Link-space formalism for network analysis

David M.D. Smith[1,2],[*] Chiu Fan Lee[1], Jukka-Pekka Onnela[1,2,3], and Neil F. Johnson[1,4]

[1]*Physics Department, Clarendon Laboratory, Oxford University, Oxford OX1 3PU, U.K.*
[2]*Said Business School, Oxford University, Oxford OX1 1HP, U.K.*
[3]*Laboratory of Computational Engineering, Helsinki University of Technology, Finland and*
[4]*Physics Department, University of Miami, Coral Gables, Florida FL 33124, USA*
(Dated: March 2, 2008)

We introduce the link-space formalism for analyzing network models with degree-degree correlations. The formalism is based on a statistical description of the *fraction of links* $l_{i,j}$ connecting nodes of degrees i and j. To demonstrate its use, we apply the framework to three growing network models, namely, random-attachment, Barabási-Albert, and a model based on a mixture of random attachment and single step random walks. For the first two models the link-space matrix can be solved analytically while the numerical solution for the third exemplifies the effect of degree-degree correlations for the resulting degree distribution. We also employ the formalism to derive the degree distributions of two very simple network decay models, more specifically, that of random link deletion and random node deletion. The formalism allows detailed analysis of the correlations within networks and we also employ it to derive the form of a perfectly non-assortative network.

Networks – in particular large networks with many nodes and links – are attracting widespread attention. The classic reviews [1, 12, 14, 25] with their primary focus on structural properties have been followed up by more recent ones addressing the role of dynamics, such as spreading and synchronization processes on networks, as well as the role of weights and mesoscopic structures, i.e. cliques and communities, within networks [8, 24]. Although several different measures for characterising networks have been presented, for example in a recent survey [11], the simple concept of vertex degree remains unrivalled in its ability to capture fundamental network properties. When comparing the degrees of connected vertices, however, one often finds that they are correlated, a quality that gives rise to a surprisingly rich set of phenomena. Degree correlations constitute a central role in network characterisation and modelling but, in addition to being important in their own right, also have substantial consequences for dynamical processes unfolding on networks. Given the increasing current interest in network dynamics, understanding structural correlations remains important and timely.

In this paper we provide a detailed mathematical formalism for modelling networks with correlations. It is built around a statistical description of inter-node linkages as opposed to single-node degrees. While correlations have been characterised in empirical and model networks, most works devoted to analytical calculations of correlations in models, as pointed out in [4], have been performed only for particular cases. We start by providing a brief overview of degree correlations for network structure and dynamics in Section I. The link-space formalism, which lies at the core of the paper, is in-troduced in Section II, where we also demonstrate its use for two well-known examples, random-attachment and Barabási-Albert (BA) preferential-attachment networks [2, 3], and solve the so-called link-space matrix analytically for these models. This allows detailed analysis of the degree-degree correlations present within these networks as is demonstrated in Section III, where we utilise the formalism to derive the form of a perfectly non-assortative network. We then consider the counter-intuitive prospect of finding steady states of decaying networks in Section IV. In Section V we introduce a simple one-parameter network growth algorithm, which is able to produce under-skewed, over-skewed and scale-free networks via a redirection process similar to the model by Krapivsky *et al.* [20]. The model is interesting in its own right in the sense that it makes use of only local information about node degrees. Here the link-space formalism allows us to identify the transition point at which over-skewed node-degree distributions switch to under-skewed distributions with respect to the BA model. We conclude in Section VI. To maintain readability, we postpone the more detailed mathematical derivations to Appendices A, B, and C, and refer to them where appropriate.

I. OVERVIEW OF DEGREE CORRELATIONS

Vertex degree correlations express the idea that the degrees of a set of adjacent vertices are statistically dependent. In general, n-vertex degree correlations, or n point correlations, can be fully characterised by the conditional probability distribution $P(k_1, k_2, \ldots, k_n | k = n)$ that a vertex of degree $k = n$ is connected to a set of n vertices with degrees $k_1, k_2, \ldots, k_n$. Two and three point correlations are of particular importance in complex networks as they can be related to network assortativity and clustering, respectively. These quantities can be used to

further classify networks based on their structural and ensuing dynamic properties. More specifically, two vertex degree correlations (two point correlations) can be fully characterised by the conditional probability $P(k'|k)$ that a vertex of degree k is connected to a vertex of degree k', meaning that the degrees of the vertices adjacent to any given edge are not statistically independent. Similarly, three vertex degree correlations (three point correlations) can be fully characterised by the conditional probability distribution $P(k', k''|k)$ that a vertex of degree k is connected to both a vertex of degree k' and a vertex of degree k''. This implies that the degrees of neighbouring nodes are not statistically independent.

Reliable estimation of $P(k'|k)$ and $P(k', k''|k)$ requires a large amount of data and, in practice, one often resorts to related measures. Instead of $P(k'|k)$, one typically reports the behaviour of the average degree of nearest neighbours, $\bar{k}_{NN}(k)$, which can be formally related to $P(k'|k)$ [27]. If $\bar{k}_{NN}(k)$ increases with k, high degree vertices tend to connect to high degree vertices. A network with this property would be described as being *assortative*. If $\bar{k}_{NN}(k)$ decreases with k high degree vertices tend to connect to low degree vertices (*disassortative*) [26]. An alternative to $\bar{k}_{NN}(k)$ is to use a streamlined, single-number measure of the assortativity using a normalised Pearson's correlation coefficient of adjacent vertex degrees as suggested by Newman [25]. To characterise three point correlations, instead of using $P(k', k''|k)$, one can rely on the clustering spectrum $\bar{c}(k)$, which can be formally related to $P(k', k''|k)$ [9].

Some of the earlier network papers focused on calculating degree distributions for different models analytically. As anecdotal examples, one can think of the highly influential paper of Barabási *et al.*, who introduced the BA-model and developed a mean-field theory for growing network models which, under the continuous degree and continuous time approximation, enabled them to calculate the degree distribution and the scaling exponents analytically [2, 3]. Dorogovtsev *et al.* considered a more general model and solved the associated discrete degree master equation analytically [13], confirming the mean-field scaling exponents for the degree distribution of the BA model of [2, 3]. A closely related model of preferential growth of clusters of particles was considered by Kullmann *et al.*, who produced a full time-dependent solution to the associated master equation [23].

While the above listed papers are all commendable achievements in their own right, they did not account for correlations between vertex degrees. As the field has matured, however, the role of vertex degree correlations has been increasingly acknowledged. Krapivsky *et al.* investigated the structural properties of different growing network models in terms of rate equations for the densities of nodes of a given degree and demonstrated that correlations emerge spontaneously between the degrees of connected nodes as the network grows [20]. They characterised this for a certain *directed* network model in terms of $n_{kl}(t)$, the fraction of nodes of total degree

k which attach to an ancestor (older) node of total degree l at time t, and derived an analytical expression for it. Barrat *et al.* introduced a framework for computing the rate equation for two vertex correlations, in the continuous degree and continuous time approximation, and also worked out the corresponding boundary condition [4]. Klemm *et al.* showed that in BA networks the average clustering of a given node is independent of its degree and tends to a constant value that vanishes in the thermodynamic limit [19]. Szabó *et al.* formulated a scaling assumption and a mean-field theory of clustering in growing scale-free networks [31]. The approach of Szabó *et al.* is amenable to different models, and the authors demonstrated the framework by writing and solving the rate equations for a simple model. It is interesting to note that the commonly observed phenomenon of decreasing average clustering coefficient as a function of degree, i.e. $\bar{c}(k)$ vs k, which reflects three point correlations as mentioned above. While this is sometimes interpreted as a signature of hierarchical structure in a network, Soffer *et al.* suggested that is is a consequence of degree-degree (two point) correlations that enter the definition of the standard clustering coefficient [41]. The authors introduced a different definition for the clustering coefficient that does not have the degree-correlation 'bias', i.e. a three point correlation measure that filters out two point correlations. Barrat *et al.* discussed a model for growing weighted networks in which structural growth is coupled to the evolution of weights in the network, and in this sense topology and weights are correlated [5].

Degree correlations have been successfully incorporated also into some slightly more specific network models. For example, Catanzaro *et al.* studied the social network of scientists based on their collaboration patterns on one of the electronic archives, found it to be assortative, and presented a model that reproduced the assortative network structure [38]. Toivonen *et al.* introduced a model of social networks that, in addition to being assortative, also produced communities with dense internal connections and, as a consequence, gave rise to high clustering [43]. Park *et al.* demonstrated that negative degree correlations in the network representation of the Internet could be partly accounted for by imposing a restriction that no two vertices have more than one edge connecting them [39]. Berg *et al.* presented a model for the evolution of protein interaction networks based on two evolutionary processes and found that the imposed link dynamics, i.e. gain and loss of interactions through mutations in existing proteins, shaped the statistical structure of the network, resulting in correlations between the connectivities of interacting proteins in accordance with empirical findings [40]. Maslov *et al.* found that in interaction and regulatory networks, links between highly connected proteins were systematically suppressed, whereas those between a highly connected and low-connected pairs of proteins were favoured, a topological organisation that increases the overall robustness of the network to perturbations [42].

