I. INTRODUCTION

The study of complex networks has expanded rapidly over the past 20 years. Many real systems have been analyzed using networks with great success, showing many non-trivial properties. Model networks have been defined to understand the origin and development of these properties from elementary principles. For instance, the Watts-Strogatz model generates networks with short average path lengths but high clustering coefficients explaining the small world phenomenon of six degrees of separation phenomenon [1]. Similarly, the Barabási-Albert (BA) model, an undirected version of the Price model [2], demonstrates that scale-free (power-law) degree distributions in real networks can arise from a combination of growth and preferential attachment [3]. These models have given significant insight into the structure of real networks. However, real systems almost never reflect the exact details of a model.

One of the most common features to study in a real network is the degree distribution [4]. The degree, $k$, of a node in a network is the number of direct connections a node has to other nodes in the network. The degree distribution, $P(k)$, is the probability distribution of the degree across all the nodes in the network. The degree distribution is said to be scale-free (the exact definition is argued over) if it displays power-law scaling such that $P(k) \propto k^{-\kappa}$, where $\kappa$ is a positive constant, often in the range $\kappa \in \{2, 3\}$ for real networks [5]. Plotting $P(k)$ vs. $k$ on a log-log scale, a power-law distribution appears as a straight line with gradient $-\kappa$.

Since the late 1990s, many real networks have been reported as having scale-free, or nearly scale-free degree distributions. This includes web-page links on the internet [6], citation networks [7], the co-occurrence of words in language [8], sexual contact [9], social networks [10], and many more. Identifying these networks as scale-free has many important consequences: (1) it gives a potential mechanistic understanding of the origin and development of these networks, notably that the network evolves according to preferential attachment, and (2) it suggests these networks have a set of important properties associated with scale-free networks. These properties include the presence of hub nodes which have degree much larger than the network average, very small network diameters [11], and resistance to errors but vulnerability to targeted attack [12].

Although the scale-free paradigm has become a hallmark of complex networks research, identifying scale-free behavior in real networks is still very controversial [13].

Significant effort has gone towards developing sophisticated statistical techniques to assess whether networks at a fixed point in time are scale-free, most notably in [4, 14]. Given the difficulty in distinguishing a power-law distribution, $P(k) \propto k^{-\kappa}$, from similar distributions such as the log-normal, $P(k) \propto k^{-1}\exp\left[-\frac{(\ln k - \mu)^2}{2\sigma^2}\right]$, and stretched exponential, $P(k) \propto k^{-\kappa}e^{-\lambda k^\tau}$, distributions, where $\mu$, $\sigma$, and $\tau$ are the mean, standard deviation, and scale factor respectively, these statistical techniques are clearly important for understanding the statistical properties of network degree distributions. Applying such techniques to a large set of real world networks, Broido and Clauset [15] found that true scale-free networks are rare, represent-
ing only about 4% of all networks [4]. However, this result was immediately rejected by others in the networks community who took issues with the methods to process the data and/or the strictness of the scale-free definition [15,16]. Arguing that scale-free networks are only well defined in the infinite system size limit, looser definitions based on the the analysis of finite networks resulted in the finding that scale-free networks are in fact not rare at all [10]. Clearly, the issue of which approach is best when analyzing network degree distributions is yet to be fully resolved.

What all these approaches have in common is that they analyze the degree distribution of a network at a fixed point in time. If such an analysis is to give insight into the mechanistic origin and evolution of a network, it would be prudent to ask whether the degree distribution is representative of the network in general during its evolution, or only for a brief period of time? Without an answer to this question, inferring the past and future evolution of a network based on the current form of its degree distribution may give misleading results.

A prominent example of a theoretical network model where the observed degree distribution appears to change over time is super-linear preferential attachment, where new nodes attach to existing nodes proportional to their degree to a power greater than 1 [18]. In the long time limit, a gelation phenomenon is observed where almost all nodes connect to a single hub node forming a star-like network. However, Krapivsky and Krioukov [19] showed that super-linear attachment has significant pre-asymptotic regimes where the degree distribution appears to be approximately scale-free.

Given the difficulty of directly identifying preferential attachment from static degree distributions, proponents of the scale-free paradigm have argued that preferential attachment can be identified directly from dynamical network data (if available) [15,20]. Numerous approaches have been introduced over the years, using a variety of different assumptions [21,27]. Most commonly, methods assume that the preferential attachment kernel follows a functional form, \( \Pi(k) \propto k^\gamma \), and primarily focus on estimating the exponent \( \gamma \) - such methods will naturally assume that the preferential attachment kernel of a network is time independent.

As an alternative approach, non-parametric methods have been proposed that do not assume a functional form. The first of these methods by Jeong et al. [22] infers the form of the attachment kernel by constructing a histogram of the degree of nodes to which new edges attach over a short observation window. However, there is no clear guide as to how to choose the start of the observation window and how long it must be - too short and the result is very noisy, too long and the result is subject to bias [27]. The method by Newman [21] avoids this problem by constructing multiple histograms over different observation windows and computing the attachment kernel by taking a weighted average over the different histograms. Although this method avoids the issue of how to choose your observation window, this approach seems to systematically underestimate the attachment kernel at large degrees [28]. This issue has since been corrected by Pham et al. [27], who also provide a comprehensive technical overview of the methods by Jeong et al. [22] and Newman [21].

For networks in which the attachment kernel is truly time independent, the method proposed by Pham et al. [27] gives an excellent fit to data. However, it is still not clearly established whether the assumption of time independence is valid for real networks. Additionally, the probability of attaching to a node in a network may be a function of a variable other than the degree - the variable of interest may be related to the degree giving the false impression that the degree of a node is the property of interest, however, how to correctly identify which feature of a node determines its attractiveness is not clear.

It is often argued that accurately calculating the attachment kernel of a growing network is important because it can help to predict the future evolution of the network. For instance, in the case of non-linear preferential attachment, where the attachment kernel is given by \( \Pi(k) \propto k^\gamma \) with positive constant \( \gamma \), it is known that for \( 0 < \gamma < 1 \), the degree distribution tends to a stretched exponential, whereas for \( \gamma > 1 \), the degree distribution displays a gelation phenomenon where a single dominant hub connects to almost all other nodes in the network [18]. In between, \( \gamma = 1 \) corresponds to traditional linear preferential attachment where the degree distribution displays power-law scaling. Hence, if we can estimate the value of \( \gamma \) for the attachment kernel of a real network, this can be used to predict the future evolution of the network.

Predictions regarding the future evolution of networks, explanations of the historical development of networks, and investigations into whether preferential attachment underlies the evolution of networks, based on measured attachment kernels, are widespread in the literature. These include studies on citation networks [29,30], protein networks [31], the bitcoin network [32], the most common words in the English language [33], social dynamics in online games [34], actor networks [22], and more. The majority of these studies make three assumptions: (1) that the degree of a node is the key feature determining a node’s attractiveness, (2) that the attachment kernel can be approximated by \( \Pi(k) \propto k^\gamma \), and (3) that the measured attachment kernel is either time independent, or that the time dependence is largely unimportant. For instance, looking at four different periods in the evolution of the American Physical Society (APS) citation network, and using the node degree (citation count) as the key variable of interest, Sheridan and Onodera found that the exponent \( \gamma \) ranges from 0.94 to 1.06 [30]. The authors assert that this implies that the attachment probabilities in the APS citation network are at least approximately time independent. However, as noted, \( \gamma < 1 \) would imply that the APS citation network’s degree distribution approaches a stretched exponential, whereas \( \gamma > 1 \) would result in a gelation effect. Since both \( \gamma < 1 \) and \( \gamma > 1 \)
were observed from the data, what does this imply for the future evolution of the network?

The aim of this paper is to illustrate the risks of assuming time independence in the rules governing the evolution of growing networks, and the risk of assuming that the node degree is the observable determining node attractiveness in a network.

Firstly, we will demonstrate how very simple network growth rules can lead to networks where the form of both the attachment kernel, and the resulting degree distribution, are not fixed in time. We will do this by introducing the “k2 model”, a simple variant of the Barabási-Albert model where new nodes do not attach to existing nodes proportionally to the number of direct neighbors a node has, but rather proportionally to the number of nodes within a distance two of the target node. This simple rule is based on the idea that well connected neighbors provide greater utility to an individual target node than poorly connected neighbors. Although this simple rule has no explicit time dependence, the correlations that form between neighboring nodes result in an implicit time dependence in the attachment kernel that means that the resulting network does not demonstrate any of the simple scaling observed in traditional network models, despite initially appearing to grow as linear preferential attachment.

