangle_\ell \gg qk \tag{23}$$

being $\langle k\rangle_\ell = 2L(\ell)/N(\ell)$ the mean degree after ℓ aggregated edges.

Introducing approximation **a.** into (20) and substituting again $\bar{N}_{k|a}(\ell) = n_{k|a}\ell$ we obtain the asymptotic solution under this approximate framework,

$$n_{k|a} \sim \left(1 + \frac{\gamma}{a/\langle a\rangle + 1 - q + \gamma}\right)^{-(k-1)} \tag{24}$$

which shows a pure exponential decay. Let us now analyze the alternative extreme condition

$$\textbf{b.} \quad \left(\frac{a}{\langle a\rangle} + 1 - q + \gamma\right)\langle k\rangle_\ell \ll qk. \tag{25}$$

This last condition is satisfied when k and q are large enough, thus (21) takes the form

$$n_{k|a} \sim k^{-1-\frac{2(\gamma+1)}{q}}. \tag{26}$$

showing that $n_{k|a}$ has power-law behaviour with exponent $\alpha = 1 + \frac{2(\gamma+1)}{q}$, resulting $\alpha > 3$ when $0 < q \leqslant 1$ and $\gamma > 0$. An important fact is that $n_{k|a}$ becomes absolutely independent of a. Accordingly, $\bar{N}_k(\ell)$ from (10) adopt the same asymptotic power-law behaviour of $n_{k|a}$,

$$\bar{N}_k \sim k^{-1-\frac{2(\gamma+1)}{q}}. \tag{27}$$

As a consequence of (27), the asymptotic large-k behaviour of $\bar{N}_k$ is sustained only on **TC** mechanism with total disregard of any particular activity distribution. Once again, the combination of population growth with a preferential attachment term (in this case coming from **TC** mechanism) gives place to power-law behaviour, just as in the original **PA** model [11]. Unfortunately, this kind of behaviour seems to be hard to detect in real social networks because condition (25) is only fulfilled for very large values of k, which are rarely achieved or are perturbed by finite size effects.

We perform again extensive numerical simulations for **GSG** model in time-aggregated representation, but this time with population growth. In figure 3, we show the good agreement between the results of numerical simulations and theoretical predictions from (21) for the particular case of constant activity rate. Extreme solutions corresponding to both conditions analyzed in the text are represented in figures 3(a) and (b) by shifted dashed lines. A very good agreement between theory and simulations is shown again in figure 4, now for power-law $F(a)$.

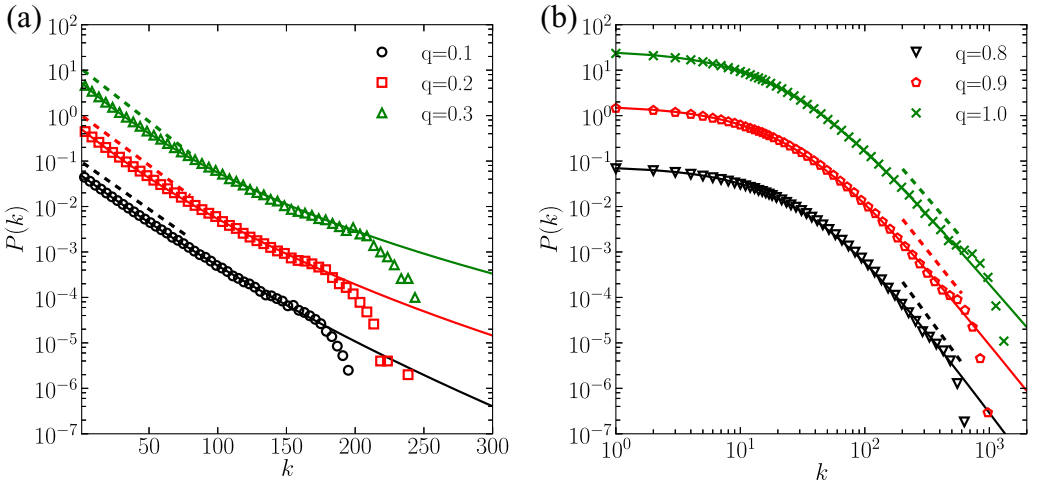

Figure 3. Degree distribution function $P_\ell(k) = \bar{N}_k(\ell)/N(\ell)$ under population growth regime and $F(a) = \delta(a - a_0)$ for time-aggregated networks with final population $N = 10^5$ and $\langle k \rangle = 20$. Symbols correspond to averages over 100 numerical simulations for: (a) $q \in \{0.1, 0.2, 0.3\}$ and (b) $q \in \{0.8, 0.9, 1.0\}$. Solid lines correspond to numerical solutions of (20) properly normalized in order to obtain $P(k)$. Asymptotic extreme solutions of (24) and (26) are plotted in shifted dashed lines. For the sake of clarity, the plots for different q values have been shifted.