Important advances have been made recently on the role of degree correlations on network dynamics. As the literature is rapidly accumulating, the following can only be a brief taster of it and is intended to convey the importance of and motivation for understanding correlations from the perspective of dynamics. Equíluz *et al.* considered clustered scale-free networks and showed that correlations play an important role in epidemic spreading [15]. According to them, prevalence, i.e. the time average of the fraction of infected individuals in the steady-state, undergoes a phase transition at a finite critical infection probability. Consequently, unlike in uncorrelated scale-free networks in which viruses with extremely low spreading probabilities can become epidemics, the presence of degree correlations changes the spreading dynamics and now the transmission probability needs to exceed a finite threshold for a virus to spread and prevail. This is compatible with the work of Boguñá *et al.* who found that the epidemic threshold in the presence of degree correlations was determined by the connectivity matrix as opposed to the degree distribution as in the case of uncorrelated networks [37]. Brede *et al.* examined the dynamical stability of networked systems by studying the spectral properties of the associated Jacobian matrix, generated from the underlying adjacency matrix, which governs the behaviour of the system to a small perturbation about an equillibrium state. They induced degree-degree correlations to Erdös-Rényi and scale-free networks by rewiring them appropriately and found that positive correlations reduced their dynamical stability [34]. Along similar lines, Bernardo *et al.* found that inducing negative degree-degree correlations in scale-free networks of non-linear oscillators affected the so-called Laplacian eigenratio with the outcome that network synchronizability improved as the network was made more dissassortative [36]. Based on this result, they conjectured that negative degree correlations may emerge spontaneously as the networked system attempts to become more stable. The same authors found similar results to hold also for weighted networks [36].

II. NODE-SPACE AND LINK-SPACE

We will now introduce the link-space formalism. Consider a simple growing network in which one node and one undirected link are added at each timestep. The process is governed by the attachment probability *kernel* Θ_j, defined as the probability that the newly introduced node links to an existing node of degree j. At some time t there exist $X_i \mid_t$ nodes of degree i and we wish to compute the expected number of nodes degree i at time $t + 1$. This *node-space master equation* can be expressed in terms of the attachment kernels and for $i > 1$ is written

$$\langle X_i \rangle \mid_{t+1} = X_i \mid_t + \Theta_{i-1} \mid_t - \Theta_i \mid_t . \tag{1}$$

For $i = 1$ we have $\langle X_1 \rangle \mid_{t+1} = X_1 \mid_t + 1 - \Theta_1 \mid_t$ since a new node of degree one is added to the existing net-

work at each timestep. We assume that there exists a steady-state in which the degree distribution $P(i)$ remains constant [44]. Under this assumption, the fraction of nodes of a given degree remains constant such that $dX_i/dt = c_i = P(i)$. It is also assumed that in the steady state, the attachment kernels are static too. We drop the notation '$\mid_t$' to indicate the steady-state and can trivially rewrite Eq.(1) as

$$c_i = \Theta_{i-1} - \Theta_i. \tag{2}$$

So far we have said nothing about the attachment mechanism, and have made the easily geneneralizable restriction that only one node and one undirected link is being added per timestep. We now follow a similar analysis, but retain the node-node linkage correlations that are inherent in many real-world systems [10, 20, 26]. Consider any link in a general network – we can describe it by the degrees of the two nodes that it connects. Hence we can construct a matrix $\mathbf{L}$ such that the element $L_{i,j}$ describes the number of links from nodes of degree i to nodes of degree j for $i \neq j$. To ease the mathematical analysis below, the diagonal element $L_{i,i}$ is assumed to be twice the number of links between nodes of degree i for the undirected graph. For undirected networks $\mathbf{L}$ is symmetric and $\sum_{i,j} L_{i,j} = 2M$, twice the total number of links M in the network. Matrix $\mathbf{L}$ represents a surface describing degree-degree correlations in the network (see Fig. 1).

The probability of selecting any node of degree $i - 1$ within the existing network for the new node to attach to is given by the attachment probability $\Theta_{i-1} \mid_t$. Suppose an $i - 1$ node is selected. The fraction of nodes of degree $i - 1$ that are connected to nodes of degree j is

$$\frac{L_{i-1,j} \mid_t}{(i-1)X_{i-1} \mid_t}.$$

The expected increase in links from nodes of degree i to nodes of degree j, through the attachment of the new node to a node of degree $i - 1$, is given by

$$\frac{\Theta_{i-1} \mid_t L_{i-1,j} \mid_t}{X_{i-1} \mid_t}.$$

Since each link has two ends, the value $L_{i,j}$ can increase by a connection to an $(i-1)$-degree node which is in turn connected to a j-degree node, or by connection to an $(j-1)$-degree node which is in turn connected to an i-degree node. We write the *link-space master equation* for the evolution of the expected number of links from i to j degree nodes, i.e. number of $i \leftrightarrow j$ links, as

$$\langle L_{i,j} \rangle \mid_{t+1} = L_{i,j} \mid_t + \frac{\Theta_{i-1} \mid_t L_{i-1,j} \mid_t}{X_{i-1} \mid_t} + \frac{\Theta_{j-1} \mid_t L_{i,j-1} \mid_t}{X_{j-1} \mid_t}$$
$$- \frac{\Theta_i \mid_t L_{i,j} \mid_t}{X_i \mid_t} - \frac{\Theta_j \mid_t L_{i,j} \mid_t}{X_j \mid_t}, \quad i,j > 1;$$

$$\langle L_{1,j} \rangle \mid_{t+1} = L_{1,j} \mid_t + \Theta_{j-1} + \frac{\Theta_{j-1} \mid_t L_{1,j-1} \mid_t}{X_{j-1} \mid_t} \tag{3}$$
$$- \frac{\Theta_1 \mid_t L_{1,j} \mid_t}{X_1 \mid_t} - \frac{\Theta_j \mid_t L_{1,j} \mid_t}{X_j \mid_t}, \quad i=1, j > 1.$$
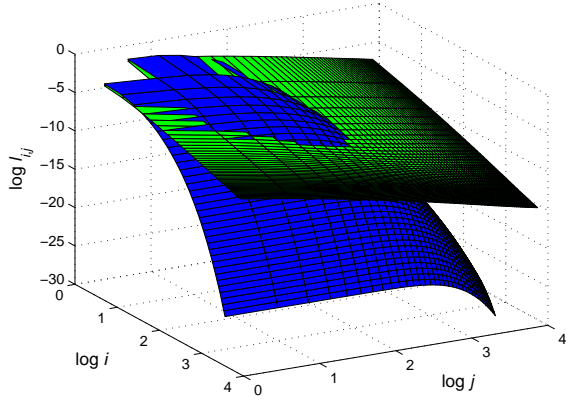
FIG. 1: (Color Online) Representation of the link-space matrix with elements $l_{i,j}$, in logarithmic coordinates, of the steady-states of the random-attachment (curved surface) and preferential-attachment (flatter surface) algorithms. The value $l_{1,1}$ is zero for networks consisting of a single component.

In the steady-state we can write this in terms of the fraction $l_{i,j} = L_{i,j}/M$ of $i \leftrightarrow j$ links as

$$
\begin{aligned}
l_{i,j} &= \frac{\frac{\Theta_{i-1}}{c_{i-1}} l_{i-1,j} + \frac{\Theta_{j-1}}{c_{j-1}} l_{i,j-1}}{1 + \frac{\Theta_i}{c_i} + \frac{\Theta_j}{c_j}}, & i,j > 1; \\
l_{1,j} &= \frac{\frac{\Theta_{j-1}}{c_{j-1}} l_{1,j-1} + \Theta_{j-1}}{1 + \frac{\Theta_1}{c_1} + \frac{\Theta_j}{c_j}}, & j > 1. \quad (4)
\end{aligned}
$$

The value $l_{1,1}$, (twice) the fraction of isolated links, is zero for all the growing networks discussed in this paper as the growth algorithms are such that the network consists of a single connected component. The notation $|_t$ has again been dropped to indicate the steady-state.

To apply the link-space formalism one starts by specifying the model dependent attachment kernel Θ_i. Substitution into the Eqs. 2 and 4 yields recurrence relations for c_i and $l_{i,j}$ respectively which can be solved analytically in some cases. The number of i-degree nodes is now given by $X_i = i^{-1} \sum_k L_{i,k}$, which allows us to retrieve the degree distribution from the link-space matrix as $P(i) = c_i = i^{-1} \sum_k l_{i,k}$. We will now demonstrate the use of the formalism to study a random attachment model and the Barabási-Albert (BA) model.

A. Random attachment model.

Consider first a random-attachment model in which at each timestep a new node is added to the network and connected to an existing node with uniform probability without any preference ('Model A' in [2, 3]). The attachment kernel is $\Theta_i = c_i$. Substituting into Eq. (2), we obtain the recurrence relation $c_{i+1} = c_i/2$, which yields

the familiar degree distribution $P(k) = 2^{-k}$. Substituting into Eq. (4) yields the recurrence relation

$$
\begin{aligned}
l_{i,j} &= (l_{i-1,j} + l_{i,j-1})/3, & i,j > 1; \\
l_{1,j} &= (c_{j-1} + l_{1,j-1})/3, & j > 1; \quad (5)
\end{aligned}
$$

with $l_{1,1} = 0$. The exact solution (see Appendix A for details) for $l_{i,j}$ is

$$
l_{i,j} = \sum_{\alpha=2}^{j} \frac{\binom{i-1+j-\alpha}{j-\alpha}}{3^{(i+j-\alpha)} 2^{(\alpha-1)}} + \sum_{\alpha=2}^{i} \frac{\binom{(i-1+j-\alpha)}{i-\alpha}}{3^{(i+j-\alpha)} 2^{(\alpha-1)}}
$$

for $i,j > 1$ and $l_{i,j} = \sum_{k=1}^{j-1} (3^k 2^{j-k})^{-1}$ for $i = 1, j > 1$, where $\binom{x}{y}$ is the conventional combinatorial 'choose' function. We shall make use of this solution to investigate the correlations of such a network in Section III.