We support these findings with a mathematical argument demonstrating that the assumption of non-linear preferential attachment for an individual node i is consistent when averaging over the whole network, but not for individual nodes. Simulations suggest that there are two scaling regimes interacting for the evolution of individual nodes. For finite \( t_i \), the time a node was added to the network, and \( t \to \infty \), the degree of individual nodes grows over time and appears to approach \( t^{2/3} \) scaling. This results in the average value of the relative attachment kernel exhibiting approximately \( k^{4/3} \) scaling for the nodes with the largest degrees, \( k^{(1)} \). In the second regime where \( t_i \to \infty \), but \( t = t_i + \epsilon \) where \( \epsilon \ll t_i \), individual nodes exhibit \( t^{1/2} \) scaling for a transient period during which the degree evolution is indistinguishable from the scaling expected by the BA model. Interestingly, this seems to result in the relative attachment kernel exhibiting linear scaling in the degree, as expected by the BA model, for nodes with small degree. For nodes with intermediate values of the degree, we observe anomalous scaling which is unexplained by our mathematical arguments. This anomalous regime becomes more prominent as the network evolves.

Secondly, to demonstrate the importance of our results for real networks, we will look at the APS citation network in a little more depth to show that this attachment kernel is also time dependent. We stress that our interest in the k2 model lies in its ability to highlight the risks of assuming a time independent attachment kernel that is a function of node degree in the rules of network growth, not in the model itself. As a result, we will limit our analysis of the k2 model to features relevant to our current discussion.

In the remainder of this paper, we will introduce the k2 model and define the observables necessary to investigate the time dependence of the network attachment kernel. We will present results for the evolution of the degree distribution and attachment kernel for the k2 model and APS citation network, and show how these results suggest that neither the form of the degree distribution, nor the attachment kernel, are time independent. We leave analysis of the time dependence of other network observables to future work. Finally, we conclude by discussing the implications our results have for the analysis of real networks.

II. METHODS

IIa. Model Definition

The k2 model is defined as a simple, undirected network. The model is initialized with a small connected network of \( m_0 \) nodes. Each time-step, a new node is created with \( m \leq m_0 \) new edges. The \( m \) edges are connected to the new node and target nodes from the network. Each target node is chosen with probability proportional to the number of neighbors within one or two steps, \( k^{(2)} \), of the target node, \( i \). We refer to \( k^{(2)} \) as the second degree of node \( i \). The attachment probability is identical to the BA model with the exception that the BA model attaches proportionally to the number of nodes within one step, \( k^{(1)} \), of the target node, \( i \). We refer to \( k^{(1)} \) as the first degree, or simply the degree, of node \( i \). Computationally, we prevent multiple edges being formed between two nodes by ensuring that the \( m \) edges added to a new node attach to \( m \) unique target nodes.

Note for clarity, whenever notation is presented with a subscript \( i \) or \( j \), for instance \( k^{(1)}_i \) or \( k^{(2)}_j \), the focus is on the value of that variable for the particular node \( i \) or \( j \). When the subscript is omitted, for instance \( k^{(1)} \) or \( k^{(2)} \), the focus is on all nodes with the same specific value of the variable in question. To avoid clutter, we use \( k \) and \( k^{(1)} \) interchangeably where appropriate.

Formally, we can define \( k^{(\ell)}_i \) as the number of unique nodes which are \( \ell \) or fewer steps from the target node \( i \), excluding node \( i \) itself. Let \( \mathcal{N}_i(t) \) be the set of nodes which are distance \( \ell \) from node \( i \) in the network at time \( t \), that is \( \mathcal{G}(t) \) which is after all nodes and edges have been added and this has \( m_0 + t \approx t \) nodes. The distance between nodes \( i \) and \( j \) is defined as the minimum number of edges which need to be crossed in order to form a continuous path from node \( i \) to node \( j \). Then we define

\[
q^{(\ell)}_i(t) = |\mathcal{N}_i(t)|, \quad (1a)
\]

\[
k^{(\ell)}_i(t) = \sum_{j=1}^{\ell} q^{(j)}_i(t). \quad (1b)
\]
FIG. 1. A sketch of a simple tree network with four labeled nodes: a, b, c & d. The values in the brackets correspond to the first degree and second degree, \((k^{(1)}, k^{(2)})\), of each individual node. The four labeled nodes have the same degree, \(k^{(1)} = 3\). In the BA model, a new node is equally likely to attach to any one of the four labeled nodes. However, sociologically, we may expect an individual to value well-connected neighbors more than isolated neighbors. The k2 model is a very simple method of adjusting a node’s importance according to how well connected it’s neighbors are. By counting the number of neighbors within distance two of a node, \(k^{(2)}\), node a, with \(k^{(2)}_a = 9\), is now clearly more important than nodes b, c & d, with \(k^{(2)}_{b,c,d} = 5\).

\[ k^{(2)}_i = q^{(1)}_i + \sum_{\alpha=1}^{\infty} k^{(1)}_{i\alpha} \]  

where \(k^{(1)}_i(t) = q^{(1)}_i(t)\). In this paper we do not consider attachment kernel’s proportional to \(k^{(\ell)}(t)\) for \(\ell > 2\). However, it is interesting to note that if the attachment kernel were proportional to \(k^{(\ell)}(t)\) and \(\ell \geq D(t)\), where \(D(t)\) is the network diameter, this attachment kernel is equivalent to random attachment until the growing network has diameter \(D(t) > \ell\).

Figure 1 illustrates the sociological motivation for the k2 model. In the BA model, a node’s importance is proportional to the number of nodes connected to it, i.e. the first degree. However, there is no consideration for whether these connected nodes are important or not. A node with 100 neighbors, each with degree \(k = 1\), is considered equally important to a node with 100 neighbors, each with multiple neighbors. This form of secondary attachment has been demonstrated to have sociological importance and is rooted in ideas related to the strength of weak ties argument by Granovetter.

In the k2 model, a node’s next nearest neighbors also contribute to the node’s importance. This reflects the idea that the resource of a node is proportional to its direct neighbors only, but the influence of a node spreads beyond its direct neighbors. In the k2 model, a first and second (next-nearest) neighbor contribute equally to a node’s \(k^{(2)}\) value. Note, in the k2 model, first neighbors are indirectly more important to a node than second neighbors since influential first neighbors contribute many second neighbors to a node, however, influential second neighbors only contribute themselves to the node’s importance.

Mathematically, we define the attachment kernel, \(\Pi\), as the function specifying the probability of attaching to a specific node in the network. In the BA model, \(\Pi^{(BA)} \propto k^{(1)}\), whereas in the k2 model \(\Pi^{(k2)} \propto k^{(2)}\). In the case of the k2 model, we can write the normalized form of the attachment kernel as

\[ \Pi^{(k2)} = \frac{k^{(2)}_i}{\sum_{j=1}^{N} k^{(2)}_j} \approx \frac{k^{(2)}_i}{\sum_{j=1}^{N} (k^{(1)}_j)^2}. \]  

In the case of \(m = 1\), the approximation is an equality, however, it is approximately true for \(m > 1\) as long as the number of non-unique second degree neighbors is small. The validity of this assumption is shown in appendix V.V.c.

For illustrative purposes, it is convenient to split the numerator of the attachment kernel into the contribution of the first degree neighbors to node \(i\), \(q^{(1)}_i = k^{(1)}_i\), and the contribution of the next-nearest neighbors to node \(i\), \(q^{(2)}_i = k^{(2)}_i - k^{(1)}_i\). Additionally, we note that \(k^{(2)}_i\) can be written as a summation over the degrees of the first degree neighbors of node \(i\),

\[ k^{(2)}_i = \sum_{\alpha=1}^{\infty} k^{(1)}_{i\alpha} \]  

where \(\alpha\) labels the \(k^{(1)}_i\) unique first neighbors of node \(i\), and \(k^{(1)}_{i\alpha}\) is the first degree of node \(\alpha\), connected to node \(i\). Hence, Eq. (2) can be rewritten for node \(i\) as

\[ \Pi^{(k2)}_i = \frac{k^{(1)}_i + \sum_{\alpha=1}^{\infty} (k^{(1)}_{i\alpha} - 1)}{\sum_{j=1}^{N} (k^{(1)}_j)^2}. \]  

which is a function of the first neighbor degree only, commonly used in network science. In Eq. (4), the first term indicates the contribution to the attachment kernel from the direct neighbors of node \(i\), and the second term indicates the contribution from next-nearest neighbors to node \(i\).