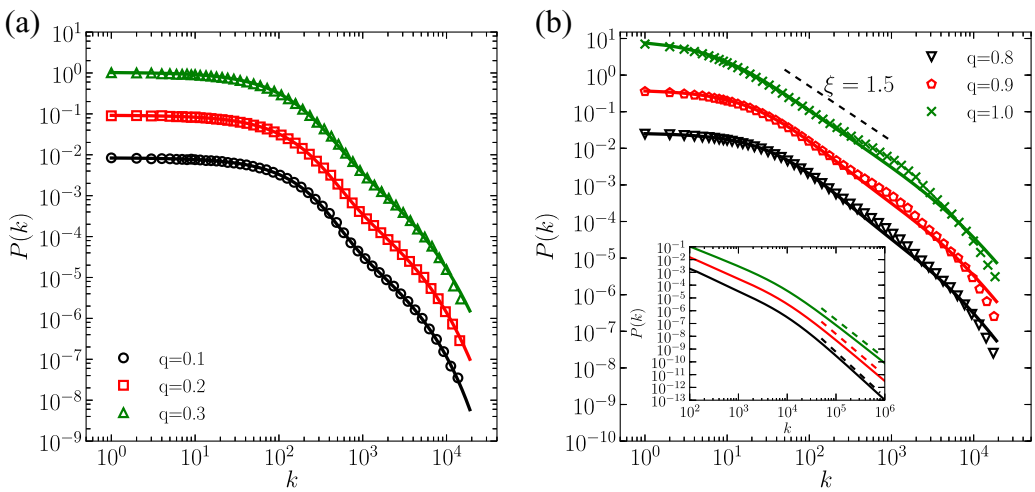

Figure 4. Degree distribution function $P_\ell(k) = \bar{N}(\ell)_k/N(\ell)$ under population growth regime and $F(a) \propto a^{-1.5}$ for time-aggregated networks with final population $N = 10^5$ and $\langle k \rangle = 200$. Symbols correspond to averages over 100 numerical simulations for: (a) $q \in \{0.1, 0.2, 0.3\}$, and (b) $q \in \{0.8, 0.9, 1.0\}$. Solid lines correspond to numerical solutions of (20) averaged through (10) and properly normalized in order to obtain $P(k)$. Asymptotic extreme solutions of (24) and (26) are plotted in shifted dashed lines. (Inset in (b)) Enlarged detail of higher k behaviour in order to compare the power-law decay for exact solutions (solid lines) with those corresponding to the approximate formulation of (27) (dashed lines). For the sake of clarity, the plots for different q values have been shifted.

6. Degree-degree correlation and clustering

6.1. Degree-degree correlation

As we have said before, social networks are usually characterized by strong degree-degree correlation and strong transitivity (i.e. high probability that the friends of my friends are also my friends). An indirect measure of degree-degree correlation can be obtained from the average degree of the neighbours of the nodes with degree k, $\bar{k}_{nn}(k)$, formally defined as [39]:

$$\bar{k}_{nn}(k) = \sum_{k'} k' P(k'|k) \tag{28}$$

where $P(k'|k)$ is the conditional probability that a node of degree k is connected to another of degree k'. Thus, when $\bar{k}_{nn}(k)$ grows with k, we say the network has *degree assortativity*. **GSG** model yields networks with degree assortativity under population growth regime, both for constant and power-law activity pdf $F(a)$, as shown in figures 5(b) and (d).

In contrast, networks yielded by **GSG** model under constant population and constant activity pdf $F(a) = \delta(a - a_0)$, show a flat plot for $\bar{k}_{nn}(k)$, as would be expected in the case of Erdös-Rényi (ER) networks (random networks with Poisson degree distribution) for which $\bar{k}_{nn}(k') = \langle k \rangle \ \forall k'$ (see figure 5(a)). The model show a very mild disassortative behaviour also under constant population but now for power-law activity pdf $F(a) \propto a^{-1.5}$ (see figure 5(c)).

The $\bar{k}_{nn}(k)$ for the time-aggregated network yielded by **AD** model has been thoroughly studied in [24]. There, the authors propose a mapping to a hidden-variable network model in order to obtain analytical expressions for some topological observables, in particular for $\bar{k}_{nn}(k)$. Here, we will follow the same path for the **GSG** model in the case of $q = 0$ and $\gamma = 0$, where it matches the conditions of the AD model. In this case, the following scaling function can be derived for the rescaled average degree of neighbours of nodes with degree k (see the appendix for a derivation),

$$\frac{2\langle a \rangle}{\langle k \rangle}\left(\bar{k}_{nn}(k) - 1\right) - 2\langle a \rangle \simeq (\langle a^2 \rangle - \langle a \rangle^2)\left(\frac{2\langle a \rangle}{\langle k \rangle}k\right)^{-1}. \tag{29}$$

In figure 6(a), we show the collapse predicted by the scaling law of (29) for solutions corresponding to different values of $\langle k \rangle$, with $q = 0$ and $\gamma = 0$. An equivalent expression of this scaling function for $q > 0$ is hard to be derived because of the local correlations induced by the **TC** mechanism. Nevertheless, in figure 6(b) we have plotted $(2\langle a \rangle/\langle k \rangle)\left(\bar{k}_{nn}(k) - 1\right) - 2\langle a \rangle$ as a function of $2\langle a \rangle k/\langle k \rangle$ for $q > 0$, showing that if this scaling function exists (as the inset in figure 6(a) for $q = 0.5$ suggest), it will have a dependence with q such that $(2\langle a \rangle/\langle k \rangle)\left(\bar{k}_{nn}(k) - 1\right) - 2\langle a \rangle \simeq F^q_{\text{scal}}(2\langle a \rangle k/\langle k \rangle)$.

6.2. Clustering coefficient

Another practical measure associated with transitivity can be given by the average clustering coefficient as a function of the degree k, $\bar{C}(k)$, defined as:

$$\bar{C}(k) = \frac{1}{N_k} \sum_{i \in \text{Deg}(k)} C_i, \tag{30}$$

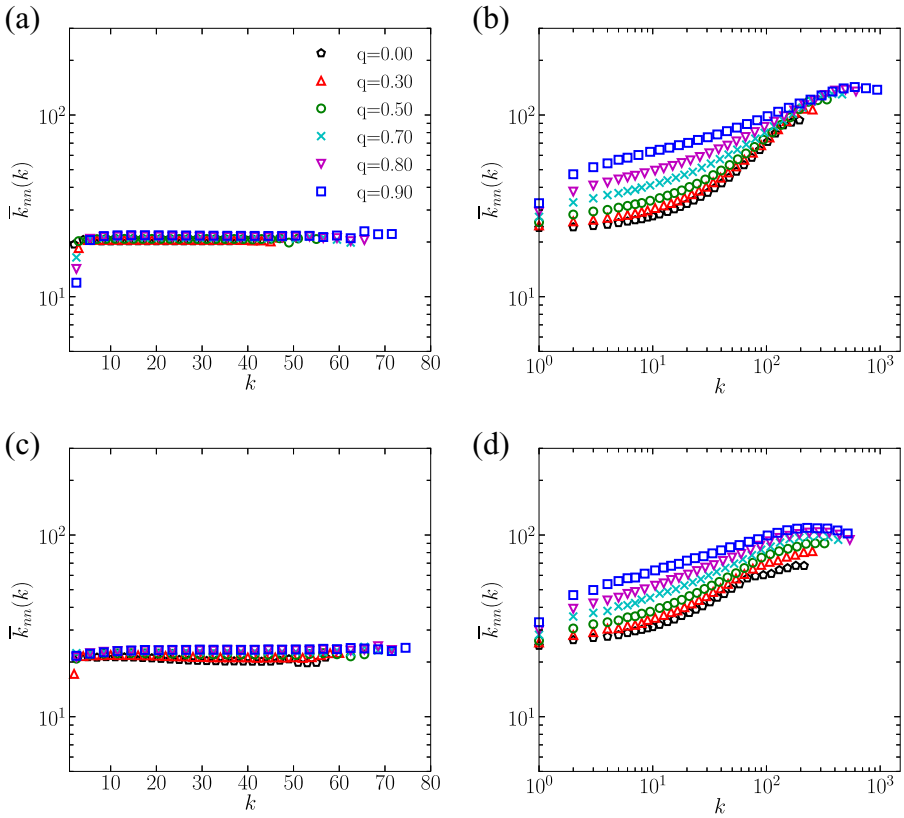

Figure 5. Average degree of the neighbours of degree k, $\bar{k}_{nn}(k)$, for networks obtained by **GSG** model, with final population $N = 10^5$ and $\langle k \rangle = 20$ in (a) and (b) while $N = 10^4$ and $\langle k \rangle = 20$ in (c) and (d). Symbols correspond to averages over 100 numerical simulations under the following conditions: (a) constant population and constant activity pdf $F(a) = \delta(a - a_0)$, (b) population growth and constant activity pdf $F(a) = \delta(a - a_0)$ (c) constant population and power-law activity pdf $F(a) \propto a^{-1.5}$, (d) population growth and power-law activity pdf $F(a) \propto a^{-1.5}$.

where $\mathrm{Deg}(k)$ is the set of all nodes of degree k, with N_k its cardinal. In (30), C_i represent the local clustering for node i, defined as the fraction of edges between neighbours of node i relative to its maximum number $k_i(k_i - 1)/2$ and reads

$$C_i = \sum_{j,k \in N_{nn}(i)} \frac{a_{jk}}{k_i(k_i - 1)}, \tag{31}$$

being $N_{nn}(i)$ the set of neighbours of node i and a_{jk} the elements of adjacency matrix $\mathbf{A}$, such that $a_{jk} = 1$ ($a_{jk} = 0$) if there is (not) an edge between nodes i and j. Clustering C_i is related to the probability of triangles occurrence with node i as one of its vertices.

Real social networks usually exhibit a scaling law for the average clustering as a function of the degree k

$$\bar{C}(k) \sim k^{-\beta} \tag{32}$$

where the observed exponents meet $\beta \lesssim 1$ [28].

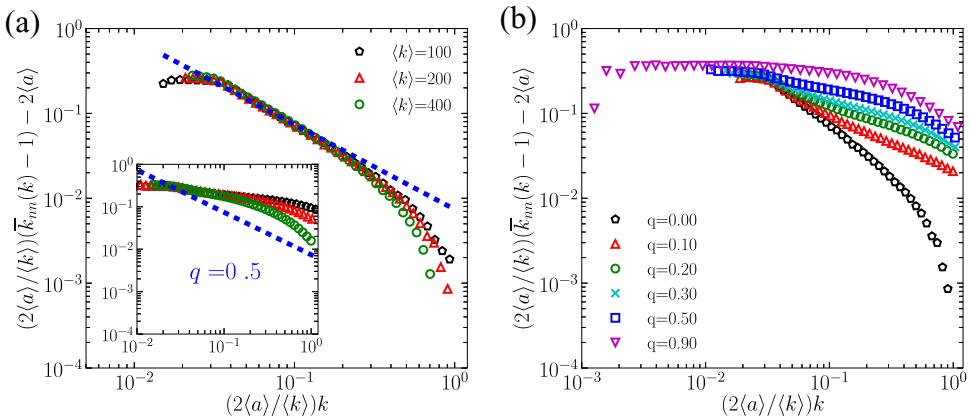

Figure 6. Rescaled average degree of the neighbours of degree k, for time-aggregated networks obtained by **GSG** model with power-law activity pdf $F(a) \propto a^{-1.5}$ and constant population $N = 10^4$. Symbols correspond to averages over 100 numerical simulations. (a) Collapse of solutions for $q = 0$ and $\langle k \rangle = 100$, 200 and 400 (the inset corresponds to $q = 0.5$). The dashed line represents the prediction of the scaling function of (29). (b) Rescaled $\bar{k}_{nn}(k)$ for different values of $q \geqslant 0$.

This scaling law behaviour is also captured by **GSG** model as shown in figure 7, where we have plotted,in all cases, a dashed line with slope -1 in log–log scale as reference. This fact also shows that time-aggregated networks obtained by **GSG** model under constant population are far from being ER random networks as figures 5(a) and (c) might have suggested. If that had been the case, $\bar{C}(k)$ would be independent of node degree k, as can be seen from its exact expression for random networks [40]:

$$\bar{C}(k) = \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{N \langle k \rangle^3}. \tag{33}$$

Finally, from the total average clustering coefficient definition:

$$\bar{C} = \frac{1}{N} \sum_{i=1}^{N} C_i = \sum_{k} P(k) \bar{C}(k), \tag{34}$$

we can confirm the expected growing nature of $\bar{C}$ with the **TC** probability q, as shown in figure 8.

