Activity ageing in growing networks

Lambiotte, R.

Link to publication

Activity ageing in growing networks

R Lambiotte

Université de Liège, Sart-Tilman, B-4000 Liège, Belgium
E-mail: renaud.lambiotte@ulg.ac.be

Abstract. We present a model for growing information networks where the ageing of a node depends on the time at which it entered the network and on the last time it was cited. The model is shown to undergo a transition from a small-world to a large-world network. The degree distribution may exhibit very different shapes depending on the model parameters, e.g. delta-peaked, exponential or power-law tailed distributions.

J. Stat. Mech. (2007) P02020

Contents

1. Introduction

The ageing of nodes is an important process for understanding the way information or social networks grow [1]–[6]. For instance, this process may be responsible for deviations to scale-free degree distributions [1] or for the non-vanishing values of the clustering coefficient observed in many networks [7]–[9]. Ageing accounts for the fact that *old* nodes lose their ability to acquire new links as time goes on, thereby limiting the number of active nodes to a small fraction of the whole network. In general, this effect embodies the notion of generation for social agents, the lifetime of information or of an article, etc. Such effects may be taken into account by attributing an age τ to nodes [2] and by assuming that their probability of receiving a link from a newly entering node depends on their age (through some decreasing function of τ) and, possibly, on other parameters such as their degree k (preferential attachment [10,11]). An alternative model [7]–[9] assumes that nodes can be deactivated with a probability proportional to k^{-1}. In this deactivation model (DM), once a node is deactivated, it is excluded from the network dynamics. The DM is appealing because it mimics the fact that less popular nodes are more easily *forgotten* than the popular ones. This is the case for citation networks [13] (e.g. where nodes are the articles and directed links are the citations of one article by another one), for instance, where highly cited papers usually continue to be cited for a long time, and vice versa. For example, papers with more than 100 citations have an average citation age of 11.7 years while the publications with more than 1000 citations have an average citation age of 18.9 years [13]. Unfortunately, the DM is unsatisfactory because the underlying mechanism for this deactivation probability, $\sim k^{-1}$, is not identified and a more fundamental model is therefore of interest.

A similar lack of clarity also occurs when one tries to justify linear preferential attachment models [10,12]. Indeed, the latter imply that entering nodes have a global knowledge of the network, i.e. they must be aware of the degrees of all previously existing nodes before connecting to one of them. This unrealistic approach can be elegantly circumvented by introducing redirection [14,15] or copying [16]–[19] mechanisms. In the simplest version, that is explained in terms of citation networks for the sake of clarity, an author who is writing the reference list for a new paper picks a random pre-existing paper. Then the author cites either the randomly selected paper (with probability $1 - r$)
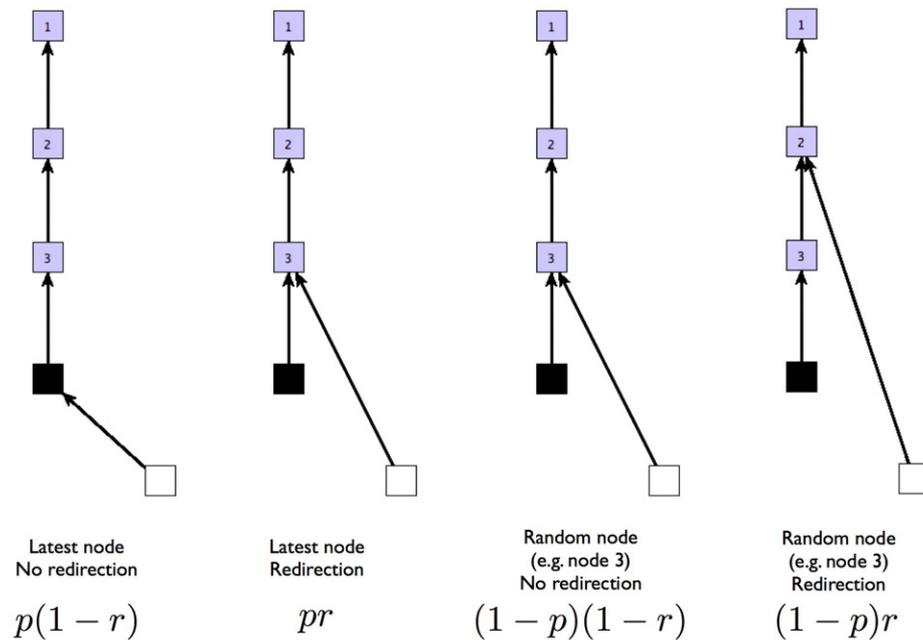
Figure 1. Four possible configurations when a new node enters the network. The latest node is darkened and the entering node is in white. With probability $p(1-r)$, the latest node receives the link from the entering node. With probability pr, the latest node is selected, but redirection takes place, so that the *father* of the latest node receives the link from the entering node. The two other possible configurations, associated with the random selection of a node (in this example, node 3), occur with probabilities $(1-p)(1-r)$ and $(1-p)r$.