It is possible to generalize the form of attachment shown in Eq. (4) by including a coefficient to the second term that adjusts the total weighting of next-nearest neighbors. This can be written as

\[ \Pi^{(k2)}_i = \frac{k^{(1)}_i + \delta \sum_{\alpha=1}^{\infty} (k^{(1)}_{i\alpha} - \delta)}{\sum_{j=1}^{N} (k^{(1)}_j + \delta \sum_{\alpha=1}^{\infty} (k^{(1)}_{j\alpha} - \delta))^2}, \]  

where \(\delta\) represents the influence of the next-nearest neighbors on the node's importance.
where we require $\epsilon \geq 0$. If $\epsilon = 0$, the k2 model reduces to the BA model, $\Pi^{(k2)}_i \rightarrow \Pi^{(BA)}_i$. Alternatively, if $\epsilon = 1$, $\delta = 1$, Eq. (5) reduces to the k2 model, Eq. (4). In this paper, to illustrate concerns about time invariance in the scaling of attachment kernels and degree distributions, we will only focus on the $\epsilon = 1$, $\delta = 1$, case shown in Eq. (4). Note that the general case presented in Eq. (5) is very closely related to the 2 levels model proposed by Dangalchev [37]. However, the 2 levels model double counts the first degree neighbors of node $i$, $\epsilon = 1$, $\delta = 0$, and in the analysis of the model, Dangalchev only looked at very small networks in which issues concerning the time invariance of the attachment kernel and degree distributions cannot be seen.

Conceptually, we can think of the k2 model as involving two separate networks. The first network is the observed network, labeled $G_{Obs}$, which corresponds to the conventional understanding of a network that would be generated by the k2 algorithm. In this representation, each node represents an agent, and an edge between two nodes represents a direct, first degree relationship between the two nodes. However, new nodes do not connect to a target node according to the node’s direct connections, but rather according to the number nodes within distance two of the target. We refer to these nodes as being within the sphere of influence of the target node. Hence, we define a second network as the influence network, labeled $G_{Inf}$, in which the nodes are equivalent to those in the observed network, but an edge between any two nodes signifies that the nodes are within each other’s sphere of influence, i.e., two nodes are connected in the influence network if the two nodes are neighbors, or next-nearest neighbors in the observed network. For illustrative purposes, edges in the influence network that exist (do not exist) in the observed network are indicated by a straight (curved) line.

![Observed Network](image1)

![Influence Network](image2)

FIG. 2. An illustration of how the k2 model can be thought of as generating two separate networks. The observed network (left) is the network of nodes which have a direct connection to each other. Here, the number on each node corresponds to the node’s first degree, $k^{(1)}_i$. In this network, the degree does not account for the importance of next-nearest neighbors. To adjust for this, we introduce the influence network (right). In this network, two nodes are connected (are in each other’s sphere of influence) if they are direct or next-nearest neighbors in the observed network. Note, that this effectively increases the importance of nodes which are central in the network. The number on each node corresponds to the degree in the influence network, or the second degree, $k^{(2)}_i$ in the observed network. For illustrative purposes, edges in the influence network that exist (do not exist) in the observed network are indicated by a straight (curved) line.

Distinguishing between the observed and influence network is useful because it allows us to draw parallels to other models of network growth. In particular, the influence network has strong similarities to the node copying mechanism studied by Lambiotte et al. [38, 39]. In the influence network, new nodes connect to a target node proportionally to the node’s degree, $k^{(2)}$. The new node then copies a fraction of the nodes attached to the target node, and forms additional edges to these copied neighbors. The neighbors of the target node which are copied correspond to those which are directly connected to the target node in the observed network. In the node copying model by Lambiotte et al. [38, 39], new nodes select a target node at random and then copy a fraction of the target node’s neighbors. As opposed to the k2 model, the copied neighbors are selected at random with probability $p$. In this respect, the node copying model where the original target node is chosen preferentially could represent a mean-field version of the k2 model, where we neglect correlations between neighboring nodes. Note however, drawing exact mathematical comparisons between the two models is difficult since it is not easy to estimate an
effective value for $p$ in the k2 model - this value is not fixed in the k2 model and will vary significantly based on the extent of clustering in a local area.

For our purposes, the introduction of the influence network will become clear in the mathematical analysis presented in section IIIb.

### IIIb. Measuring Time Invariance of Preferential Attachment

To understand how the attachment kernel of a network changes over time, it is helpful to consider relative attachment probabilities as opposed to absolute attachment probabilities. In general we can write an arbitrary attachment kernel, which is a function of the node degree only, as

$$\Pi(k; t) = \frac{f(k)}{\sum_{j=1}^{N(t)} f(k_j(t))}$$

(6)

for an arbitrary preference function $f$, where $k$ is a specific value of the degree and not the degree of a specific node $i$. The summation is over all nodes in the network. The function $f$ is time independent, however, as the network grows and more nodes are added, $N(t)$ in the denominator changes, and hence, the denominator is time dependent. Note, $f(k_i(t))$ for a specific node $i$ is not time independent, since the degree of a specific node evolves over time. Hence, we define the relative attachment kernel as

$$\phi(k, k'; t) = \frac{\Pi(k; t)}{\Pi(k'; t)} = \frac{f(k)}{f(k')}.$$  

(7)

If the attachment kernel is of the form given in Eq. (6), then the relative attachment kernel is time independent.

$$\phi(k, k') = \frac{\Pi(k; t)}{\Pi(k'; t)} = \frac{f(k)}{f(k')}.$$  

(8)

As opposed to $\Pi(k; t)$, the relative attachment kernel has no dependence on the network as a whole, but rather, is a function of the degree $k$ and $k'$ only. As a result, we can express the time dependence of the attachment kernel as

$$\frac{d\phi(k, k')}{dt} = 0.$$  

(9)

For convenience, in the following we will consider $\phi(k, 1)$, i.e., the attachment probability of connecting to a node with degree $k$ relative to a node with degree $k' = 1$. Note that, by definition, $\phi(k, k) = 1$.

Now that we have defined the relative attachment kernel, we can calculate this function for a network at different stages in its evolution. Consider the relative attachment kernel calculated at time $t$, written as $\phi_t(k, 1)$, and at time $s$, $\phi_s(k, 1)$. The subscript indicates the time at which the attachment kernel was calculated. If Eq. (9) holds, then $\phi_t(k, 1) = \phi_s(k, 1)$. For a real network, it is likely that there will be small deviations from this ideal case. Hence, we can plot the ratio of the preference functions $\phi_t(k, 1)/\phi_s(k, 1)$ against degree $k$ to gauge the extent of the time dependence. Note that the ratio is only well defined for networks which contain nodes with degree $k$ at both times $t$ and $s$. For a truly time invariant preference function, $\phi_t(k, 1)/\phi_s(k, 1) = 1$ for all $k$.

The BA model is a classic case where Eq. (9) should hold. For this particular example, $\phi_t(k, 1) = k$ for all $t$. Likewise, for non-linear preferential attachment, $\phi_t(k, 1) \propto k^\gamma$ with positive constant $\gamma$. In the case of the k2 model, Eq. (9) does not hold. Why is this the case? The key lies in the second term of Eq. (4). The preference function in the k2 model is not a function of the degree of a node, but the second degree, $k(2)$, of a node. Clearly the second degree is related to the first degree, and when analyzing the network, one could mistakenly believe that the node degree is the quantity of importance for the growth of the network. However, over time, the relation between the first and second degree changes such that correlations form in the network. These correlations mean that although the attachment kernel is not explicitly time dependent, the local network structure which determines the number of second degree neighbors for a given node $i$ is time dependent.

### IIc. Comparison to Real Data

To demonstrate the relevance of our work to real data, we investigate the evolution of the American Physical Society (APS) journals paper citation network containing over 450,000 citing articles from 1893 to 2003. The network contains over 3,000,000 directed edges representing one paper citing another. Each article in the network has a publication time stamp and, therefore, the full time evolution of the network is known.

