A likelihood approach to analysis of network data

Carsten Wiuf*†‡, Markus Brameier*, Oskar Hagberg*, and Michael P. H. Stumpf§

*Bioinformatics Research Center, University of Aarhus, Høegh-Guldbergsgade 10, Building 1090, 8000 Aarhus C, Denmark; †Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, Brendstrupgaardsvej 100, 8200 Aarhus N, Denmark; and §Centre for Bioinformatics, Imperial College London, Wolfson Building, London SW7 2AZ, United Kingdom

Edited by David O. Siegmund, Stanford University, Stanford, CA, and approved March 31, 2006 (received for review January 4, 2006)

Biological, sociological, and technological network data are often analyzed by using simple summary statistics, such as the observed degree distribution, and nonparametric bootstrap procedures to provide an adequate null distribution for testing hypotheses about the network. In this article we present a full-likelihood approach that allows us to estimate parameters for general models of network growth that can be expressed in terms of recursion relations. To handle larger networks we have developed an importance sampling scheme that allows us to approximate the likelihood and draw inference about the network and how it has been generated, estimate the parameters in the model, and perform parametric bootstrap analysis of network data. We illustrate the power of this approach by estimating growth parameters for the *Caenorhabditis elegans* protein interaction network.

biological network | importance sampling | likelihood recursion | network model | random graph

Complex biological, sociological, and technological networks vary in size, form, structure, and the mechanisms by which they grow. They are widely seen as convenient and coherent descriptions for the whole set of interactions in biological, social, or technological systems, and their empirical properties have attracted considerable attention. A range of statistical ensembles [in the sense of an "ensemble" in statistical physics (1)] of networks (or probability spaces over graphs) has been studied, notably Erdös-Renyi and scale-free random graphs (2–4). The former has been the canonical model in random-graph theory but does not capture some important aspects of real networks. These often have a fixed number of nodes (e.g., the number of proteins in an organism is fixed) and much broader degree distributions than the Poisson distribution that characterizes the degree distribution of Erdös-Renyi random graphs; i.e., some nodes have a very high degree (number of interactions), whereas most nodes have degree $k = 1$ and 2. A range of mechanistic models has been suggested where the network grows through the addition of nodes and the asymptotic shape of the degree distribution takes on the form of a power law (2).

Testing hypotheses about a network, its form, and structure and how it has evolved will be difficult, even if a plausible model for network growth can be found. Typically, the analysis of networks has therefore involved either the use of summary statistics, such as the degree distribution or the clustering coefficient, or, in the case of hypothesis tests, rewiring the network while keeping the degree of each node fixed. In the latter case each node has a number of "stumps" equal to its degree and the stumps are connected randomly to create a randomly rewired replicate, e.g., ref. 5. This procedure is well defined and meaningful, but it means that the replicates are uncorrelated (degree–degree correlations depend only on the degree sequence) and any potential coarse structure of the network (such as community structure) is ignored. Thus, although bootstrap methods can, in principle, be more informative than simple summary statistics and most structural analyses rely on them at least to some extent, it is important to keep in mind that the rewired instances of the network will often be systematically and qualitatively different from the true network. The answer to a hypothesis test might depend crucially on the part of the data that is kept fixed and the part that is changed by the bootstrapping procedure. Quite different answers might be obtained, e.g., if the skeleton of the network is fixed rather than just the (observed) degree distribution, although both approaches might appear reasonable in a given situation.

Alternatively, one might turn to likelihood methods. These methods require a probabilistic model reflecting the nature of the data and how the network has evolved. One popular broad class of mathematical models of networks and network evolution includes duplication-attachment (DA) models (4, 6, 7), where a set of parameters specifies the probabilities for including new nodes and edges. The network is considered the result of an evolutionary stochastic process such that the number of nodes has increased from a smaller number through a series of node adding events. New nodes can be (partial) copies of existing nodes and their links or completely new ones.¶ This class of models includes the Barabasi–Albert model (2) and the duplication model of Chung *et. al* (6) as special cases and can interpolate smoothly between them: both are nested inside the same DA model. Estimating their parameters (the probability of a duplication event and the probability of the duplicate node inheriting an edge from the original node, respectively) allows us, for example, to test the extent to which the assumptions of the Barabasi–Albert model are supported by the data.