7. Study cases

In order to illustrate the correspondence with **GSG** model, we analyze two experimental datasets reflecting human relationships networks in very different context: face-to-face encounters in a closed gathering and friendship relations in an online social network. For the first case, individual face-to-face contacts are detected with a time resolution of 20 s and within a distance of $\sim$1 m, through wearable active radio-frequency identification devices (RFID) placed on the chest of participants [16]. Here we analyze the publicly

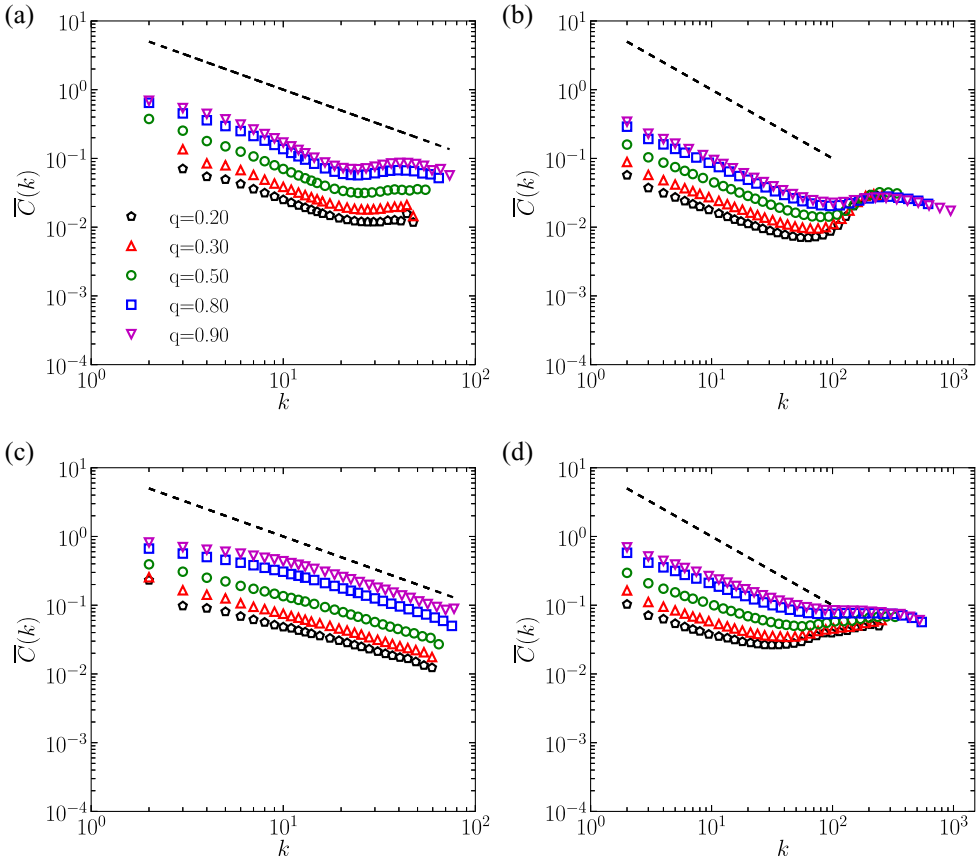

Figure 7. Average local clustering $\bar{C}(k)$ as a function of the degree k for time-aggregated networks obtained by **GSG** model, with final population $N = 10^5$ and $\langle k \rangle = 20$ in (a) and (b) while $N = 10^4$ and $\langle k \rangle = 20$ in (c) and (d). Symbols correspond to averages over 100 numerical simulations under the following conditions: (a) constant population and constant activity pdf $F(a) = \delta(a - a_0)$, (b) population growth and constant activity pdf $F(a) = \delta(a - a_0)$ (c) constant population and power-law activity pdf $F(a) \propto a^{-1.5}$, (d) population growth and power-law activity pdf $F(a) \propto a^{-1.5}$. Dashed lines corresponding to the scaling-law $\bar{C}(k) \sim k^{-1}$ are plotted as reference.

available dataset for ACM Hypertext 2009 (**HT**) conference held in Turin, Italy, with $N = 113$ nodes and $L = 2196$ unweighted edges [17,18] (we only consider time-aggregated face-to-face contacts between participants along the first meeting day). The second case corresponds to a Facebook subgraph (**FG**) comprising $N = 63731$ users from New Orleans (with larger connected component of size $N_{cc} = 63392 \approx 0.995 \times N$), interconnected by $L = 817090$ undirected edges[2] representing friendship relations between users, as collected in [15]. As in the previous case, this dataset also provides time-resolved information through the birth times of new edges. In contrast to **HT FG** exhibit population growth due to the introduction of new users, in a context without spatial constraints.

[2] Edges on Facebook can be considered as undirected because friendship requests must be explicitly accepted by the other party.

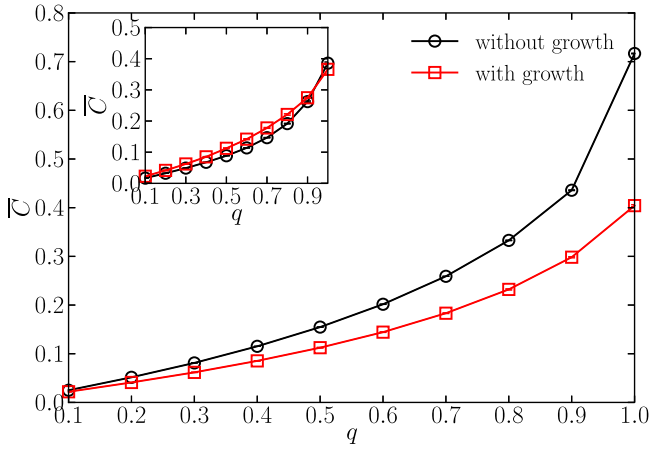

Figure 8. Average clustering coefficient $\bar{C}$ in terms of **TC** probability. All results correspond to averages over 100 realizations of **GSG** with $N = 10^4$, $\langle k \rangle = 20$ and $F(a) \propto a^{-1.5}$. Inset: results for **GSG** model with the previous parameters but constant activity $F(a) = \delta(a - a_0)$.