The network has been investigated extensively over many years and has been the key focus of several studies investigating the relationship between preferential attachment and the resulting degree distribution. Notably, Redner illustrated the “preferential attachment paradox” using the APS citation network. The paradox can loosely be explained as follows. It is known that linear preferential attachment generates power-law degree distributions. However, when analyzed, the preferential attachment in the APS citation network appears to generate a log-normal degree distribution. How is this possible? Several answers have been put forward, for example see [12, 45], but we shall use Redner’s approach here.

It is important to note that when illustrating the preferential attachment paradox, Redner first had to deduce the form of the attachment kernel and the form of the degree distribution. The degree distribution of a network at time $t$ can be measured exactly. However, the attach-
ment kernel is much harder to measure. Whenever a new article is added to the citation network, that article either cites, or does not cite, the articles in the preceding network. From this information, we cannot exactly deduce the attachment kernel, that is, the probabilities with which each node in the preexisting network was chosen or not chosen. This motivates the development of statistical techniques to deduce the form of the attachment kernel by aggregating information over time as discussed in section II. When demonstrating the paradox, Redner applied the method by Jeong et al. to suggest the functional form of the attachment kernel is linear, and hence, that the attachment kernel and the degree distribution are not consistent [22, 31].

Since then, with updated statistical methods, Sheridan and Onodera [30] claim to have resolved the preferential attachment paradox by showing that the attachment kernel of the APS citation network is in fact consistent with a log-normal degree distribution [30]. Note however, that the more sophisticated methods of Sheridan and Onodera are reliant on the assumption that the attachment kernel is time independent, and the assumption that the attachment kernel primarily depends on the first degree of a node. It is precisely these assumptions that this paper contends with. The mismatch between the observed degree distributions and attachment kernels discussed by Redner highlights the risks of inferring the functional form of one from the other.

Sheridan and Onodera do briefly allude to the importance of a time invariant attachment kernel, and measure the kernel for four separate time periods in the network’s evolution, 1901–74, 1974–88, 1988–95 and 1995–2000. From this, assuming an attachment kernel of the form \( \Pi(k) \propto k^\gamma \), they calculate values for \( \gamma \) of 0.97, 0.94, 1.05 and 1.06, and argue that this suggests that the time invariant assumption holds at least approximately true. This may well be the case – it is likely that if an attachment kernel has an implicit time dependence that the effect is small. However, for \( \gamma < 1 \), we expect to observe a stretched exponential degree distribution, whereas for \( \gamma > 1 \) we expect to observe a gelation phenomenon. So, given that we observe both \( \gamma < 1 \) and \( \gamma > 1 \) in the APS citation network, can we really predict what form the degree distribution of the APS citation network will have in the future? We would argue no, however, some authors have made such predictions based on measurements of network attachment kernels. This warrants a more careful analysis of the time dependence of network attachment kernels.

III. RESULTS

IIIa. Simulation Results

In the following, we will focus on analyzing the attachment kernel and the degree distribution of the k2 model, and the BA model as comparison. Other common net-

FIG. 3. The degree distributions for the k2 model for (a) \( m = 1 \), (b) \( m = 2 \) and (c) \( m = 3 \), during the network evolution, averaged over 100 simulations. Error bars are omitted for visual clarity. The dashed lines show the expected scaling for the BA model. Three features are observed in the development of the degree distribution. (1) The initial scaling of the k2 model is indistinguishable from the BA model. (2) As the k2 model grows, an instability forms which results in a pronounced deviation from the BA model scaling as the network gets larger. (2) This deviation is suppressed as \( m \) is increased, although the deviation is still present. As a result, for large \( m \) and small \( t \), the scaling of the k2 model is almost indistinguishable from the BA model. For small \( m \) and large \( t \), there is significant deviation between the scaling of the k2 model and the BA model.
work features such as the scaling of the diameter or the clustering of nodes we will leave for future work. We note that the size of simulations are not limited by the speed at which the networks can be generated but rather the memory required to store the second degree neighbors of nodes during the evolution of the k2 model. Each simulation is initialized with a complete graph of $m_0 = m + 1$ nodes.

Figure 4 and Fig. 5 show the degree distribution and the true relative attachment kernel for the k2 model for various $m$ at different stages in the network evolution. Both figures are averaged over 100 simulations of the k2 model. The dashed lines show the expected scaling of the degree distribution and the relative attachment kernel in the BA model. Early in the network development, there are only small differences between the behavior of the k2 model and the BA model. However, as the networks grow, a significant deviation between the k2 and BA models forms. This deviation is suppressed for larger $m$ but still present.

The deviation is most clearly seen as an excess in the probabilities of observing a node with moderate degree $k$ in the degree distribution. The degrees effected range from $k \approx 10$ to $k \approx 300$. This range is also where an excess is observed in the relative attachment probabilities. This indicates that there are multiple timescales of interest at play in the evolution of the network. Over short timescales, the growth of the k2 model appears to be close...
The true relative attachment kernel for the k2 model shown in Fig. 4 is not typically accessible for a real network. In the case of the k2 model, we know the attachment kernel because for each node, \(i\), we know both the first and second degree, \(k_i^{(1)}\) and \(k_i^{(2)}\) respectively. Figure 4 shows that for small networks, the relative attachment kernel is approximately linear. However, as the networks grow, the relative attachment kernels become increasingly non-linear. In the case of \(m = 1\), shown in Fig. 4(a), for the network at \(t = 10^6\), a node with degree \(k \approx 30\) is almost a factor of 10 more attractive, relative to the same node with degree \(k = 1\), compared to the equivalent for the network at time \(t = 10^3\). This excess in the relative probabilities grows over time for all \(m\), although the excess is significantly smaller for increasing \(m\). However, even for larger \(m\) the deviation from linear attachment kernels grows with time. Note, that if we were to fit
FIG. 7. (a) The relative attachment kernel, $\phi_t(k, 1)$, for a single simulation of the BA model with $m = 1$ as calculated using the corrected Newman method. The calculated attachment kernel is shown for the network at $t = 10^5$ and $t = 10^6$. (b) The ratio of the $t = 10^5$ to the $t = 10^6$ relative attachment kernels. Note, the ratio is only defined at $k$ values where both $\phi_{10^5}$ and $\phi_{10^6}$ are non-zero. (c) The cumulative sum of the relative attachment kernels shown in (a), $\phi_t(k, 1)$. (d) The ratio of the $t = 10^5$ to the $t = 10^6$ cumulative relative attachment kernels. The dashed lines in (a) and (c) indicate the relative attachment kernel expected for linear preferential attachment. The dashed lines in (b) and (d) indicate the ratio expected if the relative attachment kernels are time independent. The close agreement between the dashed lines and the data indicates that the relative attachment kernel is time independent for the BA model.

the relative attachment kernel, on the assumption that $\phi_t(k, 1) \propto k^\gamma$ for a positive exponent $\gamma$, and we were to omit data for large $k$ as is often done for large networks [30], we may deduce that for $t = 10^3$, the k2 model has an approximately linear, attachment kernel, whereas at $t = 10^6$, the attachment kernel is highly non-linear. If we were to use these results to infer the future scaling of the network, the data at $t = 10^3$ would suggest the network might approach a stretched exponential degree distribution, whereas from the data for $t = 10^6$, we might paradoxically infer the network is approaching a gelation state. This situation is similar to the paradox tackled by Sheridan and Onodera [30] when analyzing the APS citation network.

For added clarity, we can plot the average value of the second degree, $\bar{k}^{(2)}(k; t)$, for the nodes in the network with specific values of the degree, $k^{(1)}$, see Fig. 5. This is defined as

$$\bar{k}^{(2)}(k; t) = \frac{1}{n(k; t)} \sum_{j | k_j^{(1)} = k^{(1)}} k_j^{(2)}$$  \hspace{1cm} (10)$$

where the sum is over all nodes in the network with a specific value of the degree, $k^{(1)}$, and $n(k; t)$ is the number of nodes in the network with degree $k^{(1)}$. Note that this is equivalent to the numerator of the attachment kernel in the k2 model. The figure indicates that for all $k^{(1)}$, the average corresponding value of $k^{(2)}$ grows over time. Additionally, the figure seems to indicate two scaling regimes separated by a third cross-over region. For nodes with large $k^{(1)}$, the scaling of the second degree appears largely time independent and approximately scales as $k^{4/3}$ where the origin of the $4/3$ exponent is discussed in section IIIb, see dotted line. For nodes with small $k^{(1)}$, the second degree is time dependent, growing as the network itself grows. However, the scaling appears
approximately linear as expected for the BA model, see dashed line. Separating these regimes is a cross-over region at moderate $k^{(1)}$ showing strong time dependence. Initially there is super-linear growth in the second degree as a function of the first degree, before an extended region where the second degree appears to plateau.