B. Barabási-Albert (BA) model.

In the BA model [2, 3] at each timestep a new node is added to the network and connected to one existing node with a probability that is proportional to the degree of the node, i.e. $\Theta_i \propto i$, yielding

$$
\Theta_i \mid_t = \frac{i X_i \mid_t}{\sum_j j X_j \mid_t} = \frac{i X_i \mid_t}{2(N \mid_t -1)}, \quad (6)
$$

where the latter equality holds when the new node is added with one link. In the limit of large N (and the steady state), this is well approximated by $\Theta_i \approx i c_i/2$. Substituting this into Eq. (2) yields the recurrence relation $c_i = \frac{(i-1) c_{i-1}}{2} - \frac{i c_i}{2} = \frac{i-1}{i+2} c_{i-1}$, whose solution is $c_i = \frac{4}{i(i+1)(i+2)}$. Using the same substitution, the link-space master equations yield the recurrence relations:

$$
\begin{aligned}
l_{i,j} &= \frac{(i-1)l_{i-1,j} + (j-1)l_{i,j-1}}{2+i+j}, & i,j > 1; \\
l_{1,j} &= \frac{(j-1)c_{j-1} + (j-1)l_{1,j-1}}{3+j}, & j > 1. \quad (7)
\end{aligned}
$$

The exact solution for $l_{i,j}$ is obtained by algebraic manipulation of the link-space matrix using the previously derived degree distribution (see Appendix B), and is given by

$$
\begin{aligned}
l_{i,j} &= \frac{4(j-1)!(i-1)!}{(j+i+2)!} \Big\{ G(i+1) + 2G(i) - 3G(i-1) \\
&\quad + \frac{1}{2} \sum_{\alpha=2}^{i} (\alpha-1)(\alpha+6) \left[G(i-\alpha) - G(i-\alpha-1) \right] \Big\},
\end{aligned}
$$

where $G(x) = \begin{cases} \frac{(j+x-1)!}{x!(j-1)!} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0. \end{cases}$

The first few rows of this matrix have the form

$$
\begin{aligned}
l_{1,j} &= \frac{2(j+6)(j-1)}{j(j+1)(j+2)(j+3)}; \\
l_{2,j} &= \frac{2j(j-1)(j+10)+48}{3j(j+1)(j+2)(j+3)(j+4)}; \\
l_{3,j} &= \frac{3j^4 + 42j^3 - 3j^2 + 246j + 360}{9j(j+1)(j+2)(j+3)(j+4)(j+5)}. \quad (8)
\end{aligned}
$$

We shall use this solution to investigate the correlations with this network in Section III.

III. DEGREE CORRELATIONS AND ASSORTATIVITY

The link-space formalism allows us to address two-vertex correlations in a powerful way. Consider selecting an edge at random in the network. If we select one end of the edge at random, the probability that this node has degree j will be proportional to j since higher degree nodes have, by definition, more links connected to them than low degree nodes. Consider now only a subset of all edges with one end attached to a node of degree i. If there were no correlations present, the probability that the other end is attached a node of degree j is again proportional to j. This is the criterion of non-assortativity as described by Vázquez *et al* [33]:

$$P(j|i) \ = \ P(j) \ = \ \frac{jc_j}{2}. \qquad (9)$$

For a given network, we can write the conditional probability $P(j|i)$, i.e. the probability that a randomly chosen edge is connected to a node of degree j given that the other end is connected to a node of degree i, in terms of the link-space matrix as $P(j|i) = L_{i,j}/(iX_i)$. For total number of edges $\approx$ total number of nodes this can be approximated as $l_{i,j}/(ic_i)$. This is illustrated in Fig. 2 for the pedagogical scenarios of random and preferential attacment as well as the non-assortative criterion of Eq. (9). The average nearest-neighbor degree $\langle k_{nn} \rangle_i$ of a node of degree i can be easily obtained from the link-space matrix and is given by $\langle k_{nn} \rangle_i = (\sum_j L_{i,j})^{-1} \sum j L_{i,j} = (\sum_j l_{i,j})^{-1} \sum j l_{i,j}$ as illustrated in Fig. 3.

If the average nearest-neighbour degree $\langle k_{nn} \rangle_i$ is constant with respect to degree i, non-assortativity is implied. Certainly, the random attachment curve continues increase, implying positive assortative mixing. The preferential attachment curve in Fig. 3 appears to asymptote to a constant value (as was noted in [15]) although the fact that it decreases initially illustrates that it is not perfectly non-assortative. This is a feature that would not be captured using a normalised Pearson's correlation coefficient which would suggest non-assortativity for this network [25, 30].

A. Perfect non-assortativity

If adhering to the strict interpretation of non-assortatvity of Vázquez *et al.*, it is interesting to ask whether a perfectly non-assortative network can be generated. In [25], Newman suggests a network-generating mechanism to produce an arbitrary joint distribution of vertex degree at the ends of each link. As such, if a link-space matrix can be found such that it satisfies (9), the network it represents can be constructed.
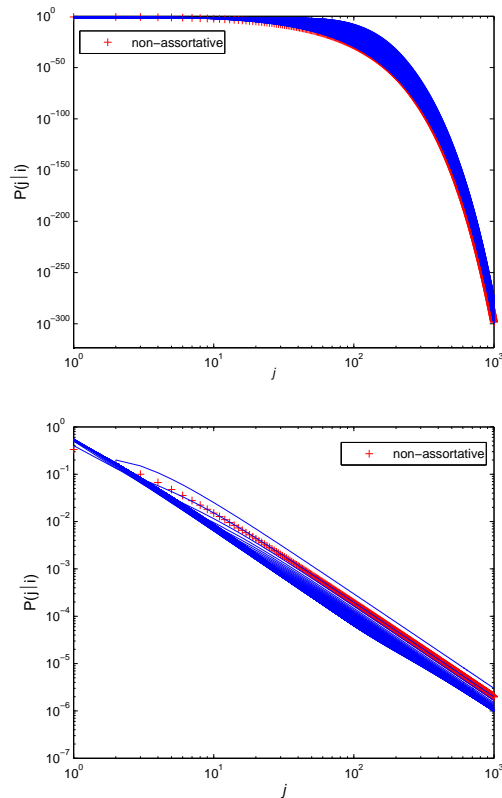

FIG. 2: (Color Online) The conditional probability that one end of a randomly selected link has degree j given that the other end is of degree i over the range $i : 1 \rightarrow 100$. The top plot is for the random attachment algorithm and below is preferential attachment. The criterion of non-assortativity of Vázquez *et al* [33] is denoted with red crosses.
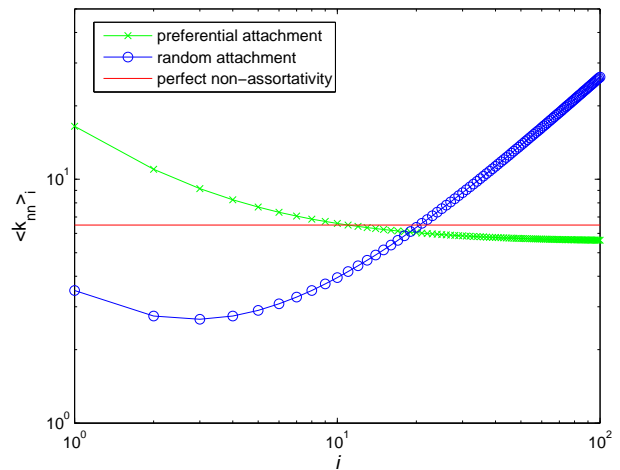

FIG. 3: (Color Online) The mean degree of the nearest neighbours of nodes of degree i, $\langle k_{nn} \rangle_i$ as a function of i for both the preferential and random attachment algorithms and the perfectly non-assortative network described in Section III A.

Assuming that such a network will not have equal numbers of nodes and links, we must rewrite the conditions of non-assortativity accordingly for total number of edges M:

$$P(j|i) = P(j) = \frac{jX_j}{2M} = \frac{jN}{2M}c_j. \qquad (10)$$

Recalling our definition that $l_{i,j} = L_{i,j}/M$, we can express this conditional probability $P(j|i)$ in terms of the link-space as

$$P(j|i) = \frac{L_{i,j}}{iX_i} = \frac{M\,l_{i,j}}{i\,N\,c_i}. \qquad (11)$$

As such, we can now write for $l_{i,j}$

$$l_{i,j} = \left(\frac{N}{M}\right)^2 \frac{i\,c_i\,j\,c_j}{2}. \qquad (12)$$

The link-space matrix which fulfills this criterion is the following surface:

$$l_{i,j} = \frac{2}{i(i+1)j(j+1)}. \qquad (13)$$

The self consistency requirements of the matrix and how it relates to node degree are such that the number of links must be less than the number of nodes. We can write the degree distribution as

$$c_i = \frac{M}{N}\frac{\sum_j l_{i,j}}{i} = \frac{2M}{Ni^2(i+1)}. \qquad (14)$$

This is clearly consistent with (13) and (12). Normalising such that the sum over the degree distribution is one, we obtain

$$\frac{M}{N} = \frac{3}{\pi^2 - 6}. \qquad (15)$$

This ratio, which is approximately 0.78, also represents half the mean degree of such a network.

IV. DECAYING NETWORKS

Although somewhat counter-intuitive it is possible to find steady states of networks whereby nodes and/or links are removed from the system. This is an idea that has been considered before [17, 28]. Aside from the obvious situation of having no nodes or edges left, there can be a steady state whereby the link-space correlation matrix and node degree distribution are static. We shall discuss the two simplest cases, that of random link removal and that of random node removal.