or one of the references within that paper (with probability r). It is straightforward to show that this purely local process generates linear preferential attachment [14]. In this paper, we proceed along the same line of thought and introduce a model, called the link activation model (LAM), that includes ageing effects. Its interpretation is quite natural for information networks, such as citation networks. The system is a growing network where, for the sake of simplicity, entering nodes have only one outgoing link (each paper cites one other paper). One assumes that only recent nodes are active but, unlike in previous models, a node is active if it has been introduced recently or if it has been cited recently. In detail, when an author cites a paper, they either select the latest paper (the paper entered at the previous time step) with probability p or a random paper with probability $1-p$. Then, with probability r, the author cites the paper cited by the selected paper. With probability $1-r$, they cite the selected paper. The model therefore depends on two parameters p and r that measure the importance of ageing and redirection processes as compared to random effects. There are four different possibilities that can take place at each time step, as summarized in figure 1. An applet allowing the dynamical visualization of the model should also be available online [20]. Let us stress that the ingredients of the model are very general and that the LAM is not limited to citation networks, but should also apply to other information networks, e.g. the Web.

Before going further, let us make the notation precise. Initially $(t = 0)$, the network is composed of one node, the seed. For the sake of coherence, the seed has an outgoing link connected to itself. At each time step t, a new node enters the network. Consequently, the total number of nodes is equal to $N_t = 1 + t$, and the number of links is also $L_t = 1 + t$.

2. Height distribution

In this section, we focus on the height distribution, the height of a node [21] being defined to be the length of the shortest path between this node and the seed. Let us denote as $H_{g;t}$ the average number of nodes at the height g. By construction, $H_{0;t} = 1$ for all times. We also define $l_{g;t}$ to be the probability that the latest node is at height g. It is straightforward to show that these quantities satisfy the coupled rate equations

$$
\begin{aligned}
H_{g;t+1} &= H_{g;t} + (1 - p)\frac{(1 - r)H_{g-1;t} + rH_{g;t}}{t + 1} + p[(1 - r)l_{g-1} + rl_g] \\
l_{g;t+1} &= (1 - p)\frac{(1 - r)H_{g-1;t} + rH_{g;t}}{t + 1} + p[(1 - r)l_{g-1} + rl_g],
\end{aligned}
\tag{1}
$$

except for $g = 1$:

$$
\begin{aligned}
H_{1;t+1} &= H_{1;t} + (1 - p)\frac{H_{0;t} + rH_{1;t}}{t + 1} + p(l_{0;t} + rl_{1;t}) \\
l_{1;t+1} &= (1 - p)\frac{H_0 + rH_{1;t}}{t + 1} + p(l_{0;t} + rl_{1;t})
\end{aligned}
\tag{2}
$$

and for $g = 0$ where one has the trivial solutions $H_0 = 1$ and $l_0 = 0$ (this is due to the fact that an entering node can only arrive at height 1 or higher). The above rate equations are derived in the usual way and generalize the equation with $p = 0$ found in [22] for instance. It is straightforward to verify that $N_t = \sum_g H_{g;t} = t + 1$ and $l_t = \sum_g l_{g;t} = 1$.

Let us first focus on the case $p = 1$, where only latest nodes are selected, and take a continuous time limit (this is justified *a posteriori* as we are interested in the long time behaviour of the model). In that case, one has to solve

$$
\begin{aligned}
\partial_t H_{g;t} &= (1 - r)l_{g-1} + rl_g \\
\partial_t l_{g;t} &= (1 - r)l_{g-1} + (r - 1)l_g.
\end{aligned}
\tag{3}
$$