In the case of a real network, we can only estimate the relative attachment kernel by observing the degree of nodes to which new nodes added to the network attach. To simulate this real-network scenario, we apply the corrected Newman method to a single simulation of the k2 model as shown in Fig. 6. Figure 6(a) shows the calculated relative attachment kernel for the k2 model at times $t = 10^5$ and $t = 10^6$. As in Fig. 4 nodes with moderate degree, $k \approx 30$, show an excess in the relative attachment kernel. In Fig. 6 deviations in the relative attachment kernel are shown explicitly by taking the ratio to the relative attachment kernels at $t = 10^5$ and $t = 10^6$. For very small degree nodes, the ratio is approximately one indicating that the attachment kernel is time independent at these degrees. Above $k = 10$, the ratio clearly deviates from one, indicating that the relative attachment kernel is not time independent. For visual clarity, and to show the effect of nodes with large degree, Fig. 6(c) & (d) show the equivalent as (a) & (b) but for the cumulative sum of the relative attachment kernel. The cumulative sum is defined as

$$
\tilde{\phi}(k, 1) = \sum_{\tilde{k} < k} \phi(\tilde{k}, 1).
$$

(11)

It is interesting to note that the estimates attachment kernel using Newman’s method is not fully consistent with the true attachment kernel shown in Fig. 4. This is because Newman’s method constructs the attachment kernel by collating multiple histograms from different times in the network evolution. The consequence is that the estimated form of the attachment kernel at $t = 10^6$ is more consistent with the true attachment kernel earlier in the evolution of the k2 model, rather than the current value of the attachment kernel.

To verify that the deviations in the relative attachment kernel are due to the evolution of the k2 model and not numerical errors, we repeat the analysis shown in Fig. 6 for a network where the relative attachment kernel is expected to be time independent. This is shown in Fig. 7 for the BA model at $t = 10^5$ and $t = 10^6$. Figure 7(a) shows that the relative attachment kernels are effectively indistinguishable for the BA model at different times in the network evolution. This is confirmed by Fig. 7(b) where the ratio of the relative attachment kernels is approximately one for all nodes with degree $k < 100$. Noise in the tail obscures the ratio for $k > 100$. Figures 7(c) & (d) show the equivalent as (a) & (b) but for the cumulative sum of the relative attachment kernel. The ratio of the cumulative relative attachment kernel is almost exactly one for all degrees except for the very largest observed in the network. However, Fig. 7(c) shows that the deviation from linear scaling expected by the BA model moves to larger $k$ as the network evolves in time. This suggests that this deviation is a finite-size effect. Overall, Fig. 7 clearly indicates that using the corrected Newman method is effective for estimating the relative attachment kernel and that the relative attachment kernel is time independent for the BA model, as expected from theory. This also suggests that the deviations in the relative attachment kernel observed in Fig. 6 are due to the structural properties of the k2 model, and not due to limitations in the method used to estimate the relative attachment kernel. Hence, we can deduce that the relative attachment kernel for the k2 model is not time independent.

IIIb. Mathematical Results

Given the complexity of the correlations that form over time in the k2 model, it is not clear how to derive an exact analytical expression for the evolution of the k2 model. However, using our knowledge of the properties of the k2 model, we can test whether the assumption of linear preferential attachment could be consistent with the mathematical properties of the k2 model.

In the following we will apply a continuum approach and consider the general case of $m \geq 1$, although we note that for $m > 1$ our result is at best an approximation. From the definition of the k2 model, we can write the evolution of the degree, $k_i^{(1)}(t)$, of a given node $i$ as

$$
\frac{dk_i^{(1)}(t)}{dt} = m \Pi_i^{(k2)}(t),
$$

(12a)

$$
\Pi_i^{(k2)}(t) \equiv \frac{k_i^{(2)}(t)}{\sum_{j=1}^{N} k_j^{(2)}(t)} \approx \frac{k_i^{(2)}(t)}{\sum_{j=1}^{N} \left( k_j^{(1)}(t) \right)^2},
$$

(12b)

where $k_i^{(1)}(t)$ is the degree of node $i$ at time $t$, which was added to the network at time $t_i(\leq t)$. The second degree, $k_i^{(2)}(t)$, is defined according to Eq. 3, the summation is over all nodes in the network, and the approximation is and equality if $m = 1$ as proven in the appendix. We can write the evolution of the second degree as

$$
\frac{dk_i^{(2)}(t)}{dt} \approx m \left( \frac{k_i^{(1)}(t)}{\sum_{j=1}^{N} \left( k_j^{(1)}(t) \right)^2} \right)^2 + \xi_i(t),
$$

(13)

where

$$
\xi_i(t) \equiv \sum_{\beta=1}^{k_i^{(1)}(t)-1} k_{i_1(\beta)}^{(1)}(t) + \cdots + \sum_{\beta=1}^{k_i^{(1)}(t)-1} k_{i_\alpha(\beta)}^{(1)}(t).
$$

(14)

The derivation of Eq. 13 is shown in the appendix. Again, the approximation is an equality in the case of $m = 1$, however, for $m > 1$ the denominator should be
approximately true as long as the number of non-unique second degree neighbors (a second degree neighbor that is connected to multiple first degree neighbors or is a first degree neighbor itself) is small relative to the number of unique second degree neighbors. Here $i_{\alpha(\beta)}$ represents the node $\alpha$ connected to node $i$. Equation (14) represents the effect from the outside of the system. Since the first degree of a node can only grow over time, $\xi_i(t)$ is a positive semi-definite monotone increasing function with respect to time $t$, that is,

$$\xi_i(t) \geq 0, \quad (15a)$$

$$\frac{d}{dt} \xi_i(t) \geq 0. \quad (15b)$$

Our aim in the following is to write $\xi_i(t)$ as a function of the first degree, $k^{(1)}$, only. To do so we can rearrange Eq. (13) to make $\xi_i(t)$ the subject,

$$\xi_i(t) = \frac{1}{m} \left( \sum_{j=1}^{N} \left( k^{(1)}_j(t) \right)^2 \right) \cdot \frac{dk^{(2)}_i(t)}{dt} - \left( k^{(1)}_i(t) \right)^2. \quad (16)$$

We can also rearrange Eq. (12a) and Eq. (12b), and take the derivative which respect to time such that

$$\frac{dk^{(2)}_i(t)}{dt} = \frac{d}{dt} \left[ \frac{1}{m} \left( \sum_{j=1}^{N} \left( k^{(1)}_j(t) \right)^2 \right) \cdot \frac{dk^{(1)}_i(t)}{dt} \right]. \quad (17)$$

Substituting Eq. (17) into Eq. (16) we find

$$\xi_i(t) = \frac{1}{m} \left( \sum_{j=1}^{N} \left( k^{(1)}_j(t) \right)^2 \right) \cdot \frac{d}{dt} \left[ \frac{1}{m} \left( \sum_{j=1}^{N} \left( k^{(1)}_j(t) \right)^2 \right) \cdot \frac{dk^{(1)}_i(t)}{dt} \right] - \left( k^{(1)}_i(t) \right)^2, \quad (18)$$

which is a function of the first degree only. Here we note that the summations in Eq. (18) correspond to the denominator in Eq. (12). This is the sum over $k^{(2)}_i(t)$ for each node in the network, and hence, corresponds to the twice the total number of edges in the influence network introduced previously. Therefore, for convenience, we introduce the following notation for the number of edges in the influence network as

$$E^{(2)}(t) = \frac{1}{2} \sum_{j=1}^{N(t)} k^{(2)}_j(t) \approx \frac{1}{2} \sum_{j=1}^{N(t)} \left( k^{(1)}_j(t) \right)^2 \quad (19)$$

where the approximation is an equality in the case of $m = 1$.

Figure 3 and Fig. 4 show the degree distribution and the rescaled attachment kernel for the k2 model obtained from simulations. As a thought experiment, let us suppose that these simulations are not for a theoretical network model but that the data represents a real world network. For the network at small times in its evolution, the degree distribution and the attachment kernel both look very similar to the degree distribution and attachment kernel expected for the BA model. Observing the similarity between the BA model and our observed network, we might naively expect that our network is growing according to linear preferential attachment.