A. Random Link Removal (RLR)

Consider an arbitrary network. At each timestep, we select a fixed number w links at random and remove them. We shall implement the link-space to investigate whether or not it is possible that such a mechanism can lead to structure. Consider the link-space element $L_{i,j}\,|_t$ denoting the number of links from nodes of degree i to nodes of degree j. Clearly this can be decreased if an $i \leftrightarrow j$ links is removed, i.e. a link that connects a degree i node to a degree j node. Also, if a $k \leftrightarrow i$ is removed and that i node has further links to j degree nodes, then those that were $i \leftrightarrow j$ links will now be $i - 1 \leftrightarrow j$, similarly for $k \leftrightarrow j$ links being removed. However, if the link removed is a $k \leftrightarrow j+1$ link and that $j+1$ degree node is connected to a i degree node, then when the $j+1$ node becomes an j node, that link will become an $i \leftrightarrow j$ link increasing $L_{i,j}$. Let us assume that we are removing links at random from the network comprising $N\,|_t$ vertices and $M\,|_t$ links. A non-random link selection process could be incorporated into the master equations using a probability kernel. The master equation for this process can be written in terms of the expected increasing and decreasing contributions:

$$\langle L_{i,j}\rangle\,|_{t+1} = L_{i,j}\,|_t + w\sum_k \frac{L_{k,j+1}\,|_t\,jL_{i,j+1}\,|_t}{M\,|_t\,(j+1)X_{j+1}\,|_t}$$
$$+w\sum_k \frac{L_{k,i+1}\,|_t\,iL_{i+1,j}\,|_t}{M\,|_t\,(i+1)X_{i+1}\,|_t}$$
$$-w\sum_k \frac{L_{k,j}\,|_t\,(j-1)L_{i,j}\,|_t}{M\,|_t\,(j)X_j\,|_t}$$
$$-w\sum_k \frac{L_{k,i}\,|_t\,(i-1)L_{i,j}\,|_t}{M\,|_t\,(i)X_i\,|_t}$$
$$-\frac{wL_{i,j}\,|_t}{M\,|_t}. \qquad (16)$$

This simplifies to

$$\langle L_{i,j}\rangle\,|_{t+1} = L_{i,j}\,|_t + \frac{wjL_{i,j+1}\,|_t}{M\,|_t}$$
$$+\frac{wiL_{i+1,j}\,|_t}{M\,|_t} - \frac{w(i-1)L_{i,j}\,|_t}{M\,|_t}$$
$$-\frac{w(j-1)L_{i,j}\,|_t}{M\,|_t} - \frac{wL_{i,j}\,|_t}{M\,|_t}, \qquad (17)$$

where the last term refers to the physical removal of an $i \leftrightarrow j$ link.

To investigate the possibility of a steady state solution, we make a similar argument as before for growing networks but this time for the process of removing w links per timestep:

$$L_{i,j}\,|_t = l_{i,j}M\,|_t$$
$$= l_{i,j}(M_0 - wt)$$
$$\frac{dL_{i,j}}{dt} = -wl_{i,j}$$
$$\approx L_{i,j}\,|_{t+1} - L_{i,j}\,|_t. \qquad (18)$$

The expression of Eq. (17) can be reduced to

$$l_{i,j} = \frac{il_{i+1,j} + jl_{i,j+1}}{i + j - 2} \qquad (19)$$

As such, the number of $1 \leftrightarrow 1$ links $l_{1,1}$ does not reach a steady state. This might be expected, as the removal process for such links requires them to be physically removed as opposed to the process by which the degree of the node at one end of the link is being reduced or increased. The value $l_{1,1}$ increases in time. However, we can investigate the properties of the rest of the links in the network which do reach a steady state by neglecting these two node components (see Appendix C for details). The link space can be written for this system for $i + j \neq 2$ as

$$l_{i,j} = \frac{A}{2^{i+j}} \frac{(i+j-3)!}{(i-1)!(j-1)!}, \qquad (20)$$

and the degree distribution subsequently derived:

$$A = \frac{4N}{M(1+ln(2))}$$
$$c_1 = \frac{ln(2)}{1+ln(2)}$$
$$c_i = \frac{1}{(1+ln(2))i(i-1)} \qquad i > 1 . \qquad (21)$$

It is interesting that the process of randomly removing links generates a network which exhibits power-law scaling of exponent 2, corresponding to a degree distribution significantly over-skewed with respect to the preferential selection process of the BA growth algorithm (exponent 3). By considering the average degree of the neighbours of nodes of degree i, denoted $\langle k_{nn} \rangle_i$ as discussed in Section III, we can see that the random link removal algorithm generates highly assortative networks.

B. Random Node Removal (RNR)

In a similar manner to Section IV A we will now discuss the possibility of creating such a steady state via a process of removing nodes (with all of their links) from an existing network at the rate of w nodes per timestep. Clearly, removing some node which has a link to an $i+1$ degree node which in turn has a link to a j degree node can increase the number of $i \leftrightarrow j$ links in the system. The other processes which can increase or decrease the number of links from i degree nodes to j degree nodes can be similarly easily explained. We consider that we select a node of some degree k for removal with some probability kernel Θ_k (as in the growing algorithms of Section II). The master equation for such a process can thus be written for general node selection kernel:

$$\langle L_{i,j} \rangle |_{t+1} = L_{i,j} |_t \ - \ w \frac{\Theta_i |_t L_{i,j} |_t}{X_i |_t} - w \frac{\Theta_j |_t L_{i,j} |_t}{X_j |_t}$$
$$+ w \sum_k \frac{\Theta_k |_t}{X_k |_t} \Bigg\{ L_{k,i+1} |_t \ \frac{i L_{i+1,j} |_t}{(i+1) X_{i+1} |_t}$$
$$+ L_{k,j+1} |_t \ \frac{j L_{i,j+1} |_t}{(j+1) X_{j+1} |_t} \Bigg\}$$
$$- w \sum_k \frac{\Theta_k |_t}{X_k |_t} \Bigg\{ L_{k,i} |_t \ \frac{(i-1) L_{i,j} |_t}{i X_i |_t}$$
$$+ L_{k,j} |_t \ \frac{(j-1) L_{i,j} |_t}{j X_j |_t} \Bigg\}. \qquad (22)$$

We now select a kernel for selecting the node to be removed, namely, the random kernel although the approach is general to any node selection procedure. As such $\Theta_k = c_k$ as a node is selected purely at random. The steady state assumptions must be clarified slightly. As before

$$L_{i,j} |_t = l_{i,j} M |_t . \qquad (23)$$

However, because we can remove more than one link (through removing a high degree node for example) we must approximate for $M |_t$. Using the random removal kernel, we can assume that on average, the selected node will have degree $\langle k \rangle |_t = \frac{2M|_t}{N|_t}$. As such, we can use this to write the number of remaining links in the network as

$$L_{i,j} |_t = l_{i,j}(M_0 - w\langle k \rangle |_t)$$
$$= l_{i,j}(M_0 - \frac{2wM |_t}{N |_t} t)$$
$$\frac{dL_{i,j}}{dt} = -\frac{2wM |_t}{N |_t} l_{i,j}$$
$$\approx L_{i,j} |_{t+1} - L_{i,j} |_t . \qquad (24)$$

Making use of the link-space identities and Eq. (24), the master equation Eq. (22) can be written in simple form:

$$l_{i,j} = \frac{i l_{i+1,j} + j l_{i,j+1}}{i + j - 2}. \qquad (25)$$

Clearly, this is identical to Eq. (19) for the random link removal model of Section IV A and, consequently, the analysis of the degree distributions will be the same too.

V. MIXTURE MODEL

In this Section we introduce a model that makes use of only local information about node degrees as microscopic mechanisms requiring global information are often unrealistic for many real-world networks [32]. It therefore provides insight into possible alternative microscopic mechanisms for a range of biological and social networks.

The link-space formalism allows us to identify the transition point at which over-skewed node-degree distributions switch to under-skewed distributions with respect to the BA preferential attachment model. While similar local algorithms have been proposed in the literature [16, 29], the strength of the approach followed here is the ability to describe the inherent degree-degree correlations [10, 20, 25, 26].

It is well known that a mixture of random and preferential attachment in a growth algorithm can produce power law degree distributions with exponents greater than three, $\gamma \in [3, \infty)$. It has often been assumed that a *one step* random walk replicates linear preferential attachment [29, 32]. This is *not* true. A one step random walk is in fact more biased towards high degree nodes than preferential attachment [45] as can be easily seen by performing the procedure on a simple hub and spoke network. In this case, the probability of arriving at the hub tends to one for increasingly large networks rather than a half as would be appropriate for preferential attachment. We can use this bias to generate networks with degree distributions that are overskewed (lower power law exponent) than the preferential attachment model. A mixture of this approach with random attachment results in a simple model that can span a wide range of degree distributions. This one-parameter network growth model, in which we simply attach a single node at each timestep, does not require prior knowledge of the existing network structure.