Mathematical models of networks have been used among other things to explain evolutionary aspects of biological networks, growth and structure of sociological networks, and how certain features of networks seem to appear naturally and globally, such as fat-tailed degree distributions, e.g., ref. 2. However, to our knowledge, mathematical models have been used only indirectly in statistical analysis; for example, by comparing the observed degree distribution to a probability model for the degree distribution (which can be seen as a composite likelihood approach), e.g., ref. 9 and references therein. In principle these models allow for a deeper and fuller statistical analysis of network data, including estimation of the set of parameters that provides the best fit to the model and hypothesis testing, and subsequently interpretation of parameters in relation to the mechanisms underlying the generation of the data (for example, the underlying biological causes and processes). Hypothesis testing is here naturally performed by using the parametric bootstrap: the null distribution is obtained by simulation of networks under the model with the estimated parameters.

Here we present a method that in principle allows us to calculate the likelihood of the full network under a given mathematical model, thereby using the full network data and all of the information embedded in the data about the network

---

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: DA, duplication attachment; IS, importance sampling.

‡To whom correspondence should be addressed. E-mail: wiuf@birc.au.dk.

¶In contrast, graphical models like Bayesian networks and chain graphs (8) consider a fixed graph that determines the probabilistic dependencies in the data; here the graph/network is the data, i.e. the graph is stochastic.

www.pnas.org/cgi/doi/10.1073/pnas.0600061103

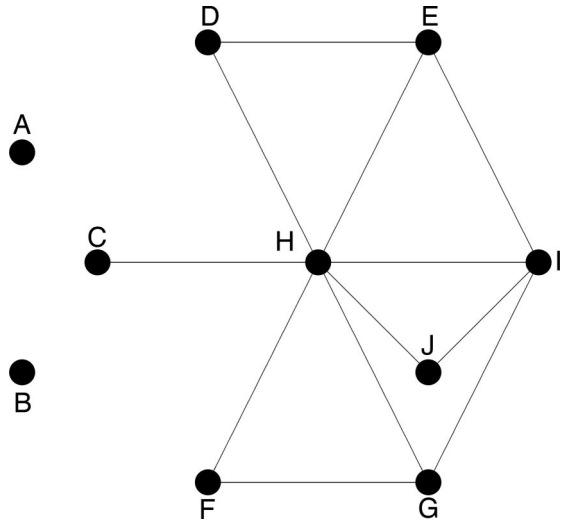


Fig. 1. An example of a graph with 10 nodes where most nodes are removable. A and B can be removed by all other nodes; C can be removed by D–J; D can be removed by E, H, and I; E can be removed by H; F can be removed by G–I; G can be removed by H; H cannot be removed; I can be removed by H; and J can be removed by E and G–I. The graph can be reduced to a single node.

structure. It has been argued (10) that the mechanistic process underlying network evolution is less important than the ability of an ensemble to qualitatively describe the data. It is, for example, well known, that protein interaction networks do not grow according to the simple preferential attachment model that describes simple scale-free models. For example, ≈120 million years before present, a whole genome duplication occurred in the lineage ancestral to *Saccharomyces cerevisiae* and other yeast species belonging to the *sensu stricto* clade and such exceedingly rare events, even though they are of fundamental biological importance, cannot be easily incorporated into a probability model of the sort used in other contexts. But even if the model is (necessarily and indeed desired) oversimplified the likelihood approach allows us to gain additional insights. For DA models, for example, we can infer a maximum-likelihood estimate for the duplication probability. This estimate, in turn, can be interpreted as an "effective" node duplication probability,[‖] which, in the case of biological networks, can be compared with estimates of the extent of historical duplication activity obtained from sequence analysis.

A General DA Model

By a graph we always mean an undirected graph without multiple edges (or links) and self-loops. A randomly growing graph is a Markov chain $\{G_t\}_{t=t_0}^{\infty}$ such that G_t is a graph of size t and the edges and nodes of G_t are subsets of the edges and nodes of G_{t+1}. This implies that one node is added at the time.

We suppose that when G_{t+1} is created, new edges are added between the new node v_{new} and the old nodes that were already present in G_t. One kind of stochastic rule that does this is duplication. An old node v_{old} is picked, and the nodes linked to v_{new} are chosen among the nodes linked to v_{old}. If, in addition, it is also possible to let v_{new} attach to v_{old}, then we speak of DA. For short, we say that v_{new} is copied from v_{old}. See Fig. 1.