In the following, we are interested in the behaviour of the average total height $G_t = \sum_{g=0}^{\infty} gH_{g;t}$. To derive it, one also needs to evaluate the behaviour of $z_t = \sum_{g=0}^{\infty} gl_{g;t}$ which is the average height of the latest node and is easily found to satisfy

$$
\partial_t z_t = (1 - r).
\tag{4}
$$

Consequently, z_t asymptotically behaves like $(1 - r)t$ and the equation for the total height G_t reads

$$
\partial_t G_t = (1 - r) + (1 - r)t.
\tag{5}
$$

This equation leads to the asymptotic behaviour $G_t = ((1 - r)/2)t^2$. This implies that the average height $g_t \equiv G_t/(N + 1) \simeq G_t/t$ asymptotically increases linearly with time. Moreover, the redirecting process slows down the growth of the network (see figure 2).
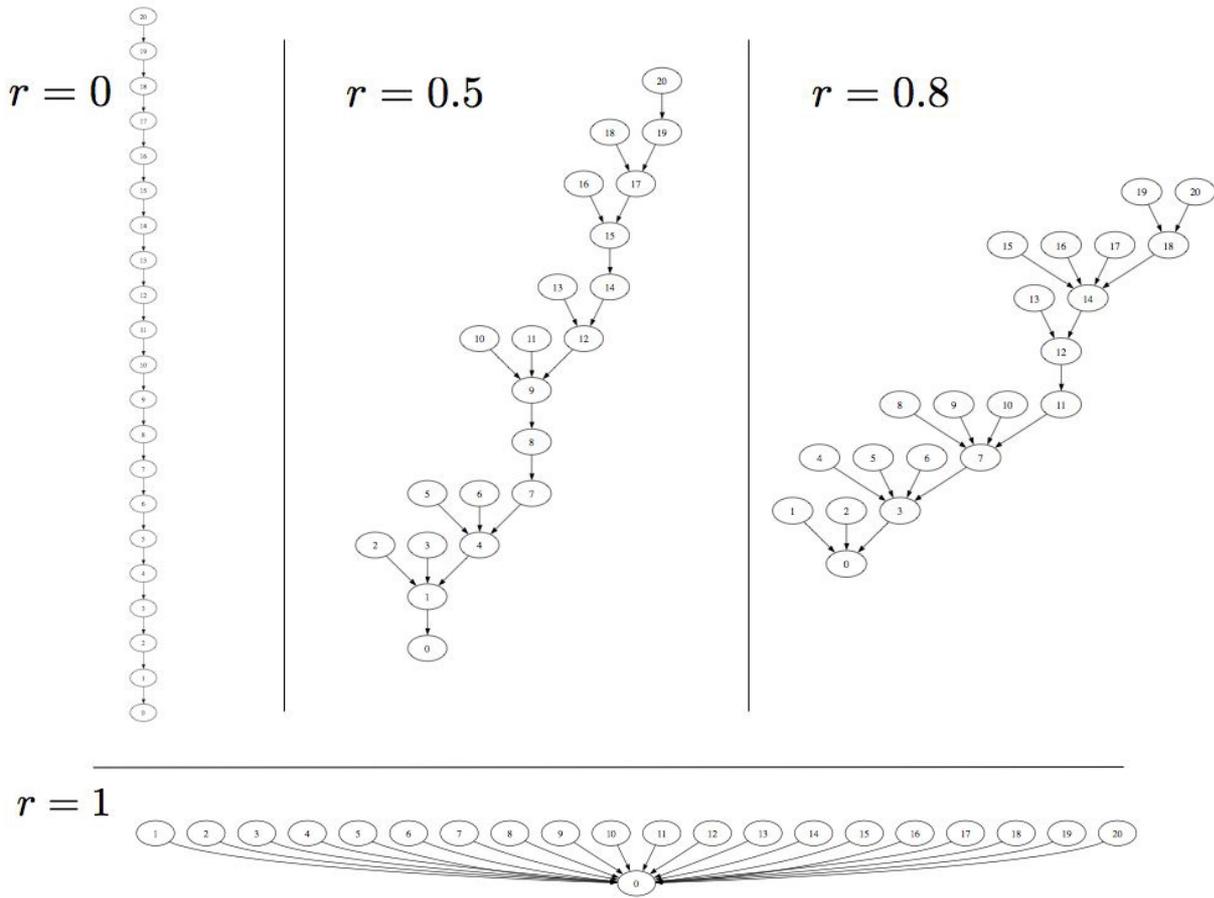
Figure 2. Typical realizations of the model when $p = 1$. In that case, the average height g increases in a large-world way, i.e. linearly with time $g_t = ((1 - r)/2)t$, and one observes a large range of behaviours, from an aligned network ($r = 0$) to a star network ($r = 1$).