Explicitly, the algorithm proceeds as follows: (i) pick a node κ within the existing network at random; (ii) with probability a make a link to that node; otherwise (iii) pick any of the neighbours of κ at random and link to that node. Hence this algorithm resembles an object or 'agent' making a short random-walk. This is very similar to the Growing Network with Redirection model introduced by Krapivsky *et al* [20] except that the model here employs undirected links which necessitates a random (as opposed to deterministic) walk. Fig. 4 shows examples of the resulting networks and the corresponding (culmulative) degree distributions. Interestingly, $a = 0$ yields a network that is dominated by hubs and spokes while $a = 1$ yields the random attachment network. Intermediate values of a yield networks which are neither too ordered nor too disordered. For $a \sim 0.2$, the algorithm generates networks whose degree-distribution closely resembles the BA preferential-attachment network (see Figs. 4 and 5).

Our analysis of the algorithm starts by establishing the attachment probability Θ_i, which in turn requires properly resolving the one-step random walk. The link-space formalism provides us with an expression for the probability P'_i associated with performing a random walk of length one and arriving at a degree i node [30]:

$$P'_i = \frac{X_1 \mid_t L_{1,i} \mid_t}{N_t X_1 \mid_t} + \frac{X_2 \mid_t L_{2,i} \mid_t}{2 N_t X_2 \mid_t} + \frac{X_3 \mid_t L_{3,i} \mid_t}{3 N_t X_3 \mid_t} + \dots \quad (26)$$

This can also be written [30] as $P'_i = \frac{i X_i \mid_t}{N_t} \langle \frac{1}{k_{nn}} \rangle_i$ where the average is performed over the neighbours of nodes
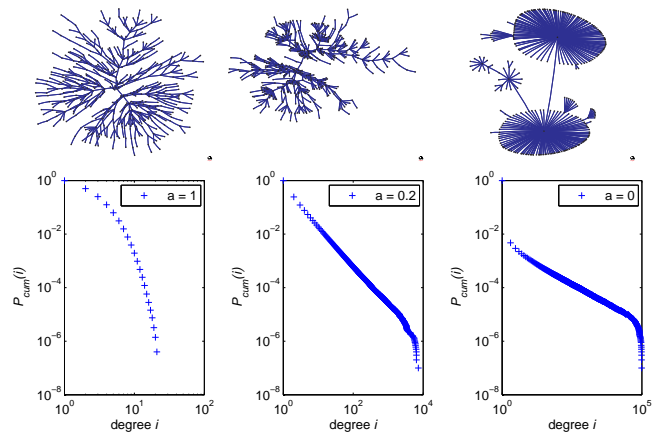

FIG. 4: (Color Online) Top: Networks generated using the one-parameter, local information growth algorithm with $a = 1$ (left), $a = 0.2$ (center), $a = 0$ (right). Initial network seed comprised two nodes and one link. Graphs drawn with Pajek [6]. Bottom: Culmulative degree distributions for the same networks grown to $100,000$ nodes and ensemble-averaged over 100 networks per a value.

with degree i. Note that this quantity does *not* replicate preferential attachment, in contrast to what is commonly thought [29, 32]. Defining β_i as

$$\beta_i \equiv \frac{1}{i \, c_i} \langle \frac{1}{k_{nn}} \rangle_i = \frac{\sum_k \frac{L_{i,k}}{k}}{\sum_k L_{i,k}} = \frac{\sum_k \frac{l_{i,k}}{k}}{\sum_k l_{i,k}} \;, \quad (27)$$

yields

$$\Theta_i = a c_i \; + \; (1-a)\beta_i i c_i \;. \quad (28)$$

To investigate a possible steady-state, we could substitute the above equations into Eqs. (2) and (4). However the non-linear terms resulting from β imply that a complete analytical solution for $l_{i,j}$ and c_i is difficult. We leave this as a future challenge, but stress that our formalism can be implemented in its non-stationary form Eq. (4) numerically by iteration with very good efficiency [30], yielding the degree distributions shown in Fig. 5.

We can now use the link-space formalism to deduce the parameter value at which our algorithm yields the BA degree distribution. At this value $a = a_c$, the node-degree distribution goes from over-skewed to under-skewed in relation to the BA model. For this parameter value, the attachment probability to nodes of various degrees is equal for both our mixture algorithm and the BA model. Using Eqs. (6) and (28), we have $\frac{i}{2} = a_c \; + \; (1 - a_c)\beta_i i$, and hence for large i this yields $a_c = 1 - \frac{1}{2\beta_i}$. We could then proceed to use the exact solution of the link-space equations for the preferential-attachment algorithm, in order to infer β_i in the high i limit. However, since β_i can be expanded in terms of $l_{i,j}$ as shown in Eq. (27) and $l_{i,j}$ decays very rapidly as i, j become large, we can obtain a good approximation by using only the first two terms of Eq. (8). This yields $\beta \sim 0.66$. Hence the critical
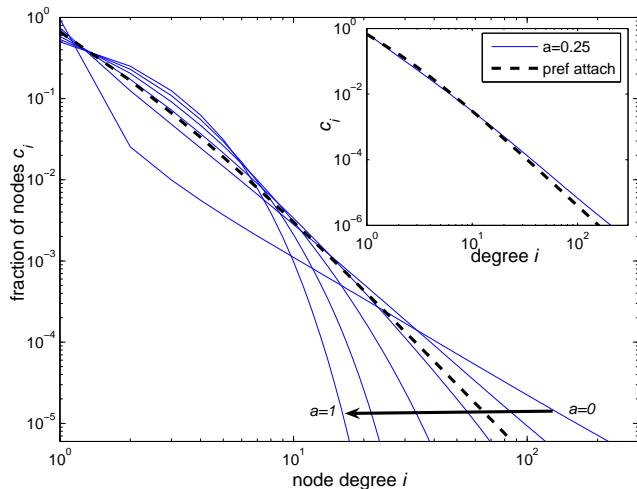
FIG. 5: (Color Online) Degree distributions generated using the link-space analysis of our one-step algorithm. Dashed line: results for the Barabási-Albert (BA) [2] preferential-attachment algorithm. Our one-step algorithm is able to reproduce these BA results with parameter value $a = 0.25$ (see inset).

value at which this simple model replicates preferential attachment is $a_c = 0.25$, in excellent agreement with the inset plot of Fig. 5.

VI. CONCLUSION

In conclusion, we have developed a new formalism which accurately accounts for the node-node linkage correlations in networks. We have employed the formalism to produce analytic solutions to the link-space correlation matrix for both the random attachment and BA models of network growth as well as deriving the form of a perfectly non-assortative network. We have also illustrated the possiblility of a steady state degree distribution and link-space for decaying networks. The link-space formalism has allowed us to accurately describe a simple one-parameter network growth algorithm which is able to reproduce a wide variety of degree distributions without any global information about node degrees.

Whilst the present paper focuses on illustrating the use of the link-space formalism for a variety of model networks, we stress that its applicability is far more general.
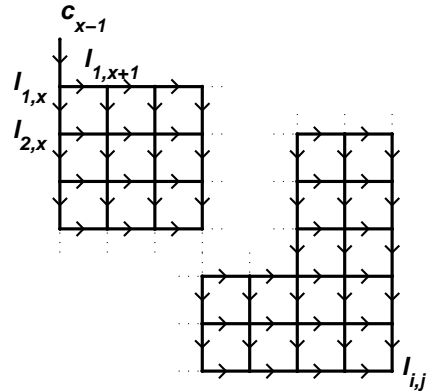
FIG. 6: The paths of probability flux from c_{x-1} influencing element $l_{i,j}$. Each arrow (step) represents a further factor of $\frac{1}{3}$.

APPENDIX A: EXACT SOLUTION OF RANDOM ATTACHMENT

The link-space master equation for the random attachment algorithm whereby one new node is added to the existing network with one new link per timestep can be expressd for he steady state as:

$$l_{i,j} = \frac{l_{i-1,j} + l_{i,j-1}}{3} \qquad i,j > 1$$
$$l_{1,j} = \frac{c_{j-1} + l_{1,j-1}}{3} \qquad j > 1$$
$$l_{1,1} = 0. \tag{A1}$$

It is easy to populate the link-space matrix numerically, just from the degree distribution $c_i = 2^{-i}$ obtained from solving the node-space recurrence relation and from the link-space master equation Eq. (A1).

At first glance, the solution to the master equation A1 would be of the form (check by substitution)

$$l_{i,j} = \frac{b}{4 \; 2^{i+j} \; 3^{i+j}}. \tag{A2}$$

However, the boundary conditions, which could be interpreted as influx of probability into the diffusive matrix, are such that this does not hold. We can actually solve the link-space for this model exactly. Consider the values c_i as being influxes of probability into the top and left of the link-space matrix. We can compute the effect of one such element on the value in the matrix $l_{i,j}$. Each step in the path of probability flux reflects an extra factor of $\frac{1}{3}$. First, we will consider the influx effect from the top of the matrix as in Fig. 6. The total path length from influx c_{x-1} to element $l_{i,j}$ is simply $i+j-x$. The number of possible paths between c_{x-1} and element $l_{i,j}$ can be expressed as $\binom{j-x+i-1}{j-x}$. We can similarly write down the
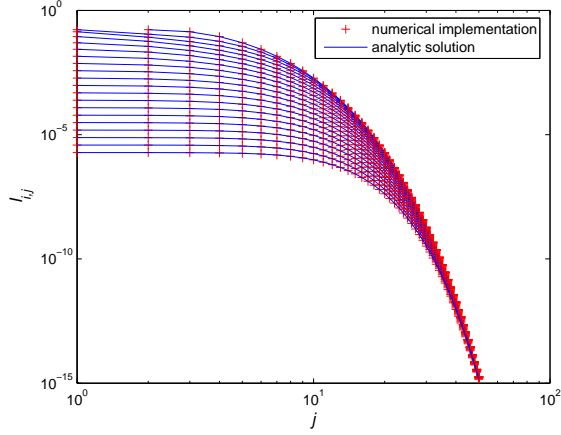
FIG. 7: (Color Online) Comparison of the numerically derived link-space matrix (markers) and the analytic solution (solid lines) for random node attachment.

paths and lengths for influxes into the left hand side of the matrix. As such, the first row (and column) can be described as

$$
\begin{aligned}
l_{1,j} &= \frac{c_{j-1} + l_{1,j-1}}{3} \\
&= \frac{2^{-(j-1)} + l_{1,j-1}}{3} \\
&= \sum_{k=1}^{j-1} \frac{1}{3^k 2^{j-k}}.
\end{aligned} \tag{A3}
$$

Subsequent rows can be similarly described. So, for $i, j > 1$ an element can be written:

$$
\begin{aligned}
l_{i,j} &= \frac{l_{i-1,j} + l_{i,j-1}}{3} \\
&= \sum_{\alpha=2}^{j} \frac{\binom{i-1+j-\alpha}{j-\alpha}}{3^{(i+j-\alpha)} 2^{(\alpha-1)}} + \\
&\quad \sum_{\alpha=2}^{i} \frac{\binom{i-1+j-\alpha}{i-\alpha}}{3^{(i+j-\alpha)} 2^{(\alpha-1)}}.
\end{aligned} \tag{A4}
$$

The results can be observed in Fig. 7.