Consider a graph G_t and let $\delta(G_t, v)$ denote the graph with v deleted, i.e., v and all links to it are removed. A node v in G_t is said to be removable if G_t can be created by copying a node in $\delta(G_t, v)$. If G_t contains removable nodes, it is said to be reducible, otherwise G_t is irreducible.

A surprising fact, which we prove in *Supporting Text*, which is published as supporting information on the PNAS web site, is that if removable nodes are deleted repeatedly from G_t until the graph is irreducible, we always end up with graphs that are isomorphic to each other irrespectively of the order the nodes were removed. This fact will be important when computing the likelihood.

We say that a randomly growing graph is a DA model if

1. G_{t_0} is irreducible.
2. $\mathsf{P}(G_{t+1}|G_t) > 0$, if, and only if, G_{t+1} can be obtained by copying one node in G_t.

In the next section we consider for simplicity a specific DA model, although the theory applies generally to all DA models (including the Barabasi–Albert model mentioned above).

The Likelihood

Assume the model is described by a possibly vector-valued parameter θ and let G_t be an observation from the model. We are interested in finding the maximum-likelihood estimate of θ. To that end we calculate the likelihood as a function of θ. Let $\mathcal{R}(G_t)$ be the set of removable nodes in G_t, then

$$L(\theta, G_t) = \frac{1}{t} \sum_{v \in \mathcal{R}(G_t)} \omega(\theta, G_t, v) L(\theta, \delta(G_t, v)), \qquad [1]$$

where

$$\omega(\theta, G_t, v) = \mathsf{P}_\theta(G_t | \delta(G_t, v)).$$

The factor $1/t$ is the probability that v is the last added node.

Thus, in principle, it is possible to compute the likelihood recursively. As soon as we arrive at an irreducible graph, it is isomorphic to the true starting graph and has the same likelihood as the true starting graph. If we could end up with nonisomorphic graphs (potentially with different number of nodes), then a nontrivial distribution must have been required for the initial graph.

In the remainder of the section, we consider the following DA model. Let

$$\theta = (\pi, p, q, r).$$

The model is defined by the following two rules.

1. Choose v_{old} in G_t uniformly. Create a link between v_{new} and any node that links to v_{old} with probability p. Link v_{old} to v_{new} with probability q.
2. Choose v_{old} in G_t uniformly. Create a link between v_{old} and v_{new} with probability r.

In every step, we follow rule 1 with probability π, and rule 2 with probability $1 - \pi$. We call this model duplication with random attachment.

To compute the likelihood, we make a list of all nodes in $\mathcal{R}(G_t)$. To every node v in the list there is a nonempty set $\mathcal{C}(G_t, v)$ of nodes w in G_t such that v could have been copied from w. Let $e(w, v) = 1$ if there is an edge between w and v and let $e(w, v) = 0$ otherwise. Further, let $d(v)$ be the degree of node v. Then

$$\omega(\theta, G_t, v) = \frac{1}{t-1} \sum_{w \in \mathcal{C}(G_t, v)} \omega(\theta, G_t, v, w),$$

where

$$\omega(\theta, G_t, v, w) = \pi p^{d(v)-e(w,v)}(1-p)^{d(w)-d(v)} q^{e(w,v)}(1-q)^{1-e(w,v)} + (1-\pi)[I_{\{d(v)=0\}}(1-r) + I_{\{d(v)=1\}} e(w, v) r],$$

[‖] This interpretation is similar in spirit to subsuming details of the demographic history of a population into the effective population size N_e in population genetics, see e.g. ref. 11.

and I is the indicator function. Even though this, in principle, provides the means to compute the likelihood, the method is computationally too intensive even for moderately sized networks; the number of recursive calls that has to be made becomes astronomical in a short time. For a fully connected graph (or a graph with no connections at all) the number of recursive calls is easily seen to be $\lfloor (e-1)t! \rfloor$, where e is Euler's number, $e \approx 2.71$, and $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x. Keeping a list of already calculated likelihood values reduces the number of calls to $t\,2^{t-1} - t + 1$, (see *Supporting Text*), but this number is still extremely large even for small values of t. In terms of computational time it is worth mentioning that each look-up in the list can be done in at most $\log_2(2^t - 1) \approx t$ operations, because there are at most $2^t - 1$ entries (the number of nonempty subsets of t objects) in the list. Fig. 2 provides some examples. For the investigated parameter values the number of calls appears to increase subexponentially, whereas the number for a complete graph is superexponential. Still the numbers become very large; e.g., for $t = 20$ the number of calls are of the order of 10^6 compared with 10^7 for a complete graph.