7.1. Topological properties

As can be seen in figures figure 9(a), $P(k)$ for **HT** results narrow and short-tailed with small-k Poissonian-like behaviour (see figure 9(a)) as obtained for **GSG** model under constant population with the same N and $\langle k \rangle$ (see figure 9(b)). Instead, **FG** network exhibit a heavy-tailed degree distribution with large-k power-law behaviour $P(k) \sim k^{-\alpha}$ with $\alpha \approx 3.4$ as shown figure 10(a), also compatible with a network generated by **GSG** model but now under population growth regime (see figure 10(b)). Additionally, the average clustering $\bar{C}(k)$ and the neighbours average degree $\bar{k}_{nn}(k)$ for **HT** and **FG** also exhibit the same qualitative behaviour of those generated by **GSG** model under constant and population growth regimes, respectively, as shown in figures 9(c)–(e) and figures 10(c)–(e).

In figures 9(b), (d) and (f) we present a semi-qualitative comparison with $P(k)$, $\bar{C}(k)$ and $\bar{k}_{nn}(k)$, respectively, for a time-aggregated network obtained by the **GSG** model with power-law activity. For the comparison with the **HT** network, we assume a constant population with $N = 113$ and $L = 2196$ ($\langle k \rangle = 38.9$), i.e. the values obtained from the original dataset. The triadic-closure probability was fixed in $q = 0.8$ for both comparisons with **HT** and **FG** an appropriate value considering the analysis presented in section 7.2. Thus, the exponent ξ of $F(a) \sim a^{-\xi}$ becomes the only free parameter. We have performed runs of **GSG** model for $\xi \in [0.5, 3]$ with variable step $\delta\xi = 0.01 - 0.1$ in order to obtain the best fit (MLE method) of the degree distribution $P(k)$. We proceeded in the same way with the comparison of figure 10, but now assuming a growing population with $N = 63731$ and $L = 817090$ as final values. In the latter case, the selected growth rate was $\gamma = N/L$ with an initial population $N_i = 100$.

7.2. Growth pattern: Triadic closure

Both **HT** and **FG** provide time-resolved information through the birth times of added edges. By virtue of this particular feature, we can address the problem of edges growth

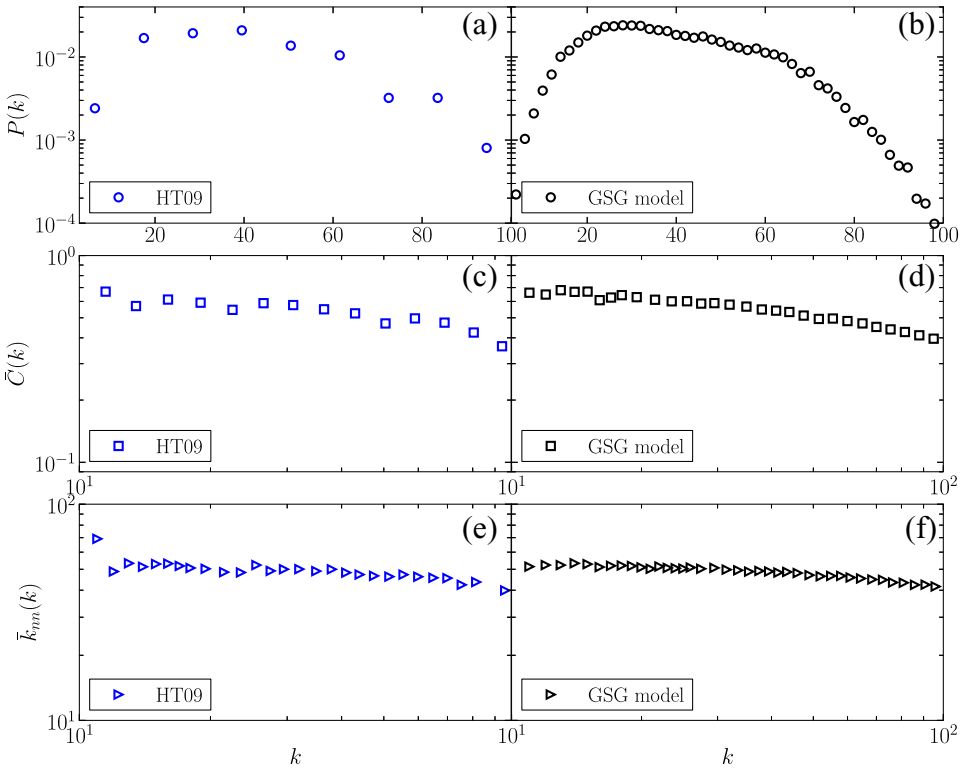

Figure 9. Degree distribution $P(k)$, average clustering $\bar{C}(k)$ and neighbours average degree $\bar{k}_{nn}(k)$ for (a),(c) and (e) **HT** time-aggregated contacts network and (b),(d) and (f) **GSG** model under constant population with $N = 113$, $\langle k \rangle = 38.9$, **TC** probability $q = 0.8$ and power-law activity pdf with exponent $\xi = 0.79$.

mechanism. Let $\mathbf{l_t} = (i, j)_t$ define an edge between nodes $i = \mathbf{l_t}(1)$ and $j = \mathbf{l_t}(2)$ emerged at time t and let $d(\mathbf{l_t}, t)$ be the distance between nodes $\mathbf{l_t}(1)$ and $\mathbf{l_t}(2)$ at time t, immediately before the occurrence of $\mathbf{l_t}$. Those edges $\mathbf{l_t}$ for which $d_t(\mathbf{l_t}, t) = 2$, called *transitive edges*, are the product of a **TC** mechanism (or *cyclic closure* for $d(\mathbf{l_t}, 2) > 2$ [25]).

Then, we record $d(\mathbf{l_t}, t)$ for each edge and define $N_d(T)$ as the aggregated number of edges $\mathbf{l_t}$ with $d(\mathbf{l_t}, t) = d$ for $t < T$. Let $d(\mathbf{l_t}, t) = 0$ when node $\mathbf{l_t}(1)$ or $\mathbf{l_t}(2)$ is a newcomer and $d(\mathbf{l_t}, t) = \infty$ when there is no path between $\mathbf{l_t}(1)$ and $\mathbf{l_t}(2)$ previous to $\mathbf{l_t}$. Obviously $N_{d=1}(T) = 0$ because multiple edges between nodes are forbidden.