APPENDIX B: EXACT SOLUTION OF PREFERENTIAL ATTACHMENT

In the BA model, when adding one new link with one undirected node per timestep, the steady-state solution of the node-space master equation leads to the degree distribution

$$
c_i = \frac{4}{i(i+1)(i+2)}, \tag{B1}
$$

which can be checked easily by substitution. Substituting the attachment probability kernel $\Theta_i \approx i c_i / 2$ into the

link-space master equation Eq. (4), we obtain the master equations for the link-space for this network growth algorithm:

$$
\begin{aligned}
l_{i,j} &= \frac{(i-1)l_{i-1,j} + (j-1)l_{i,j-1}}{2+i+j} \quad i, j > 1 \\
l_{1,j} &= \frac{(j-1)c_{j-1} + (j-1)l_{1,j-1}}{3+j} \quad j > 1 \\
l_{1,1} &= 0.
\end{aligned} \tag{B2}
$$

It is easy to populate the matrix numerically just by implementing the node and link-space equations. At first glance, the solution to the master equation Eq. (7) would be of the form (check by substitution)

$$
l_{i,j} = \frac{w}{i(i+1)j(j+1)}, \tag{B3}
$$

where w is a constant. This would imply for the degree distribution

$$
c_i = \frac{w}{i^2(i+1)}. \tag{B4}
$$

Summing over the entire node-space would give $w = \frac{6}{\pi^2 - 6}$. However, the boundary conditions (which could be interpreted as influx of probability into the diffusive matrix) are such that this solution does not hold. This is evident when comparing the node degree distribution with Eq. (B1), which compares to simulated networks well.

We can obtain an exact (although somewhat less pretty) solution by tracing fluxes of probability around the matrix and making use of the previously derived degree distribution. We can rewrite our link-space master equation, (B2), in terms of vertical and horizontal components:

$$
\begin{aligned}
l_{i,j} &= \Psi_{i,j} l_{i-1,j} + \Upsilon_{i,j} l_{i,j-1} \\
l_{1,j} &= \Upsilon_{i,j}\big(l_{1,j-1} + c_{j-1}\big) \\
l_{1,1} &= 0,
\end{aligned} \tag{B5}
$$

where trivially

$$
\begin{aligned}
\Psi_{i,j} &= \frac{\frac{\Theta_{i-1}}{c_{i-1}}}{1 + \frac{\Theta_i}{c_i} + \frac{\Theta_j}{c_j}} \\
\Upsilon_{i,j} &= \frac{\frac{\Theta_{j-1}}{c_{j-1}}}{1 + \frac{\Theta_i}{c_i} + \frac{\Theta_j}{c_j}}.
\end{aligned} \tag{B6}
$$

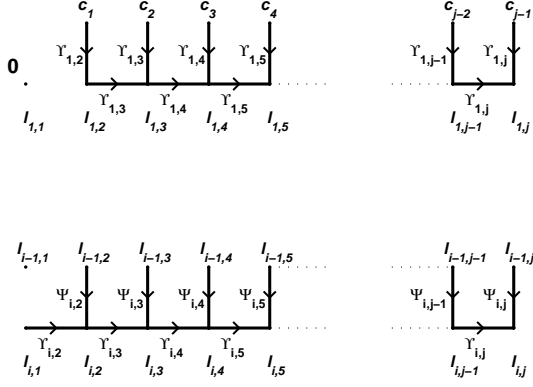By considering the probability fluxes as shown in Fig. 8, we can write the individual elements in the link-

FIG. 8: The components of flux of probability around the link-space matrix. This is general to any attachment kernel. An element within the link-space matrix can be built up from contributing elements and the appropriate factors.

space matrix as

$$
\begin{aligned}
l_{1,j} &= \sum_{\alpha=2}^{j}\left(c_{\alpha-1}\prod_{x=\alpha}^{j}\Upsilon_{1,x}\right)\\
l_{i,j} &= \sum_{\alpha=2}^{j}\left(l_{i-1,\alpha}\Psi_{i,\alpha}\prod_{x=\alpha+1}^{j}\Upsilon_{i,x}\right)\\
&\quad + l_{i,1}\prod_{x=2}^{j}\Upsilon_{i,x}\\
l_{1,1} &= 0.
\end{aligned}
\tag{B7}
$$

Note that we have yet to introduce the attachment probability kernels and the analysis so far is general. Using the preferential attachment probability

$$
\Theta_i \approx \frac{ic_i}{2},
\tag{B8}
$$

we can write our component-wise factors for the master equation as

$$
\begin{aligned}
\Psi_{i,j} &= \frac{i-1}{i+j+2}\\
\Upsilon_{i,j} &= \frac{j-1}{i+j+2}.
\end{aligned}
\tag{B9}
$$

Substituting Eq. (B9) into Eq. (B7) and using the previously derived degree distribution of Eq. (B1) yields for

the first row

$$
\begin{aligned}
l_{1,j} &= \sum_{\alpha=2}^{j}\left(\frac{4}{\alpha(\alpha-1)(\alpha+1)}\prod_{x=\alpha}^{j}\frac{x-1}{x+3}\right)\\
&= \frac{4(j-1)!}{(j+3)!}\sum_{\alpha=2}^{j}(\alpha+2)\\
&= \frac{2(j+6)(j-1)}{j(j+1)(j+2)(j+3)}.
\end{aligned}
\tag{B10}
$$

Subsequent rows can be written as

$$
\begin{aligned}
l_{i,j} &= l_{i,1}\frac{(j-1)!(3+i)!}{(2+i+j)!}\\
&\quad + \sum_{\alpha=2}^{j}l_{i-1,\alpha}(i-1)\frac{(j-1)!}{(2+i+j)!}\frac{(1+i+\alpha)!}{(\alpha-1)!}\\
&= \frac{(j-1)!}{(2+i+j)!}\left\{(3+i)!l_{i,1}\right.\\
&\quad + \left.(i-1)\sum_{\alpha=2}^{j}l_{i-1,\alpha}\frac{(1+i+\alpha)!}{(\alpha-1)!}.\right\}
\end{aligned}
\tag{B11}
$$

Rewriting this gives

$$
l_{i,j} = \frac{(j-1)!}{(2+i+j)!}\left\{K_i + E_{i-1,j}\right\},
\tag{B12}
$$

where the meaning of K_i and $E_{i-1,j}$ is transparent. Clearly, we can write Eq. (B12) for $l_{i-1,\alpha}$ as

$$
l_{i-1,\alpha} = \frac{(\alpha-1)!}{(1+i+\alpha)!}\left\{K_{i-1} + E_{i-1,\alpha}\right\}.
\tag{B13}
$$

Substituting Eq. (B13) into Eq. (B11) yields

$$
l_{i,j} = \frac{(j-1)!}{(2+i+j)!}\left\{K_i + (i-1)\sum_{\alpha=2}^{j}\left(K_{i-1} + E_{i-1,\alpha}\right)\right\}.
\tag{B14}
$$

In order to solve this recurrence relation, we define an operator for repeated summation, $S_{j,\alpha}^{n}$, such that

$$
\begin{aligned}
S_{j,\alpha}(f(\alpha)) &= \sum_{\alpha=2}^{j}f(\alpha)\\
S_{j,\alpha}^{n}(f(\alpha)) &= \sum_{\alpha_n=2}^{j}\sum_{\alpha_{n-1}=2}^{\alpha_n}\sum_{\alpha_{n-2}=2}^{\alpha_{n-1}}\cdots\sum_{\alpha_3=2}^{\alpha_4}\sum_{\alpha_2=2}^{\alpha_3}\sum_{\alpha=2}^{\alpha_2}f(\alpha).
\end{aligned}
\tag{B15}
$$

The first subscript denotes the initial variable to be summed over and the second the final limit. A few ex-

amples of this operation will clarify its use:

$$
\begin{aligned}
S_{j,\alpha}^0(f(\alpha)) &= f(\alpha) \\
S_{j,\alpha}(1) &= j - 1 \\
S_{j,\alpha}^2(1) &= S_{j,\alpha}(\alpha) - S_{j,\alpha}(1) \\
&= \frac{1}{2}(j^2 - j) \\
S_{j,\alpha}^3(1) &= \frac{1}{6}(j^3 - j) \\
S_{j,\alpha}(\alpha) &= \frac{j(j+1)}{2} - 1 \\
S_{j,\alpha}^2(\alpha) &= \frac{1}{6}(j^3 + 3j^2 - 4j)
\end{aligned}
\tag{B16}
$$