This is expected, as redirection favours the connection to nodes closer to the seed. In the limiting case $p = 1$, where the process is easily shown to lead to a star network (i.e. all the nodes are connected to the seed), one finds $G_t = t \Leftrightarrow g = 1$.

Let us now focus on the more general case $p < 1$ which reads in the continuum time limit

$$
\begin{aligned}
\partial_t H_{g;t} &= (1 - p)\frac{(1 - r)H_{g-1} + rH_{g;t}}{t + 1} + p[(1 - r)l_{g-1} + rl_g] \\
\partial_t l_{g;t} &= (1 - p)\frac{(1 - r)H_{g-1} + rH_{g;t}}{t + 1} + p[(1 - r)l_{g-1} + rl_g] - l_g.
\end{aligned}
\tag{6}
$$

By using

$$
\frac{H_0 + rH_{1;t}}{t + 1} + \sum_{g>1} g\frac{(1 - r)H_{g-1} + rH_{g;t}}{t + 1} = r\frac{H_0}{t + 1} + (1 - r) + \frac{G_t}{t + 1},
\tag{7}
$$

and neglecting terms $\sim t^{-1}$, one obtains the following set of equations for the above defined average quantities:

$$\partial_t G_t = (1-p)\left(1-r+\frac{G_t}{t}\right) + p(1-r+z_t)$$
$$\partial_t z_t = (1-p)\left(1-r+\frac{G_t}{t}\right) + p(1-r+z_t) - z_t. \tag{8}$$

It is easy to simplify equation (8) into

$$t\partial_t G_t = (1-r)t + (1-p)G_t + ptz_t$$
$$t\partial_t z_t = (1-r)t + (1-p)G_t + (p-1)tz_t. \tag{9}$$

Numerical integration of the above set of equations and our knowledge of the previous simplified cases (e.g. equation (5)) suggest looking for solutions of the form $G_t = Ct\log(t)$, $z_t = C\log(t) + K$. By inserting these expressions into equations (9) and keeping leading terms in the long time limit $t \gg 1$, one finds the conditions

$$C = K = \frac{1-r}{1-p}, \tag{10}$$

which cease to be valid when $p = 1$, in agreement with the solution of equation (5). Consequently, the average height g_t asymptotically grows logarithmically with time $g_t = ((1-r)/(1-p))\log(t)$. This result should be compared with the linear regime $g_t = ((1-r)/2)t$ arising when $p = 1$. Let us stress that such a transition from a large-world ($g_t \sim t$) to a small-world [23, 24] ($g_t \sim \log(t)$) network has already been observed in another model with ageing [9] and is associated with the crossover from a structured network, reminiscent of a one-dimensional line, to an unstructured network. The above solution is in agreement with the prediction $g_t = (1-r)\log(t)$ arising in a model without ageing [22].

3. Degree distribution

Let us denote by $N_{k;t}$ the average number of nodes with k incoming links. For the sake of clarity, we first focus on three simplified cases, $r = 0$, $p = 0$ and $p = 1$, before deriving results for general values of the parameters.

When $r = 0$, there is no possible redirection and the stochastic mechanism takes place during the selection of a node. With probability p, the latest node, which has by definition zero incoming links, receives the link of the entering node, while with probability $1-p$, a random node receives this link. Consequently, the rate equation for $N_{k;t}$ [25] reads

$$\partial_t N_k = (1-p)\frac{N_{k-1} - N_k}{N} + p(\delta_{k,1} - \delta_{k,0}) + \delta_{k,0}, \tag{11}$$

where the last delta term accounts for the degree distribution of the newly entering node. We look for a stationary solution of the distribution $n_k = N_k/N$ which is determined by the recurrence relations

$$(1-p)(n_{k-1} - n_k) + p(\delta_{k,1} - \delta_{k,0}) + \delta_{k,0} - n_k = 0. \tag{12}$$

Its solution is easily found to be

$$n_0 = \frac{1-p}{2-p}$$

$$n_1 = \frac{1-p}{2-p}n_0 + \frac{p}{2-p}$$

$$n_k = \left(\frac{1-p}{2-p}\right)^{k-1} n_1, \qquad \text{for } k > 1.$$

(13)

When $p = 0$, one recovers the exponential solution $n_k = (1/2)^{k+1}$. For increasing values of p, the tail of the distribution remains exponential, but its core is more and more peaked around $k = 1$. In the limiting case $p = 1$, the solution goes to a peaked distribution $n_k = \delta_{k1}$ that corresponds to an aligned network (see figure 2).