The distance distribution $P_T(d)$ is formally defined as

$$P_T(d) = \lim_{L \to \infty} \frac{N_d(T)}{L} \tag{35}$$

with $L = \sum_{i \in \mathbb{N}_0} N_i(T)$ the total number of edges.

Clearly, $P_T(d)$ depends on the temporal ordering of $\{\mathbf{l_t}\}_{t \leqslant T}$ and this is why it provides valuable information about edges growth mechanism. In order to verify this statement, we compare $P_T(d)$ for the actual edges succession $\{\mathbf{l_1}, \mathbf{l_2}, \ldots, \mathbf{l_T}\}$, with the average $\langle P_T(d) \rangle_{\text{rand}}$ over 100 random permutation $\{\mathbf{l}_{\sigma_t}\}_{t=1,\ldots,T}$, where $\sigma_t \in \text{Perm}(T)$ is a random permutation function belonging to the group of all permutation of T index $\text{Perm}(T)$. At this point, it is important to note that all succession $\{l_{\sigma_t}\}_{t=1,\ldots,T}$ gives exactly the same final time-aggregated network at time T.

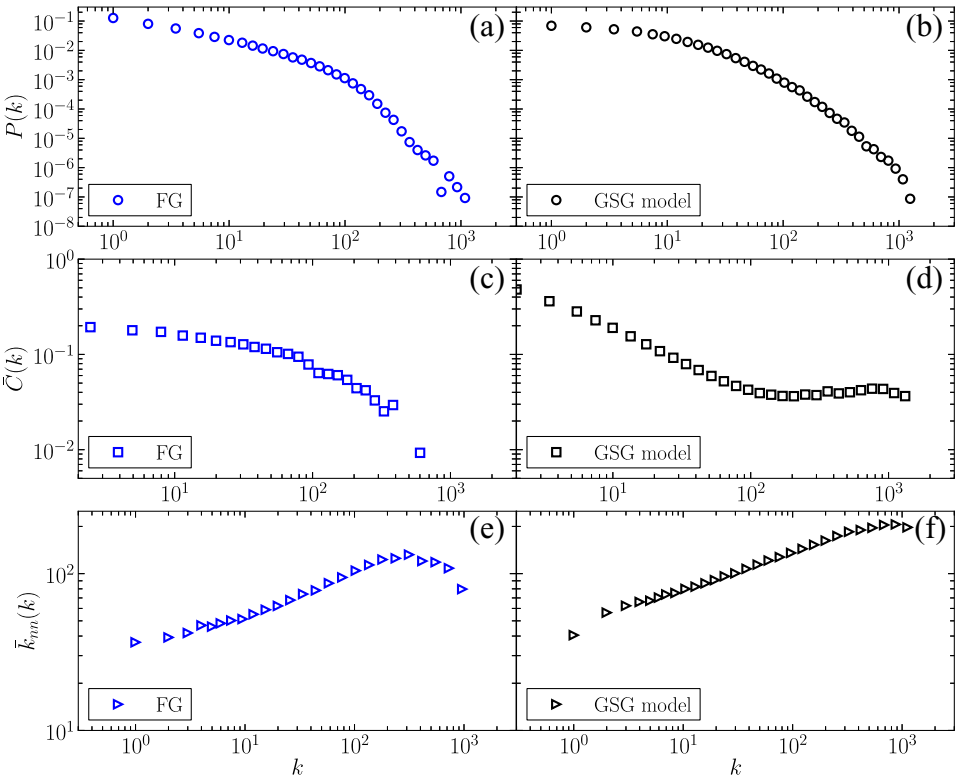

Figure 10. Degree distribution $P(k)$, average clustering $\bar{C}(k)$ and neighbours average degree $\bar{k}_{nn}(k)$ for (a),(c) and (e) **FG** time-aggregated contacts network and (b),(d) and (f) **GSG** model under population growth with $N = 63731$, $\langle k \rangle = 25.6$, **TC** probability $q = 0.8$ and power-law activity pdf with exponent $\xi = 1.5$.

Statistically significant differences are observed between $P_T(d)$ and $\langle P_T(d) \rangle_{\text{rand}}$ for both **HT** and **FG** as shown in figure 11. In particular, both plots in figure 11 show $P_T(d) < \langle P_T(d) \rangle_{\text{rand}}$ for $d > 2$ but $P_T(2) > \langle P_T(2) \rangle_{\text{rand}}$. The strong deviation for $d = 2$ can be quantified through the z-score value, in this case defined as

$$z = \frac{P_T(2) - \langle P_T(2) \rangle_{\text{rand}}}{\sigma_{\text{rand}}}, \tag{36}$$

resulting in $z_{\text{HT}} = 32$ and $z_{\text{FG}} = 185$. On the other hand, we have obtained a large fraction of transitive edges with $P_T(2) = 0.91(2)$ for **HT** and $P_T(2) = 0.77(4)$ for **FG**. These facts state a strong memory effect in the mechanism of edges growth with a significant predominance of transitive edges.

8. Summary and discussion

In this work, we have introduced a stochastic growth model (**GSG**) for dynamical social networks assuming agents with heterogeneous activity rate. As a distinctive characteristic, the model assumes non-Markovian agents performing connections through a mixed mechanism, including a **TC** connectivity process. As a main contribution, the

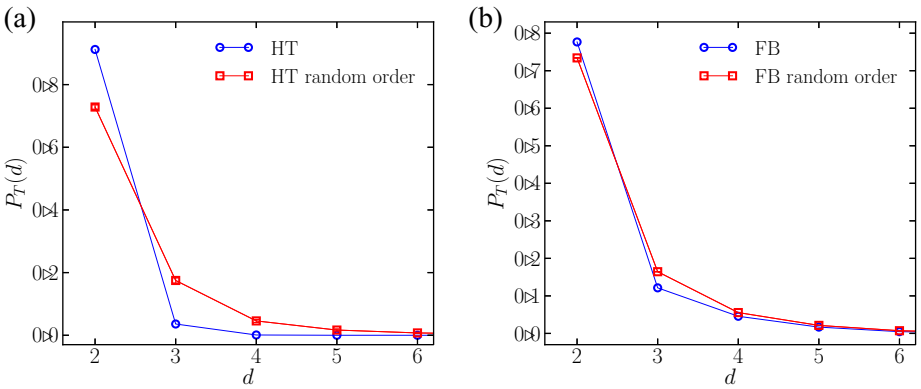

Figure 11. Distance probability distribution $P_T(d)$ as a function of distance d (blue circles) and its average calculated over 100 random permutations of the actual time-ordered edges succession (red squares) for: (a) **HT** face-to-face data set, and (b) **FG** online friendship relations subgraph. Standard error bars are smaller than symbols.