We can use this operator in our expression for the element $l_{i,j}$ in Eq. (B14) and expand to the easily derived value $E_{2,\alpha}$:

$$
\begin{aligned}
l_{i,j} &= \frac{(j-1)!}{(2+i+j)!}\left\{ K_i + (i-1)S_{j,\alpha}\big(K_{i-1} + E_{i-1,\alpha}\big)\right\} \\
&= \frac{(j-1)!}{(2+i+j)!}\left\{ K_i + (i-1)K_{i-1}S_{j,\alpha}(1) \right. \\
&\quad \left. +(i-1)(i-2)S_{j,\alpha}^2\big(K_{i-2} + E_{i-2,\alpha}\big)\right\} \\
&= \frac{(j-1)!}{(2+i+j)!}\left\{ K_i + (i-1)K_{i-1}S_{j,\alpha}(1) \right. \\
&\quad +(i-1)(i-2)K_{i-2}S_{j,\alpha}^2(1) \\
&\quad +(i-1)(i-2)(i-3)K_{i-3}S_{j,\alpha}^3(1) + ... \\
&\quad +(i-1)(i-2)(i-3)*...*2K_2 S_{j,\alpha}^{i-2}(1) \\
&\quad \left. +(i-1)(i-2)*....*2 S_{j,\alpha}^{i-2}(E_{2,\alpha})\right\}.
\end{aligned}
\tag{B17}
$$

We can express $E_{2,\alpha}$ in terms of the operator S too:

$$
E_{2,\alpha} = 4S_{\alpha,\alpha}\big(2S_{\alpha,\alpha}(1) + S_{\alpha,\alpha}(\alpha)\big).
\tag{B18}
$$

As such, the element $l_{i,j}$ can be expressed as

$$
\begin{aligned}
l_{i,j} &= \frac{(j-1)!}{(2+i+j)!}\left\{ \sum_{m=2}^{i} \frac{(i-1)!}{(m-1)!} K_m S_{j,\alpha}^{i-m}(1) \right. \\
&\quad \left. +4(i-1)!\big((2S_{j,\alpha}^i(1) + S_{j,\alpha}^i(\alpha)\big)\right\}.
\end{aligned}
$$

This form is somewhat obtuse as the calculation of the operator values is less than obvious. However, we can transform to a more easily interpreted operator $W(n)$ analogous to S but with different limits such that

$$
\begin{aligned}
W_{j,\alpha}(f(\alpha)) &= \sum_{\alpha=1}^{j} f(\alpha) \\
W_{j,\alpha}^n(f(\alpha)) &= \sum_{\alpha_n=1}^{j}\sum_{\alpha_{n-1}=1}^{\alpha_n}\sum_{\alpha_{n-2}=1}^{\alpha_{n-1}}\cdots\sum_{\alpha_3=1}^{\alpha_4}\sum_{\alpha_2=1}^{\alpha_3}\sum_{\alpha=1}^{\alpha_2} f(\alpha).
\end{aligned}
\tag{B19}
$$

Whilst, at first glance, it looks like little progress has been made, we only need evaluate the repeated operation on initial function $f(\alpha) = 1$. This is exactly solvable and as such, we can rewrite in terms of a function $G(n)$ dropping the superfluous alpha subscripts:

$$
\begin{aligned}
G_j(n) &= W_j^n(1) \\
G_j(n) &= \begin{cases} \frac{(j+n-1)!}{n!(j-1)!} & \text{for } n \geq 0 \\ 0 & \text{for } n < 0. \end{cases}
\end{aligned}
\tag{B20}
$$

The inductive proof associated with Eq. (B20) can be understood by path counting for some repeating binomial process and is an intrinsic property of the combinatorial choose coefficient. Consider a repeated coin toss over x steps. The number of ways of achieving $n + 1$ successes after these x iterations is $\binom{x}{n+1}$. Now, the occurrence of this last success could have happened on the $n + 1^{\text{th}}$ iteration or the following one, or any of the subsequent iterations till the xth one. For the last successful outcome to occur on the mth step, n successful outcomes must have occurred in the previous steps. The number of ways this could have occurred is $\binom{m-1}{n}$. Clearly, summing over all possible m values, the total possible paths resulting in $n + 1$ successes must equate to $\binom{x}{n+1}$.

For clarity, this is depicted in Fig. 9. To reach point **B** from **A** in the binomial process, one of the steps w, x, y or z must be traversed, after which there is only one route to **B**. As such, the number of paths between **A** and **B** utilising step w is the same as the number of paths between **A** and **W**. Similarly, the number of paths between **A** and **B** utilising step x is the same as the number of paths between **A** and **X** and so on. The number of paths between **A** and **B** can be built expressed as the sum of the paths $\mathbf{A} \rightarrow \mathbf{W}$, $\mathbf{A} \rightarrow \mathbf{X}$. $\mathbf{A} \rightarrow \mathbf{Y}$ and $\mathbf{A} \rightarrow \mathbf{Z}$.

We incorporate this behaviour into our proof for the solution of $G_j(n)$.

$$
\begin{aligned}
G_j(n) &= \binom{j+n-1}{n} \\
G_j(n+1) &= \sum_1^{j} G_j(n) \\
&= \sum_1^{j} \binom{j+n-1}{n} \\
&= \binom{j+n}{n+1}
\end{aligned}
\tag{B21}
$$

As $G(1) = \binom{j}{1} = j$, this inductive proof holds for all n. As such, the following relations hold:

$$
\begin{aligned}
S^n(1) &= G(n) - G(n-1) \\
S^n(j) &= G(n+1) - G(n-1)
\end{aligned}
\tag{B22}
$$

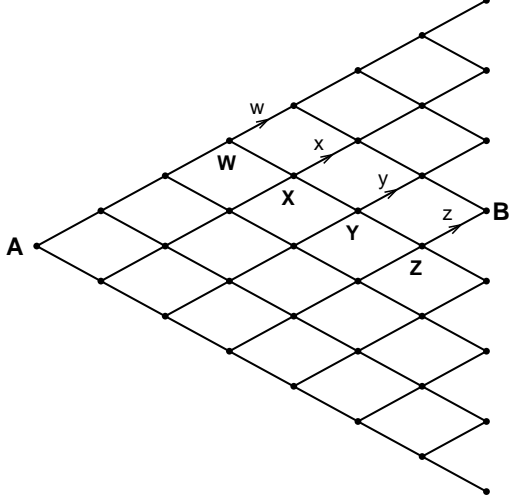We can now write the element in our link-space matrix for preferential attachment exactly as our function $G(n)$

FIG. 9: A binomial process over seven steps. The number of paths between **A** and **B** can thus be expressed as the sum of the paths $\mathbf{A} \to \mathbf{W}$, $\mathbf{A} \to \mathbf{X}$, $\mathbf{A} \to \mathbf{Y}$ and $\mathbf{A} \to \mathbf{Z}$.
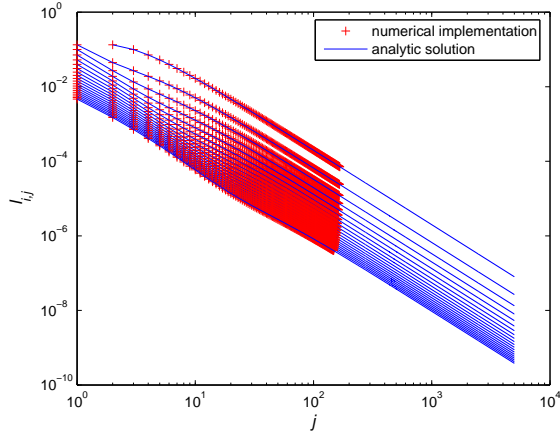

FIG. 10: (Color Online) Comparison of the numerically derived link-space matrix (markers) and the analytic solution (solid lines) for preferential attachment.

is easily evaluated.

$$
l_{i,j} = \frac{4(j-1)!(i-1)!}{(j+i+2)!}\left(G(i+1) + 2G(i) - 3G(i-1)\right.
$$
$$
\left. + \frac{1}{2}\sum_{k=2}^{i}(k-1)(k+6)\Big(G(i-k) - G(i-k-1)\Big)\right).
$$
$$
\tag{B23}
$$

APPENDIX C: DEGREE DISTRIBUTION OF DECAYING NETWORK

The link-space master equation for the steady state of the decaying networks of Section IV is written as

$$
l_{i,j} = \frac{il_{i+1,j} + jl_{i,j+1}}{i+j-2}. \tag{C1}
$$

As such, the number of $1 \leftrightarrow 1$ links does not reach a steady state for this process. This might be expected, as the removal process for such links requires them to be physically removed as opposed to the process by which the degree of the node at one end of the link being reduced or increased. As such, the value $l_{1,1}$ would increase in time. However, we can investigate the properties of the rest of the links in the network which do reach a steady state by neglecting these two-node components. In a similar manner to [20] we can make use of a substitution to find a solution to this recurrence equation, namely, for $i + j \neq 2$ as

$$
l_{i,j} = m_{i,j}\frac{(i+j-3)!}{(i-1)!(j-1)!}. \tag{C2}
$$

As such, we can obtain

$$
m_{i,j} = m_{i+1,j} + m_{i,j+1}. \tag{C3}
$$

This has a trivial solution

$$
m_{i,j} = \frac{A}{2^{i+j}} \tag{C4}
$$

The link-space for this system can be written for $i+j \neq 2$ as

$$
l_{i,j} = \frac{A}{2^{i+j}}\frac{(i+j-3)!}{(i-1)!(j-1)!}. \tag{C5}
$$

We can use the form of Eq. (C5) to infer the shape of the degree distribution. Recalling that

$$
c_i = \frac{M}{N}\frac{\sum_k l_{i,k}}{i}, \tag{C6}
$$

we can write for the degree distribution, neglecting the two node components in the network, as

$$
c_1 = \frac{MA}{N}\sum_{k=2}^{\infty}\frac{(k-2)!}{(k-1)!2^{1+k}}
$$
$$
c_i = \frac{MA}{Ni}\sum_{k=1}^{\infty}\frac{(i+k-3)!}{(i-1)!(k-1)!2^{i+k}}, \quad i > 1. \tag{C7}
$$