For preferential attachment (linear or non-linear), it is known that, on average, the degree of a given node $i$ evolves in time as a power function given by

$$k^{(1)}_i(t) = m \left( \frac{t}{t_i} \right)^\delta, \quad t \geq t_i, \quad (20)$$

where $t_i$ is the time at which node $i$ was added to the network, and $\delta = 1/2$ for linear preferential attachment [46].

In the case of sub- (super-) linear preferential attachment, $\delta < 1/2$ ($\delta > 1/2$). Let us assume Eq. (20) holds for our network and test whether this form for the evolution of the degree of node $i$ is consistent with the mathematical form of the k2 model. First, we substitute Eq. (19) into Eq. (18),

$$\xi_i(t) = \frac{2}{m} E^{(2)}(t) \frac{d}{dt} \left[ \frac{2}{m} E^{(2)}(t) \frac{d}{dt} \left( k^{(1)}_i(t) \right) \right] - \left( k^{(1)}_i(t) \right)^2. \quad (21)$$

In the case of the k2 model where one node is added to the network at each time step, we initialize our network such that $t_j = j$ and note that the number of nodes in the network at time $t$ is given by $N(t) = m_0 + t \approx t$.

Using this initialization, we now calculate the value of $E^{(2)}(t)$ by approximating the sum as an integral and substituting in Eq. (20),

$$E^{(2)}(t) \approx \frac{1}{2} \sum_{j=1}^{t} \left( k^{(1)}_j(t) \right)^2 \approx \frac{m^2}{2} \int_1^t dt' \left( \frac{t'}{\bar{v}} \right)^{2\delta} \approx \frac{m^2}{2(1-2\delta)} \left[ \left( t' \right)^{1-2\delta} \right]_1^t. \quad (22)$$

There are three cases for the different possible values of $\delta$.

Case (i) $2\delta < 1$. Corresponding to sub-linear preferential attachment, this scenario is likely to be irrelevant for the k2 model since the influence network cannot grow slower than the original network in the BA model. However, in this case you would find linear growth in the number of edges in the Influence network

$$E^{(2)}(t) \approx \frac{m^2}{2(1-2\delta)} t. \quad (23)$$
Here $E^{(2)}(t)$ is dominated by the youngest nodes (created at the largest times $t_i$) as the older nodes grow too slowly.

Case (ii) $2\delta = 1$. Corresponding to linear preferential attachment, this would be the case in the BA model,

$$E^{(2)}(t) \approx \frac{m^2}{2} t \ln(t). \quad (24)$$

Case (iii) $2\delta > 1$. Corresponding to super-linear preferential attachment where there is some enhancement over linear preferential attachment. For the $k_2$ model, this scenario is plausible since we know that for any given node $k_i^{(2)}(t) \geq k_i^{(1)}(t)$. In this case we find

$$E^{(2)}(t) \approx \frac{m^2}{2(2\delta - 1)} (t^{2\delta} - t), \quad (25)$$

where the growth in the number of edges in the influence network is dominated by the oldest nodes in the network. Note, we have kept the linear term in Eq. (25) for reasons which will become apparent later in this section.

Let us assume case (iii) is valid for the $k_2$ model. Substituting Eq. (25) into Eq. (21) we find,

$$\xi(t) = \frac{m^3}{(2\delta - 1)^2} \left( \frac{1}{t_i} \right)^\delta \left[ (3\delta - 1)t^{5\delta - 2} - (4\delta - 1)t^{3\delta - 1} + \delta t^\delta \right] - m^2 \left( \frac{t}{t_i} \right)^{2\delta}$$

$$= a_1 t^{5\delta - 2} + a_2 t^\delta - a_3 t^{3\delta - 1} - a_4 t^{2\delta}, \quad (26)$$

where in the final line we have grouped all the constants for each term into a single positive prefactor, $a_1$ to $a_4$. Recall that the $k_2$ model requires that $\xi(t)$ is a positive, semi-defined monotonically increasing function, and note that Eq. (26) is only valid for $\delta > 1/2$.

As $t \to \infty$, the first term of Eq. (26) will dominate the second if $5\delta - 2 > \delta, \delta \geq 1/2$. Likewise, the third term will dominate the fourth if $3\delta - 1 \geq 2\delta, \delta \geq 1$. Hence, as $t \to \infty$, the first term is the dominant positive term and the fourth term is the dominant negative term. To ensure $\xi_i(t) \geq 0$ for all $t > t_i$, this requires the first term to grow faster than the fourth term giving $5\delta - 2 > 2\delta$, corresponding to $\delta \geq 2/3$.

Returning to Eq. (13) and substituting in Eq. (26) and Eq. (25), we can also write

$$\frac{d k_i^{(2)}(t)}{dt} = m(2\delta - 1) \frac{m^2 (t/t_i)^{2\delta} + \xi_i(t)}{m^2(t^{2\delta} - t)}. \quad (27)$$

We have established above that to satisfy Eq. (15a), the leading term of $\xi_i(t)$ must scale as $t^{5\delta - 2}$, and $\delta \geq 2/3$. However, as a consequence of the rules of the $k_2$ model, at time $t > t_i$, node $i$ can gain no more than $m$ new edges in the influence network in any given time step (i.e. $k_i^{(2)}(t + 1) - k_i^{(2)}(t) \leq m$). Hence, strictly for $t > t_i$, we require

$$\frac{d k_i^{(2)}(t)}{dt} \leq m, \quad (28)$$

which is only satisfied if the denominator of Eq. (27) grows at least as fast as the numerator of Eq. (27). This requires $t^{2\delta} \geq t^{5\delta - 2}$ as $t \to \infty$. Hence, $2\delta \geq 5\delta - 2$, giving $\delta \leq 2/3$. Combining the conditions in Eq. (15a) and Eq. (28), we find that a power function of the form given in Eq. (20) can only satisfy the requirements of the $k_2$ model if $\delta = 2/3$.

To test the validity of our argument, we simulate the growth in the number of edges in the influence network using the $k_2$ model. This is shown in Fig. 8 for $m = 1, 2, 3$. The figure shows that, at large $t$, the number of edges in the influence network scales as approximately $t^{4/3}$ corresponding to $\delta = 2/3$, in agreement with our mathematical prediction.

Although the results so far give some indication that the growth of the $k_2$ model has a simple scaling form, further analysis appears to contradict these findings. Firstly, we can simulate the $k_2$ model and track the degree of specific nodes added to the network over time. This is shown in Fig. 9 for $m = 1, 2, 3$. The data has been averaged over $10^4$ simulations with the shaded regions indicating the standard deviation. Note, to facilitate the use of a log-log plot, the lower bound of the shaded regions have been capped at $k_i^{(1)}(t) = m$. The dashed lines indicate the expected scaling for linear preferential attachment which is known to exhibit power-function scaling with an exponent $\delta = 1/2$. A large number of simulations are required since there is significant variation between the evolution of individual nodes across simulations – only with a very large sample size can the average evolution of node $i$ be observed. Note, due to computational constraints, simulations are terminated at $t = 2 \times 10^5$ time steps.

A number of conclusions can be drawn from Fig. 9. Firstly, for a transient period after being added to the network, the average degree evolution of a node appears to scale as $t^{1/2}$ which is the expected scaling for linear preferential attachment. Data is averaged over 100 simulations. Error bars are negligible and have therefore been omitted. For large $t$, the number of edges in the influence network scales approximately with $t^{4/3}$ as indicated by the dashed line, although there are small deviations from this simple scaling.
although we note that the integral in Eq. (22) is dominated to contradict the $\delta$ preferential attachment in the BA model. This appears the use of a log-log plot.

preferential attachment, see Eq. (20). Note, the lower limit of the power function scaling that would be expected from linear deviation across the simulations. The dashed lines indicate the shaded region around the solid lines indicates the standard evolution derived from the $k_2$ model simulations. The

to facilitate this transition takes place increases with the time nodes are added to the network.

This suggests that the true functional form for the degree evolution in the $k_2$ model comprises of two competing terms, the first scaling as $t^{2/3}$ which is suppressed by $t_i$, and a second term which scales as $t^{1/2}$ which is suppressed by $t$. We hypothesize that this implies two scaling regimes: For fixed $t_i$ and $t \to \infty$, the scaling of the degree evolution is dominated by $t^{2/3}$ to ensure that $E(t) \propto t^{1/3}$ as $t \to \infty$. For $t_i \to \infty$ and $t = t_i + \epsilon$ where $\epsilon \ll t_i$, the degree evolution of node $i$ is dominated by the $t^{1/2}$ term. Note, however, that we have not been able to calculate an exact result within the current framework.