GSG model gives rise to a time-aggregated representation with the topological features expected for real social networks. In the particular case with $q = 0$ and $\gamma = 0$, **GSG** recover some of the topological properties of the time-aggregated networks for **AD** model [24].

We have obtained the degree distribution from a general analytical framework based on a rate equation approach, for both constant population and population growth regimes. In particular, we have shown explicitly that the **TC** mechanism not only allows one to increase the average clustering coefficient but also shapes the degree distribution. Additionally, we have traced a parallel with the hidden-variable approach performed in [24] for **AD** model.

Recently, Karsai *et al* [36] have shown that the addition of long-term memory in agents affect their contacts dynamic. In the same vein, we have shown that the agent's memory also leaves its mark on the topology of time-aggregated networks.

Acknowledgments

We thanks P Balenzuela and M Otero for helpful comments. ADM is grateful to Universidad de Buenos Aires for financial support through its postgraduate fellowship program. ADM and COD acknowledge partial financial support from Universidad de Buenos Aires through its project UBACYT 2012–2015.

Appendix A. A brief comparison with hidden-variable model

The class of models with hidden variables assume an intrinsic feature of the nodes that completely determines the topological properties of the network [23]. The model start with N disconnected nodes tagged with a hidden variable h from a probability distribution $\rho(h)$.

Then, the edges are added with a symmetric connection probability $r(h_i, h_j)$. Thus, both functions $\rho(h)$ and $r(h_i, h_j)$ fully determine the topological properties of a *Markovian random network*. These topological properties are obtained in terms of the hidden variable h, then a *propagator* $g(k|h)$ provides the transformation from h to the usual degree k. In fact, the propagator $g(k|h)$ represents the conditional probability that a node with a hidden variable h ends up with degree k. In this way, the degree distribution can be written as:

$$P(k) = \sum_h g(k|h)\rho(h). \tag{A.1}$$

It is also possible to obtain the average degree of the neighbours of a node with degree k as [23]

$$\bar{k}_{nn}(k) = 1 + \frac{1}{P(k)} \sum_h g(k|h)\rho(h)\bar{k}_{nn}(h), \tag{A.2}$$

where $\bar{k}_{nn}(h)$ is defined as

$$\bar{k}_{nn}(h) = \frac{N}{\bar{k}(h)} \sum_{h'} \rho(h')\bar{k}(h')r(h, h'), \tag{A.3}$$

with $\bar{k}(h')$ the average degree of a h'-node given by

$$\bar{k}(h) = N \sum_{h'} \rho(h')r(h, h'). \tag{A.4}$$

Finally, the average clustering as a function of the degree k is written as

$$\bar{C}(k) = \frac{1}{P(k)} \sum_h \rho(h)g(k|h)\bar{C}(h) \tag{A.5}$$

where $\bar{C}(h)$ is defined as

$$\bar{C}(h) = \sum_{h',h''} P(h'|h)r(h', h'')P(h''|h), \tag{A.6}$$

with $P(h'|h)$ the conditional probability of connection between a h-node with a h'-node, given by

$$P(h'|h) = \frac{N\rho(h')r(h, h')}{\bar{k}(h)}. \tag{A.7}$$

In the **GSG** model, the activity a plays the role of hidden variable, with the activity pdf $F(a)$ taking the place of $\rho(h)$ in a continuous formulation. However, the time-aggregated networks from **GSG** model are not a Markovian random network because of the local correlations induced by the **TC** mechanism (except for the case with $q = 0$). We have been able to obtain $\bar{N}_{k|a}$ as an equivalent to the propagator of the hidden-variable model for any $q \geqslant 0$. Nevertheless, in this case it is not possible, in general, to derive all the degree correlations only from the connection probability $r(a, a')$ of the hidden-variable model. This is because the introduction of the TC connection mechanism adds degree correlations beyond those imposed by the activity-driven model only. However, we can still completely describe the topological properties of the network from $F(a)$ and $r(a, a')$ if $q = 0$.

Finally, we will give a sketch for the derivation of the scaling function of (29). In this particular case, we have $q = 0$ and, for this reason, we are able to obtain an approximate

expression for the probability $r_\ell(i,j)$ that two nodes i and j with activities a_i and a_j, respectively, become connected after ℓ added edges. Let $s_\ell(i,j) = 1 - r_\ell(i,j)$ be the probability that no connection has been created between this two nodes up to ℓ added edges. Thus, we can write $s_\ell(i,j)$ as [24]

$$s_\ell(i,j) = \sum_{z_i,z_j} P_\ell(z_i)P_\ell(z_j)\left[\prod_{n=1}^{z_i}\left(1-\frac{1}{N-n}\right)\right]\left[\prod_{m=1}^{z_j}\left(1-\frac{1}{N-m}\right)\right], \quad \text{(A.8)}$$

with $P_\ell(z_i)$ ($P_\ell(z_j)$) the probability that z_i (z_j) connections have been emerged from i (j) after ℓ added edges, given by

$$P_\ell(z_i) = \binom{\ell}{z_i}\left(\frac{a_i}{N\langle a\rangle}\right)^{z_i}\left(1-\frac{a_i}{N\langle a\rangle}\right)^{\ell-z_i}, \quad \text{(A.9)}$$

being $a_i/(N\langle a\rangle)$ the probability to choose the node i with activity a_i from the population. Assuming $z_i, z_j \ll N$ in (A.8), we have

$$\prod_{n=1}^{z_{i,j}}\left(1-\frac{1}{N-n}\right) \simeq \left(1-\frac{1}{N}\right)^{z_{i,j}}. \quad \text{(A.10)}$$