In the case $p = 0$, the LAM reduces to the usual model with redirection for which it is well known [14] that the degree distribution evolves as

$$\partial_t n_k = r[(k-1)n_{k-1} - kn_k] + (1-r)(n_{k-1} - n_k) + \delta_{k,0} - n_k.$$

(14)

The stationary solution is therefore found by recurrence:

$$(rk + 2 - r)n_k = (rk + 1 - 2r)n_{k-1}.$$

(15)

This stationary solution has a power-law tail $k^{-\nu}$ whose exponent ν is obtained by inserting the form $n_k \sim k^{-\nu}$ into the above equation. Keeping the leading terms in k^{-1}, i.e. $(k-1)^{-\nu} = k^{-\nu}(1 - 1/k)^{-\nu} \simeq k^{-\nu}(1 + \nu/k)$, one has to solve

$$(rk + 2 - r)k^{-\nu} = (rk + 1 - 2r)k^{-\nu}(1 + \nu/k),$$

(16)

so that one recovers the value $\nu = (1 + r)/r$ derived in [14].

The case $p = 1$ is slightly more complicated, due to the fact that the selected node is always the latest node. Consequently, one also has to focus on the quantity A_k that is the average number of nodes with degree k that are cited by the latest node. By construction, this quantity satisfies $\sum_k A_k = 1$ (because there is only one latest node by construction and this latest node has only one outgoing link) and the system is described by the coupled set of equations

$$\partial_t N_k = r(A_{k-1} - A_k) + (1-r)(\delta_{k,1} - \delta_{k,0}) + \delta_{k,0}$$

$$\partial_t A_k = rA_{k-1} + (1-r)\delta_{k,1} - A_k.$$

(17)

Let us note that the equations for N_k and A_k are quite similar, except for their loss term. This is due to the fact that all nodes that do not receive a link at a time step are cited by nodes that cease to be the latest node by construction. The stationary values of A_k

$$A_0 = 0$$

$$A_k = r^{k-1}(1-r), \qquad \text{for } k > 0$$

(18)

and of the distribution n_k

$$n_0 = r$$

$$n_k = r^{k-1}(1-r)^2, \qquad \text{for } k > 0$$

(19)

are found by recurrence. In the case $r = 0$, one recovers the distribution $n_k = \delta_{k1}$ of the aligned network. Before going further, let us stress that the LAM exhibits a very

rich phenomenology, with a degree distribution that can behave like a delta peak, an exponential or a power law depending on the parameters.

By putting together the contributions of the above limiting case, it is straightforward to write a set of equations for general values of p and r:

$$\partial_t N_k = (1-p)\left[r\frac{(k-1)N_{k-1} - kN_k}{N} + (1-r)\frac{N_{k-1} - N_k}{N}\right]$$
$$+ p[r(A_{k-1} - A_k) + (1-r)(\delta_{k,1} - \delta_{k,0})] + \delta_{k,0}$$
$$\partial_t A_k = (1-p)\left[r\frac{(k-1)N_{k-1}}{N} + (1-r)\frac{N_{k-1}}{N}\right]$$
$$+ p[rA_{k-1} + (1-r)\delta_{k,1}] - A_k,$$

(20)

whose stationary solutions are found by resolving the recurrence relations

$$0 = (1-p)[r((k-1)n_{k-1} - kn_k) + (1-r)(n_{k-1} - n_k)]$$
$$+ p[r(A_{k-1} - A_k) + (1-r)(\delta_{k,1} - \delta_{k,0})] + \delta_{k,0} - n_k$$
$$0 = (1-p)[r(k-1)n_{k-1} + (1-r)n_{k-1}]$$
$$+ p[rA_{k-1} + (1-r)\delta_{k,1}] - A_k.$$

(21)

It is possible to write the formal solution of the second relation:

$$A_0 = 0$$
$$A_k = (pr)^{k-1}p(1-r) + \sum_{i=1}^{k}(pr)^{i-1}(1-p)[r(k-i) + (1-r)]n_{k-i}.$$

(22)

After inserting this solution into the first equation of equation (21), looking for a solution of the form $n_k \sim k^{-\nu}$ and keeping the leading terms in k^{-1}, it is a straightforward but lengthy procedure to get the expression

$$\nu = \frac{1 + r - 2pr}{r - pr}.$$

(23)

This solution is well defined when $p \neq 1$ and $r \neq 0$ and recovers the result derived above when $p = 0$. It is important to note that the tail of the distribution behaves like a power law for any other value of the parameters. Let us also note that equation (23) is a monotonically increasing function of p, for fixed values of r, so that ageing mechanisms have a tendency to diminish the number of nodes with very high degrees. This can be understood by noting that ageing diminishes the probability for old nodes being cited, while these old nodes are typically those with the highest degree.