It is interesting to consider the origin of the $t^{1/2}$ scaling. Our results are inconclusive, however, if we let $t_i \to \infty$ and set $t = t_i + \epsilon$ with $\epsilon \ll t_i$, a Taylor expansion of Eq. (26) gives

$$\xi_i(t) = b_1 t_i^{4\delta - 2} - b_3 t_i^{2\delta - 1} + O(\epsilon),$$

(29)

with positive constants $b_1$ and $b_3$, revealing that $\delta \geq 1/2$, rather than $\delta \geq 2/3$, is sufficient for ensuring that $\xi_i(t = t_i) \geq 0$ as $t_i \to \infty$.

The mathematical argument presented here does not provide a rigorous proof of the limiting behavior of the $k_2$ model – it is unclear how one would derive such a proof. However, the result does indicate that the inclusion of simple nearest neighbor correlations in the rules of network growth introduce restrictions on the future development of a network. Despite initially appearing to grow as linear preferential attachment, this initial scaling breaks down as the network grows. In the case of real networks this may happen at an early stage in the evolution of a network. However, as demonstrated using the $k_2$ model, the transient time during which the model appears to grow according to linear preferential attachment may be significant - it is not uncommon to analyze real networks with $10^4 - 10^5$ nodes, yet in the case of the $k_2$ model, the network is still in this transient period. This supports the primary argument of this paper that assuming a real network follows simple scaling rules risks misleading conclusions about the future or past evolution of the network.

### IIIc. APS Citation Network

Before analysing the APS citation network, we must restate that our claim is not that the APS citation network follows the dynamics outlined in the $k_2$ model. Rather, our aim is to illustrate that the attachment kernel is not time independent.
Figure 10 shows the degree distribution for the APS citation network in the years 1960 and 2003. The degree distribution follows a log-normal distribution as shown in [30,41]. There are small deviations in the degree distribution between the networks in 1960 and 2003.

Figure 11(a) shows the relative attachment kernel, \( \phi_t(k, 1) \), for the APS citation network up to the years 1960 and 2003 as calculated using the corrected Newman method. The data has not been binned. Recall that, by definition, \( \phi_0(1, 1) = 1 \). If the attachment kernel of the APS citation network is time independent, then we expect \( \phi_{2003}(k, 1) \approx \phi_{1960}(k, 1) \) for \( k \) values where both functions are defined. However, Fig. 11(a) suggests that the time independence condition does not hold - \( \phi_{1960}(k, 1) \) appears to grow faster than \( \phi_{2003}(k, 1) \) up to a cutoff at \( k \approx 200 \). This difference is most clearly seen in the ratio of the relative attachment kernels shown in Fig. 11(b). If the ratio at a given value of \( k \) has value \( \alpha \), this implies that the probability of attaching to a node with degree \( k \) in the year 1960 was \( \alpha \) times larger than the probability of attaching to a node with degree \( k \) in the year 2003, relative to the probability of attaching to a node with degree \( k = 1 \) in either year. Figures 11(c) & (d) show the equivalent as Fig. 11(a) & (b) for the cumulative relative attachment kernel. The consistent deviations in the ratio from one indicate that the APS citation network does not have a time independent relative attachment kernel. This is not in principle surprising since scientists ability to access literature will have changed significantly from 1960 to 2003, notably with the growth of the internet.

IV. DISCUSSION & CONCLUSION

Studying the topology of network degree distributions has become a hallmark of network analysis in the last two decades. The aim of such analysis is simple - to elucidate the processes underlying the formation and growth of a network, and to understand the implications for dynamics taking place on such a network. Amongst the core ideas in network topology is “preferential attachment” [3], the undirected version of Price’s “cumulative advantage” for directed networks [2]. If nodes in a growing network receive connections from new nodes proportionally to their degree, the resulting degree distribution will exhibit power-law scaling.

Although the underlying idea is simple, finding preferential attachment in real networks has proved controversial. Most commonly, authors look for the characteristic power-law scaling in a network degree distribution. Many networks have been identified as showing the necessary scale-free property including web-page links on the internet [6], citation networks [7], the co-occurrence of words in language [8], sexual contact [9] and social networks [10]. However, with a power-law distribution closely resembling related distributions such as the log-normal or stretched exponential, other authors suggest that power-law degree distributions are far rarer in real networks [4].

An alternative approach to analysing a network degree distribution is to explicitly measure the temporal evolution of a network [5,22,26,27]. This approach is far less frequently used since the necessary data is often missing for real networks. When data is available, the probability that new nodes added to a network attach to nodes with a given degree can be estimated - the attachment kernel. If the attachment kernel is linear, this implies the degree distribution will exhibit power-law scaling, and likewise, if we observe power-law scaling, we expect a linear attachment kernel. However, all of this assumes that the degree of a node is the key observable which determines the future evolution of a network. In reality, networks may evolve according to a completely unrelated node observable. In the case of the k\(^2\) model, the observable used is related to the node degree but it is one which breaks the breaks simple scaling assumed by preferential attachment.

If all we are interested in is a network at a fixed point in time, both approaches work well. However, the risk arises if we analyse a network at time \( t_1 \) and want to use the results to make a statement about the network at a later time, \( t_2 \). For instance, consider super-linear preferential attachment. As shown by [Krapivsky and Krioukov 19], in the asymptotic limit super-linear preferential attachment results in a network where all nodes connect to a single dominant hub node. However, in the pre-asymptotic regime the degree distribution shows power-law scaling. So if we observe a real network with power-law scaling, how do we know whether the network in the future will continue to show power-law scaling indicative of linear preferential attachment, or whether we will start to observe a gelation phenomena associated with super-linear preferential attachment? Without more information about the network’s evolution the answer is unclear. Having access to the attachment kernel should solve this problem - in papers analysing the attachment kernel of real networks, it is not uncommon to find statements along the lines of “in the case of super-linear at-
FIG. 11. (a) The relative attachment kernel, $\phi_t(k, 1)$, for the APS citation network as calculated using the corrected Newman method. The calculated attachment kernel is shown for the network up to the years 1960 (blue up triangles) and 2003 (red down triangles). (b) The ratio of the 1960 to 2003 relative attachment kernels. Note, the ratio is only defined at $k$ values where both $\phi_{1960}$ and $\phi_{2003}$ are non-zero. (c) The cumulative sum of the relative attachment kernels shown in (a), $\tilde{\phi}_t(k, 1)$. (d) The ratio of the 1960 to 2003 cumulative relative attachment kernels. The dashed lines in (b) and (d) indicate the ratio expected if the relative attachment kernels are time independent. The deviations from the dashed line indicate that the relative attachment kernel is not time independent for the APS citation network.

tachment, the degree distribution is predicted to converge over time to a state in which a single node dominates all other nodes". For an ideal network where the attachment kernel is always a fixed super-linear function of the node degree, this statement is true. However, it is not clear whether the attachment kernel for a real network is fixed in time. If the attachment kernel is *time dependent*, we are not aware of methods for estimating the attachment kernel which take this into account.

In this paper, we have tried to bring attention to these issues by demonstrating how very simple network growth rules can break both the time-invariance of the network degree distribution and the time-invariance of the attachment kernel. We have done this by introducing the k2 model, a simple variant of the Barabasi-Albert model of network growth where the attractiveness of a node is correlated to the attractiveness of a node’s neighbors. Even though such a network growth rule does not contain an explicit time dependence, the formation of clusters means that a node’s attractiveness is strongly dependent on the local structure of the network.

The k2 model shows that for small networks, the degree distribution appears approximately power-law, and the attachment kernel is approximately linear, both of which are consistent with preferential attachment. Likewise, the degree evolution of most nodes in the network appear to initially grow as a power function scaling as $t^{1/2}$, as expected in the BA model.