Then, substituting (A.9) in (A.8) together with the approximation of (A.10), we obtain $r_\ell(a_i,a_j)$ as

$$r_\ell(a_i,a_j) = 1 - s_\ell(a_i,a_j) \simeq 1 - \left(1-\frac{a_i}{N^2\langle a\rangle}\right)^{\ell}\left(1-\frac{a_j}{N^2\langle a\rangle}\right)^{\ell}, \quad \text{(A.11)}$$

that can be approximated as

$$r_\ell(a_i,a_j) \simeq 1 - \exp\left(-\frac{\ell}{N^2\langle a\rangle}(a_i+a_j)\right). \quad \text{(A.12)}$$

Mapping $\rho(h) \to F(a)$, $r(h,h') \to r_\ell(a,a')$, $g(k|h) \to \bar{N}_{k|a}$ and replacing (A.12) in (A.3) and (A.4), we arrive to the scaling function of (29) by the same way as described in [24].

References

[1] Pastor-Satorras R and Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200–3
[2] Kuperman M and Abramson G 2001 *Phys. Rev. Lett.* **86** 2909–12
[3] Newman M E J 2002 *Phys. Rev.* E **66** 016128
[4] Liu Z and Hu B 2005 *Europhys. Lett.* **72** 315
[5] Miller J C 2009 *J. R. Soc. Interface* **6** 1121–34
[6] Volz E M, Miller J C, Galvani A P and Meyers L A 2011 *PLoS Comput. Biol.* **7** e1002042
[7] Newman M E J and Park J 2003 *Phys. Rev.* E **68** 036122
[8] Girvan M and Newman M E J 2002 *Proc. Natl Acad. Sci. USA* **99** 7821–6
[9] Medus A, Acuña G and Dorso C O 2005 *Physica* A **358** 593–604
[10] Newman M E J 2002 *Phys. Rev. Lett.* **89** 208701
[11] Barabási A L and Albert R 1999 *Science* **286** 509–12
[12] Krapivsky P L, Redner S and Leyvraz F 2000 *Phys. Rev. Lett.* **85** 4629–32
[13] Newman M E J 2001 *Proc. Natl Acad. Sci. USA* **98** 404–9
[14] Barabasi A L, Jeong H, Neda Z, Ravasz E, Schubert A and Vicsek T 2002 *Physica* A **311** 590–614
[15] Viswanath B, Mislove A, Cha M and Gummadi P K 2009 On the evolution of user interaction in Facebook *Proc. of the 2nd ACM Workshop on Online Social Networks (Barcelona, Spain, August 2009)* ed Crowcroft J *et al* (New York: ACM) pp 37–42

AQ1

[16] Cattuto C, den Broeck W V, Barrat A, Colizza V, Pinton J F and Vespignani A 2010 *PLoS One* **5** e11596
[17] Isella L, Stehlé J, Barrat A, Cattuto C, Pinton J F and den Broeck W V 2011 *J. Theor. Biol.* **271** 166–80
AQ2 [18] www.sociopatterns.org
[19] Perra N, Gonçalves B, Pastor-Satorras R and Vespignani A 2012 *Sci. Rep.* **2** 469
[20] Bianconi G and Barabási A L 2001 *Europhys. Lett.* **54** 436–42 (http://dx.doi.org/10.1209/epl/i2001-00260-6)
[21] Caldarelli G, Capocci A, De Los Rios P and Muñoz M A 2002 *Phys. Rev. Lett.* **89** 258702 (scale-free networks without growth—'good get richer')
[22] Dorogovtsev S N, Mendes J F F and Samukhin A N 2000 *Phys. Rev. Lett.* **85** 4633–6
[23] Boguna M and Pastor-Satorras R 2003 *Phys. Rev.* E **68** 036112
[24] Starnini M and Pastor-Satorras R 2013 *Phys. Rev.* E **87** 062807
[25] Kossinets G and Watts D J 2006 *Science* **311** 88–90
[26] Szell M and Thurner S 2010 *Soc. Netw.* **32** 313–29
[27] Szell M, Lambiotte R and Thurner S 2010 *Proc. Natl Acad. Sci. USA* **107** 13636–41
[28] Klimek P and Thurner S 2013 *New J. Phys.* **15** 063008
[29] Albert R and Barabasi A L 2000 *Phys. Rev. Lett.* **85** 5234
[30] Davidsen J, Ebel H and Bornholdt S 2002 *Phys. Rev. Lett.* **88** 128701
[31] Holme P and Kim B J 2002 *Phys. Rev.* E **65** 026107
[32] Vázquez A 2003 *Phys. Rev.* E **67** 056104
[33] Toivonen R, Onnela J P, Saramäki J, Hyvönen J and Kaski K 2006 *Physica* A **371** 851–60
[34] Moriano P and Finke J 2013 *J. Stat. Mech.* **6** P06010
[35] Liben-Nowell D and Kleinberg J 2007 *J. Am. Soc. Inform. Sci.* **58** 1019–31
[36] Karsai M, Perra N and Vespignani A 2014 *Sci. Rep.* **4** 4001
[37] Krapivsky P L, Rodgers G J and Redner S 2001 *Phys. Rev. Lett.* **86** 5401–4
[38] Muchnik L, Pei S, Parra L C, Reis S D S, Andrade J S, Havlin S and Makse H A 2013 *Sci. Rep.* **3** 1783
[39] Pastor-Satorras R, Vázquez A and Vespignani A 2001 *Phys. Rev. Lett.* **87** 258701
[40] Dorogovtsev S N 2004 *Phys. Rev.* E **69** 027104

QUERIES

Page 21
AQ1
Please check the details for any journal references that do not have a blue link as they may contain some incorrect information. Pale purple links are used for references to arXiv e-prints.

Page 22
AQ2
Please update the year and publication details if appropriate in reference [18].