Importance Sampling (IS)

IS is an efficient simulation (variance reduction) technique that in many cases provides ways to approach quantities for which exact or numerical results are otherwise difficult to obtain, e.g., refs. 12 and 13 for a comprehensive treatment. The IS scheme to be implemented here for computing the likelihood is inspired by the recursion relation (1) that allows the likelihood to be written as an expectation over a Markov chain (see below). Our approach makes direct use of the recursive form of the likelihood function. Similar schemes has been proposed in different contexts; see, for example, refs. 14–16 for proposed schemes for inference on the so-called coalescent (17). We apply IS to random graph models, but the flexibility of IS [or sequential IS (13)] schemes makes them powerful tools for the statistical analysis of (biological) network models.

We rewrite Eq. **1** in the form:

$$L(\theta; G_t) = \frac{1}{t} \sum_{v \in \mathcal{R}(G_t)} \frac{\omega(\theta_0, G_t, v)}{\omega(\theta_0, G_t)} S(\theta_0, \theta, G_t, v) L(\theta; \delta(G_t, v)), \qquad [2]$$

where for $t > t_0$

$$\omega(\theta_0, G_t) = \sum_{v \in \mathcal{R}(G_t)} \omega(\theta_0, G_t, v),$$

and

$$S(\theta_0, \theta, G_t, v) = \frac{1}{t} \frac{\omega(\theta_0, G_t)}{\omega(\theta_0, G_t, v)} \omega(v, G_t, v).$$

Using lemma 1 in ref. 14 (the lemma relates generally to Markov chains with a stopping rule), the likelihood in Eq. **2** can be written as an expectation

$$L(\theta; G_t) = \mathsf{E}_{\theta_0}\left[\prod_{s=t_0}^{t} S(\theta_0, \theta, G_s, v) \right], \qquad [3]$$

where

$$S(\theta_0, \theta, G_{t_0}, v) = L(\theta; G_{t_0}) \equiv 1.$$

The usability of Eq. **3** rests on the fact that all irreducible graphs derived from G_t are isomorphic and thus have the same likelihood. In Eq. **3**, the expectation is with respect to the probabilities

$$\frac{\omega(\theta_0, G_s, v)}{\omega(\theta_0, G_s)},$$

$s = t_0 + 1, \ldots, t$, that define a Markov chain on graphs. This Markov chain is not the same as the one defined by the DA model. However, it motivates the following simulation scheme:

1. Let $G_t^{(i)} = G_t$.
2. For $s = t - 1, \ldots, t_0 + 1$, choose v_i with probability proportional to $\omega(\theta_0, G_s^{(i)}, v_i)$ and let $G_{s-1}^{(i)} = \delta(G_s^{(i)}, v_i)$.
3. Let $l_{\theta_0}^{(i)}(\theta) = \Pi_{s=t_0}^{t} S(\theta_0, \theta, G_s^{(i)}, v_i)$.
4. Repeat steps 1–3 N times and approximate $L(\theta, G_t)$ by

$$\hat{L}(\theta, G_t) = \frac{1}{N} \sum_{i=1}^{N} l_{\theta_0}^{(i)}(\theta). \qquad [4]$$

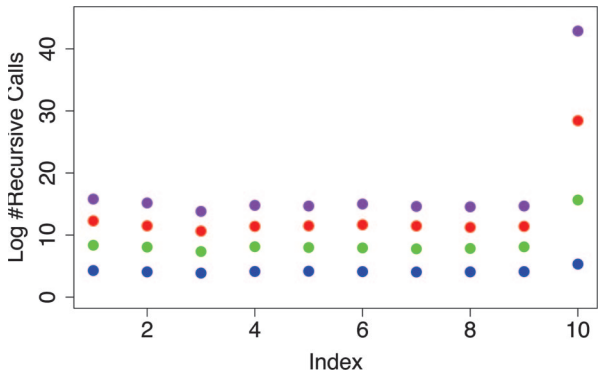
Each of the N draws is called a path. The value θ_0 is the so-called driving value (12), which can be chosen either arbitrarily or in some other conditioned way, e.g., by using summary statistics. Fig. 3 provides an example for $t = 10$. There are at least two things that are worth pointing out. First, the overall form of the simulated likelihood curve is similar to the true likelihood curve even for small N and thus the relative likelihood $L(\theta; G_t)/L(\theta_1; G_t)$ (for fixed θ_1) might be estimated accurately even for small N (depending on t); this observation is also made by other authors (14, 18). Second, the driving value influences the accuracy of the simulated likelihood curve. A driving value close to the true value is likely to provide faster convergence to the true likelihood curve than a driving value far from the true value. However, for all θ_0 and θ the convergence is of order $\sqrt{N}$, because Eq. **4** is an unbiased estimator of the likelihood. Hence, the rate of convergence depends solely on the standard deviation of the terms $l_{\theta_0}^{(i)}(\theta)$ in Eq. **4**.