However, analysis of the k2 model demonstrates that power function scaling of the form found in linear preferential attachment is only consistent with the k2 model for individual nodes during a transient period after initially entering the network. The oldest nodes appear to demonstrate $t^{2/3}$ scaling which would be consistent with non-linear preferential attachment. This is the only power function scaling which satisfies the requirements of the k2 model for $t \to \infty$, but finite $t_i$. These two regimes are separated by an cross-over region that, so far, remains
unexplained. This anomalous cross-over region grows over
time and is associated with gelation to specific communi-
ties where new nodes immediately acquire a large influ-
ence, but start with a small degree. The interplay of these
different scaling regimes indicates that, had the k2 model
been a real network, the analysis of the model at any given
point in time may have resulted in misleading conclusions
about the future evolution of the network, with different
pieces of evidence pointing towards linear preferential at-
tachment, non-linear preferential attachment, or another
unexplained form of non-linear attachment.

The k2 model is an idealised network growth model,
however, changes in the degree distribution and the at-
tachment kernel can also be seen in real data, such as
the APS citation network. Although the effect is small,
our analysis shows that the degree distribution and the
network attachment kernel are not time independent. Is
this in itself problematic? Not if we take care when draw-
ing conclusions from these observables. However, when
papers conclude that observing super-linear attachment
will result in a gelation phenomena, this approach risks
misleading results. To the best of our knowledge, there
has been no systematic study of how the attachment ker-
nel of growing networks varies over time. This should be
a key priority moving forward.

To conclude, we have introduced a simple network
model and analysed the APS citation network to demon-
strate the risks of assuming time-invariant degree distri-
butions and attachment kernels when making predictions
about the future evolution of networks. Our aim is not
to suggest that the degree distribution and attachment
kernel are not valuable network observables. However, our
analysis suggests far more care needs to go into analysing
how network topology changes with time when attach-
ment rules are not stationary. Future work should focus
on developing improved tools to better understand these
phenomena.

ACKNOWLEDGMENTS

M.F. thanks Renaud Lambiotte, Henrik Jeldtoft Jensen
and Vaiva Vasiliauskaite for their comments. M.F. thanks
the hospitality of the Miyake laboratory at Tokyo Institute
of Technology where much of this work was initiated
and carried out. M.F. gratefully acknowledges a PhD
studentship from the Engineering and Physical Sciences
Research Council through Grant No. EP/N509486/1. KO
was supported by the Japan Society for the Promotion of
Science (KAKENHI), Grant-in-Aid for Scientific Research
(C), Grant Number 18K11500.

M.F., J.L., S.A. and K.O. proposed the project and
defined the k2 model. M.F. and J.L. ran simulations and
analysed the k2 model. K.O., T.S.E. and M.F. devel-
oped the mathematical background of the model. S.A.,
K.O. and Y.M. proposed the second degree as a network
variable of interest. M.F. proposed challenging the as-
sumptions of simple scaling in growing networks using the
k2 model. J.L. analysed the APS citation network. K.O.,
K.Y., Y.M., T.S.E. and K.C. supervised the project. M.F.
wrote the manuscript. All authors commented on and
approved the final manuscript.
Thus, the denominator of Eq. (2) can be rewritten as

\[ \frac{k_{i_m}^{(1)}(t)}{k_{j_{n_i}}^{(1)}(t)} \]

Then, the k2 model with \(m = 1\) and the neighbors \(i_\alpha(\beta)\) connected to node \(i_\alpha\). Here, \(\alpha = 1, \ldots, k_i^{(1)}(t)\), and \(\beta = 1, \ldots, k_\alpha^{(1)}(t) - 1\). Then, the k2 model with \(m = 1\) is applied to nodes \(i_\alpha\) as follows:

\[ \frac{d}{dt} k_{i_\alpha}^{(1)}(t) = \frac{k_{i_m}^{(2)}(t)}{\sum_{j=1}^{N} k_{j_{n_i}}^{(2)}(t)} \quad \text{for } \forall \alpha. \]  

Next, by using Eq. (34), the numerator of the right side of Eq. (34) can be rewritten as

\[ k_{j_{n_i}}^{(2)}(t) = k_{i_\alpha}^{(1)}(t) + \sum_{\beta=1}^{k_{i_\alpha}^{(1)}(t) - 1} k_{i_\alpha(\beta)}^{(1)}(t). \]

Therefore, we obtain the dynamical equations of node \(i_\alpha\) as follows:

\[ \frac{d}{dt} k_{i_\alpha}^{(1)}(t) = \frac{k_{i_m}^{(1)}(t) + \sum_{\beta=1}^{k_{i_\alpha}^{(1)}(t) - 1} k_{i_\alpha(\beta)}^{(1)}(t)}{\sum_{j=1}^{N} \left( k_{j_{n_i}}^{(1)}(t) \right)^2}. \]  

Summing \(k_{i_{1\alpha}}^{(1)}(t)\) from \(\alpha = 1\) to \(\alpha = k_i^{(1)}(t)\) for each side, we obtain

\[ \frac{d}{dt} \left( \sum_{\alpha=1}^{N} k_{i_\alpha}^{(1)}(t) \right) = \frac{\sum_{\alpha=1}^{N} k_{i_\alpha}^{(1)}(t) + \sum_{\beta=1}^{k_{i_\alpha}^{(1)}(t) - 1} k_{i_\alpha(\beta)}^{(1)}(t)}{\sum_{j=1}^{N} \left( k_{j_{n_i}}^{(1)}(t) \right)^2}. \]  

Recall that, \(k_{i_\alpha}^{(1)}(t) = \sum_{\alpha=1}^{N} k_{i_\alpha}^{(1)}(t)\). Thus, Eq. (37) can be modified as follows:

\[ \frac{d}{dt} k_{i_\alpha}^{(2)}(t) = \frac{\sum_{\alpha=1}^{N} k_{i_\alpha}^{(1)}(t) + \sum_{\alpha=1}^{N} \sum_{\beta=1}^{k_{i_\alpha}^{(1)}(t) - 1} k_{i_\alpha(\beta)}^{(1)}(t)}{\sum_{j=1}^{N} \left( k_{j_{n_i}}^{(1)}(t) \right)^2} \]

\[ = \left( k_{i_\alpha}^{(1)}(t) \right)^2 + \sum_{\alpha=1}^{N} \sum_{\beta=1}^{k_{i_\alpha}^{(1)}(t) - 1} k_{i_\alpha(\beta)}^{(1)}(t) \]

\[ \sum_{j=1}^{N} \left( k_{j_{n_i}}^{(1)}(t) \right)^2 \]

This is equivalent to Eq. (13).

Vb. Derivation of Eq. (13)

Vc. Equality of denominator in Eq. (12b)

Our mathematical analysis is reliant on the following equality

\[ \sum_{j=1}^{N} k_{j_2}^{(2)}(t) = \sum_{j=1}^{N} \left( k_{j_2}^{(1)}(t) \right)^2, \]

which for \(m = 1\) is proven in appendix Va. For \(m = 1\), this equality is not exact but should be approximately true as long as the number of non-unique second degree neighbors of node \(i\) is small (i.e. second degree neighbors of node \(i\) which are connected to multiple first degree neighbors of \(i\), or are connected to both a first degree neighbor and node \(i\) itself). We can computationally verify whether this assumption is valid for \(m > 1\). Figure 12 plots the ratio of the two sums, for \(m = 1, 2, 3\), defined as

\[ \frac{S_2(t)}{S_1(t)} = \frac{\sum_{j=1}^{N} k_{j_2}^{(2)}(t)}{\sum_{j=1}^{N} \left( k_{j_2}^{(1)}(t) \right)^2}. \]

against time. The figure has been averaged over 100 simulations of the k2 model. Note, the standard deviation of the ratio is very small. As such, the behavior of the ratio for a single simulation is consistent with the behavior observed when averaging over many simulations. Figure 12
Fig. 12. The ratio of the sum over the second degree of each node in the network to the sum over the first degree squared for each node indicated in the network, see Eq. (40). The ratio equals one for \( m = 1 \), and converges to one for \( m > 1 \).
[40] American Physical Society, “APS Data Sets for Research,”
https://journals.aps.org/datasets, accessed: 2018-11-06.
[41] S. Redner, Physics Today 58, 49 (2005).
[42] H. Zhu, X. Wang, and J.-Y. Zhu, Physical Review E 68, 056121 (2003).
[43] K. B. Hajra and P. Sen, Physica A 346, 44 (2005).
[44] S. Goldberg, H. Anthony, and T. Evans, Scientometrics 105, 1577 (2015), arXiv:1408.2970
[45] M. Golosovsky and S. Solomon, Physical Review E 95, 012324 (2017).
[46] M. Newman, Networks (OUP Oxford, 2018).