This can be easily solved by considering the Maclaurin Series [7] of the function $ln(1+x)$ (as a historical note, this was done by Mercator as early as 1668) and evaluating for $x = -\frac{1}{2}$:

$$
ln(1+x) = \sum_{k'=1}^{\infty}\frac{(-1)^{k'}}{k'}x^{k'}
$$
$$
ln\left(\frac{1}{2}\right) = -\sum_{k'=1}^{\infty}\frac{1}{k'\,2^{k'}}
$$
$$
= -ln(2). \tag{C8}
$$

Rearranging Eq. (C7) and letting $k = k' + 1$ yields

$$
\begin{aligned}
c_1 &= \frac{MA}{N} \sum_{k=2}^{\infty} \frac{1}{(k-1)2^{1+k}} \\
&= \frac{MA}{N} \sum_{k'=1}^{\infty} \frac{1}{k'2^{k'+2}} \\
&= \frac{MA}{4N} ln(2).
\end{aligned}
\tag{C9}
$$

For node degree greater than one, the solution of (C7) requires a simple proof by induction. Consider the function $Q(i')$ defined for positive integer i' as

$$
Q(i') = \sum_{k'=0}^{\infty} \frac{\binom{i'+k'}{k'}}{2^{k'+i'}}
\tag{C10}
$$

We can write $Q(i'+1)$ with similar ease as

$$
Q(i'+1) = \sum_{k'=0}^{\infty} \frac{\binom{i'+k'+1}{k'}}{2^{k'+i'+1}}.
\tag{C11}
$$

As such, we obtain

$$
\begin{aligned}
2Q(i'+1) - Q(i') &= \sum_{k'=0}^{\infty} \left(\binom{i'+k'+1}{k'} - \binom{i'+k'}{k'} \right) \frac{1}{2^{k'+i'}} \\
&= \sum_{k'=1}^{\infty} \binom{i'+k'}{k'-1} \frac{1}{2^{k'+i'}}.
\end{aligned}
\tag{C12}
$$

A quick substitution of $k'' = k' - 1$ leads to:

$$
\begin{aligned}
2Q(i'+1) - Q(i') &= \sum_{k''=0}^{\infty} \binom{i'+k''+1}{k''} \frac{1}{2^{k''+i'+1}} \\
&= Q(i'+1) \\
\Rightarrow \quad Q(i'+1) &= Q(i') \\
Q(1) &= 2 \\
&= Q(i'),
\end{aligned}
\tag{C13}
$$

which holds of for all i'. Rearranging Eq. (C7) and using some simple substitutions, $i' = i - 2$ and $k' = k - 1$, we can derive the following:

$$
\begin{aligned}
c_i &= \frac{MA}{Ni} \sum_{k=1}^{\infty} \frac{(i+k-3)!}{(i-1)!(k-1)!2^{i+k}} \\
&= \frac{MA}{Ni(i-1)} \sum_{k=1}^{\infty} \frac{(i+k-3)!}{(i-2)!(k-1)!2^{i+k}} \\
&= \frac{MA}{Ni(i-1)} \sum_{k'=0}^{\infty} \frac{\binom{i'+k'}{k'}}{2^{i'+k'+3}} \\
&= \frac{MA}{4N} \frac{1}{i(i-1)}.
\end{aligned}
\tag{C14}
$$

This degree distribution for the network without the two node components (the $1 \leftrightarrow 1$ links) can then be expressed as

$$
\begin{aligned}
A &= \frac{4N}{M(1+ln(2))} \\
c_1 &= \frac{ln(2)}{1+ln(2)} \\
c_i &= \frac{1}{(1+ln(2))i(i-1)},
\end{aligned}
\tag{C15}
$$

where the last line holds for all $i > 1$.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[2] A.L. Barabási and R. Albert, Science **286**, 509 (1999).

[3] A.L. Barabási, R. Albert and H. Jeong, Physica A **272**, 173 (1999).

[4] A. Barrat and R. Pastor-Satorras, Phys. Rev. E **71**, 036127 (2005).

[5] A. Barrat, M. Barthélemy and A. Vespignani, Phys. Rev. E **70**, 066149 (2004).

[6] V. Batagelj and A. Mrvar, Connections **21**, 47 (1998).

[7] M.L. Boas,*Mathematical methods in the physical sciences*, Wiley (2006).

[8] S. Boccaletti, V. Latorab, Y. Morenod, M. Chavezf and D.-U. Hwanga, Physics Reports 424, 175-308 (2006).

[9] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003).

[10] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman and S. H. Strogatz, Phys. Rev. E **64**, 041902 (2001).

[11] L. da F. Costa, F. A. Rodrigues, G. Travieso and P. R. Villas Boas, cond-mat/0505185.

[12] S.N. Dorogovtsev and J.F.F. Mendes, *Evolution of Networks* (Oxford University Press, Oxford, 2003).

[13] S.N. Dorogovtsev, J.F.F. Mendes and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).

[14] S.N Dorogovtesev and J.F.F. Mendes, Adv. Phys. **51**,

1079 (2002).

[15] V. M. Eguíluz and K. Klemm, Phys. Rev. Lett. **89**, 108701 (2002).

[16] T.S. Evans and J.P. Saramaki, Phys. Rev. E **72**, 26138 (2005).

[17] N. Farid and K. Christensen, New Journal of Physics **8** 212 (2006).

[18] P. Holme and B.J. Kim, Phys. Rev. E **65**, 026107 (2002).

[19] K. Klemm and V.M. Eguíluz, Phys. Rev. E **65**, 057102 (2002).

[20] P.L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).

[21] P. L Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).

[22] P.L. Krapivsky, S. Redner and F. Leyvraz, Phys Rev. Lett. **85**, 4629 (2000).

[23] L. Kullmann and J. Kertész, Phys. Rev. E **63**, 051112 (2001).

[24] M.E.J. Newman, A.L. Barabási and D.J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, Princeton, 2006).

[25] M.E.J. Newman, SIAM Rev. **45** 167 (2003).

[26] M.E.J. Newman Phys. Rev. Lett. **89**, 208701 (2002).

[27] R. Pastor-Satorras, A. Vázquez and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).

[28] M. Salathé, R.M. May and S. Bonhoeffer, RJ. R. Soc. Interface (2005) 2, 533ñ536.

[29] J.P. Saramaki and K. Kaski, Physica A **341**, 80 (2004).

[30] For mathematical details, derivations and further information, see D.M.D. Smith, C.-F. Lee, N.F. Johnson and J.-P. Onnela, physics/0702010 (2006).

[31] G. Szabó, M. Alava and J. Kertész, Phys. Rev. E **67**, 056102 (2003).

[32] A. Vázquez, Phys. Rev. E **67**, 056104 (2003).

[33] A. Vázquez, R. Pastor-Satorras and A. Vespignani, cond-mat/00112400 (2001).

[34] M. Brede and S. Sinha, cond-mat/0507710.

[35] M. di Bernardo, F. Garofalo and F. Sorrentino, cond-mat/0506236.

[36] M. di Bernardo, F. Garofalo and F. Sorrentino, cond-mat/0504335.

[37] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **66**, 047104 (2002).

[38] M. Catanzaro, G. Caldarelli and L. Pietronero, Phys. Rev. E **70**, 037101 (2004).

[39] J. Park and M. E. J. Newman, Phys. Rev. E **68**, 026112 (2003).

[40] J. Berg, M. Lässig and A. Wagner, Phys. Rev. E **4**, 51 (2004).

[41] S. N. Soffer and A. Vázquez, Phys. Rev. E **71**, 057101 (2005).

[42] S. Maslov and K. Sneppen, Science **296**, 910 (2002).

[43] R. Toivonen, J.-P. Onnela, J. Saramäki and J. Hyvönen and K. Kaski, Physica A **371**, 851 (2006).

[44] The master equation does not represent the evolution of an ensemble average of networks. Each specific realisation will have its own evolution of attachment kernel which cannot be descirbed by the ensemble average. The existence of a steady state to the master equation does not necessarily imply that specific realisations converge upon it in the large N limit but simply that the solution is static with respect to the growth algorithm. In practice, however, a master equation approach often yields good analytic agreement with even single realisations [30].

[45] A random walk of two steps in a network will be biased towards selecting low degree nodes. Previous studies have (unwittingly) relied on this phenomenon, using a two (or more) step random walk to select two (or more) nodes within the existing network for a new node to attach to with two (or more) links. The opposing biases effectively cancel such that the resulting algorithm can replicate preferential attachment. This was studied by Saramaki and Kaski [29] and independently by Evans [16]. For a more detailed analysis see [30].