We calculated the average run time for one path for different network sizes. Graphs with different numbers of nodes were generated with parameter $\theta_1 = (1,0.66,0.33,0)$ or $\theta_2 = (1,0.33,0.33,0)$ and a path was drawn 15 times for each parameter value by using driving value, $\theta_0 = \theta_i, i = 1, 2$. The observed run time was approximately polynomial with an estimated degree of 2.34 and 2.84, respectively; see *Supporting Text* for a description of the algorithm and Fig. 6, which is published as supporting information on the PNAS web site, for a plot of run times. For θ_1 the average number of links per node increases with network size, whereas for θ_2 it stabilizes and is lower than the average for θ_1. Apparently, in these cases it has the opposite effect on the run times.

Application

We applied the IS method to protein interaction data from *Caenorhabditis elegans* (19). The largest connected component was selected (2,368 nodes; as described in ref. 20), and the data were analyzed by fixing three of four parameters for the sake of demonstration: $\pi = 1$, $q = 0.33$, and $r = 0$; $\theta_0 = (1,0.66,0.33,0)$ was used as driving value; and p was varied between 0 and 1. Fig. 4 shows simulated likelihood curves. The maximum-likelihood estimate of p is $\approx$0.28. In other words, when assuming the model and the other parameters are correct only 28% of all links survive when a node is copied.

Two things transpire from the likelihood curves: The first is that the variance of the contribution from one path $l_{\theta_0}^{(i)}(\theta)$ (see Eq. **4**) is small compared with the log-likelihood $L(\theta, G_t)$ itself. We take this observation as evidence that even with a small number of paths the importance sampled likelihood is a good approximation to the true likelihood. Possibly it is a large sample (or network) size effect. The second thing is that the confidence intervals (CI) of the maximum-likelihood estimator of p appear to be wide. The CIs are likely to be even wider if all four parameters are estimated. After removing all removable nodes



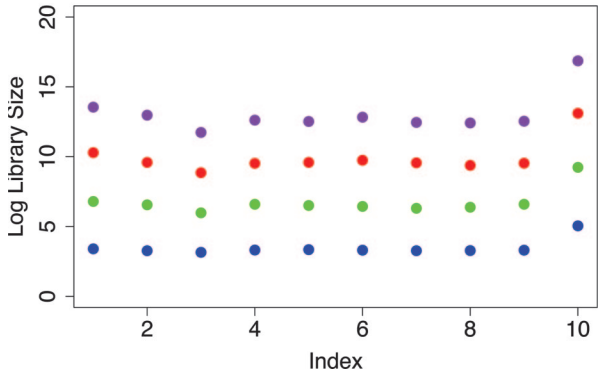


Fig. 2. Complexity of likelihood recursion. (*Upper*) The logarithm of the number of recursive calls made when calculating the likelihood of a graph with Eq. **1** and a library for storing already computed likelihoods of subgraphs for different parameter values. Index 1: $\theta = (\pi, p, q, r) = (0,0,0,0.25)$; index 2: (0,0,0,0.5); index 3: (0,0,0,0.75); index 4: (0.5,0.25,0.5,0.5); index 5: (0.5,0.5,0.5,0.5); index 6: (0.5,0.75,0.5,0.5); index 7: (1,0.25,0.5,0); index 8: (1,0.5,0.5,0); and index 9: (1,0.75,0.5,0). Index 10 shows the log of the number of calls for a complete graph or a graph with no links at all. This number is $\log_e(t2^{t-1} - t + 1)$, where t is number of nodes. $t = 5$ nodes (blue), 10 nodes (green), 15 nodes (red), and 20 nodes (purple). The average over 30 random graphs was computed for each value of t and θ, and all graphs were generated from a single node. (*Lower*) Shown is the log of the library size when calculating the same likelihood as in *Upper*. The number for index 10 is $\log_e(2^t - 1)$.