4. Discussion

In this paper, we have presented a simple model for growing networks with ageing. This link activated model incorporates the fact that articles remain present in the *collective memory* as long as they are cited or read. Namely, articles that are the most likely to be cited are those that have been published recently or those that have been cited recently. In other words, all sorts of articles that are *current* and may have regularly triggered the reader's curiosity. This natural process is shown to lead to a rich behaviour for the

network structure, that leads to a transition from a large-world to a small-world network. Moreover, various kinds of asymptotic stationary degree distributions may be reached depending on the model parameters: a delta peak that corresponds to a one-dimensional lattice, exponential-like distributions or power-law tailed distributions. Let us stress the fact that the LAM is quite general and should apply to many situations involving a competition between multiplicative effects (the *rich get richer*) and ageing. Apart from citation networks that have been discussed above, one may think of *short-lived* information Web-pages. A typical example is *digg.com* where users propose a new information/article and are subject to the voting of the whole community of users. Usually, information loses its appeal within a few hours or days.

Acknowledgments

References

[1] Amaral L A N, Scala A, Barthélémy M and Stanley H E, 2000 *Proc. Natl Acad. Sci. USA* **97** 11149
[2] Dorogovtsev S N and Mendes J F F, 2000 *Phys. Rev.* E **62** 1842
[3] Hajra K B and Sen P, 2004 *Phys. Rev.* E **70** 056103
[4] Hajra K B and Sen P, 2005 *Physica* A **346** 44
[5] Lehmann S, Jackson A D and Lautrup B, 2005 *Europhys. Lett.* **69** 298
[6] Cattuto C, Loreto V and Servedio V D P, 2006 *Europhys. Lett.* **76** 208
[7] Klemm K and Eguíluz V M, 2002 *Phys. Rev.* E **65** 036123
[8] Eguíluz V M and Klemm K, 2002 *Phys. Rev. Lett.* **89** 108701
[9] Tian L, Zhu C-P, Shi D-N and Gu Z-M, 2006 *Phys. Rev.* E **74** 046103
[10] Barabási A-L and Albert R, 1999 *Science* **286** 509
[11] Krapivsky P L, Redner S and Leyvraz F, 2000 *Phys. Rev. Lett.* **85** 4629
[12] Price D J de S, 1965 *Science* **149** 510
[13] Redner S, 2005 *Phys. Today* **58** 49
[14] Krapivsky P L and Redner S, 2005 *Phys. Rev.* E **63** 066123
[15] Kleinberg J, Kumar R, Raghavan P, Rajagopalan S and Tomkins A, 1999 *Lecture Notes in Computer Science* vol 1627 (New York: Springer) p 1
[16] Krapivsky P L and Redner S, 2005 *Phys. Rev.* E **71** 036118
[17] Valverde S and Solé R V, 2005 *Europhys. Lett.* **72** 858
[18] Ohno S, 1970 *Evolution by Gene Duplication* (New York: Springer)
[19] Ispolatov I, Krapivsky P L and Yuryev A, 2005 *Phys. Rev.* E **71** 061911
[20] dyncanLAM.jar, also available on http://www.lambiotte.be
[21] Ben-Naim E, Krapivsky P L and Majumdar S N, 2001 *Phys. Rev.* E **64** R035101
[22] Lambiotte R and Ausloos M, 2007 *Europhys. Lett.* **77** 58002 [physics/0612148]
[23] Watts D J and Strogatz S H, 1998 *Nature* **393** 440
[24] Cohen R and Havlin S, 2003 *Phys. Rev. Lett.* **90** 058701
[25] Krapivsky P L and Redner S, 2003 *Statistical Mechanics of Complex Networks* (Berlin: Springer) pp 3–22, http://physics.bu.edu/∼redner/pubs/ps/sitges.ps