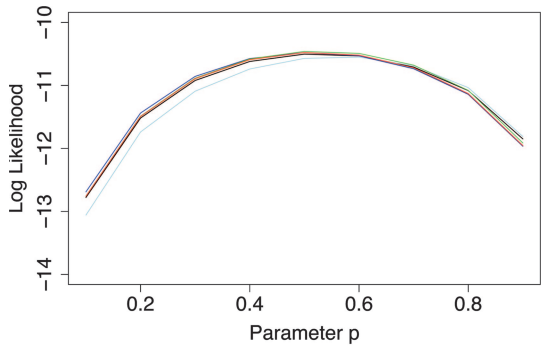
there are still 735 nodes left, which is a considerable amount. Both degree sequences could well be explained by a power law with degree ≈2 (see Fig. 5).

Discussion

While the inference of e.g., coexpression networks and the statistical inference and analysis of graphical models and Bayesian networks have received great attention from the statistics community (see e.g., refs. 21–23), the analysis of experimentally derived molecular networks has largely relied on ad hoc descriptions and summary statistics. This situation contrasts with the social sciences, where networks are much smaller and inferential procedures have been used with some considerable success (24). Thus there is a rich amount of literature on the statistical inference of networks and, perhaps to a lesser extent, also on the analysis of smaller (social) network structures.

In this article we have attempted to calculate the full likelihood of a network, G_t, assuming a mathematical model of network growth that was inspired by basic evolutionary processes.** For the class of DA models we consider here, we can show that it is possible to calculate the likelihood without specifying a distribution on the initial graph, simply because it turns out that the any two paths consistent with G_t initiate with isomorphic graphs. One can also choose to think of the likelihood of G_t as conditional on the initial graph isomorphism.

**Duplication is a straightforward biological process. For the attachment process there is, apart from horizontal gene transfer in bacteria, no easy biological equivalent; rather it is convenient to subsume all nonduplication processes into this mechanism.



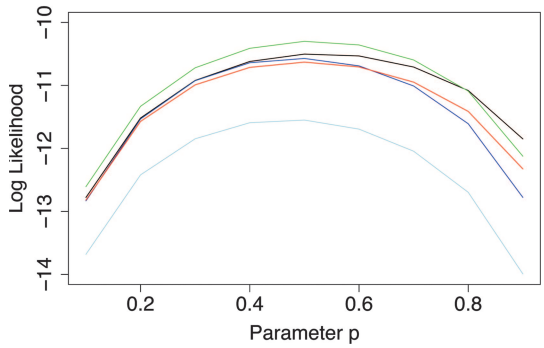


Fig. 3. Likelihood curve for the graph in Fig. 1 for variable p, but with all other parameters fixed to the true values. The graph is generated with parameter $\theta = (1,0.66,0.33,0)$. (*Upper*) The likelihood curve is calculated by using IS with driving value $\theta_0 = \theta$, i.e., $p_0 = 0.66$. The average over several paths is computed. $n = 10$ paths (light blue), 100 paths (blue), 1,000 paths (green), and 10,000 paths (red). For 100,000 the curve is almost impossible to distinguish from the true likelihood curve that is shown in black. (*Lower*) Here the driving value is $\theta_0 = (1,0.33,0.33,0)$, i.e., $p_0 = 0.33$. The average over several paths is computed. $n = 10$ paths (light blue), 100 paths (blue), 1,000 (green), and 10,000 (red). For 100,000 the curve is almost impossible to distinguish from the true likelihood curve that is shown in black.

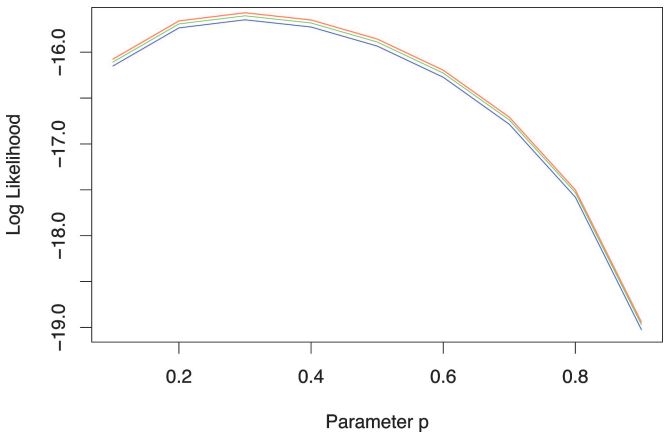


Fig. 4. Shown is the likelihood curve (in units of thousands) of the *C. elegans* data set of 2,368 nodes. After removing all removable nodes the graph comprises 735 nodes. The likelihood curve was generated by using 1,000 paths and driving value $\theta = (1,0.66,0.33,0)$. All parameters but p were fixed: $\pi = 1$, $q = 0.33$, and $r = 0$. Average of 10 paths (blue), 100 paths (green), and 1,000 paths (red). Each path took ≈400 central processing unit sec on our machine.

STATISTICS

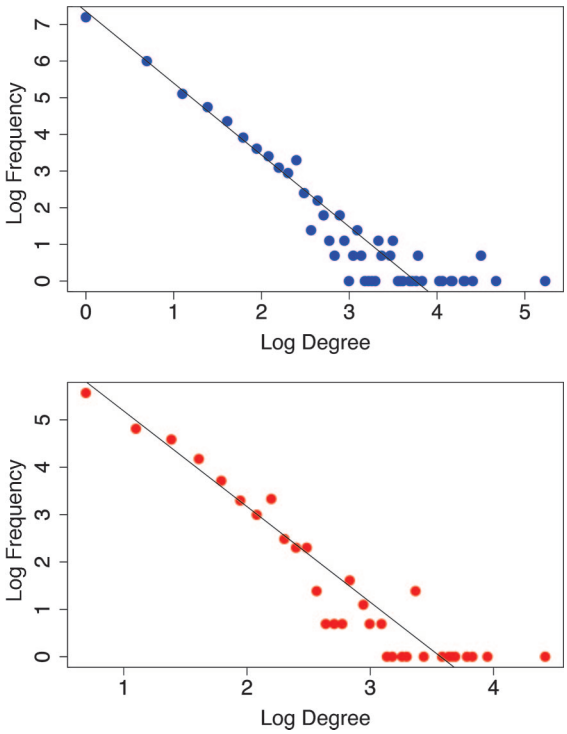


Fig. 5. Shown is the degree distribution of the full *C. elegans* data set (*Upper*) and the reduced data set (*Lower*). Both data sets look like the degree distribution can be described by a power law with coefficient ≈2. In the full data set there are no nodes of degree 0 because the network is connected, and in the reduced data set there are no nodes of degrees 0 and 1, because they can always be removed.

We further showed that the likelihood, in principle, could be calculated by using a recursion, but also that this approach is computationally too demanding to be practical for even moderate networks. As an alternative, we suggested adopting an IS approach that samples paths consistent with $\mathcal{G}_t$ and that the likelihood can be computed from such paths. In our implementation this approach also runs into time constraints but only for graphs exceeding at least 2,500 nodes.

Our work leaves room for improvements and also raises some questions. The application raises the question of whether the DA models are adequate models for describing the *C. elegans* network. In previous work (9) we have shown that power laws describe the degree distribution of the *C. elegans* data statistically better than other types of distributions (derived from normal, exponential, and other standard distributions) (see also Fig. 5). However, the initial (irreducible) graph comprises almost one-third of the nodes in the entire network (735/2,368 = 31%), implying that there are loops and cycles in the network that are not consistent with how DA models build up graphs (see also *Lemma 6* in *Supporting Text*).

This observation leads us to our second point: development of more realistic models. We have stuck with the class of DA models because they are widely used and discussed in the literature. Generalization to directed DA models should be straightforward, but also models that evolve by other mechanisms than duplication and attachment should be possible to handle in a similar way to that described here. Very general models allowing for insertion and deletion of edges at any time are straightforwardly handled by the theory because the graph eventually is reduced to a single node. However, this simplicity is at the cost of computational complexity because the number of paths from G_t to G_{t_0} (containing a single node) is now even larger and may quickly become unmanageable. More importantly, such a model is biologically implausible. Nature is not likely to remove functions and interactions without having reasonable substitutes for them. Models that allow for moderate deletions of nodes and/or edges are biologically much more realistic. For example, one could allow a link between two nodes, v and w, to be removed if a copy v' of v exists that also has a link to w. Such features are biologically tractable but potentially require more bookkeeping for calculating the likelihood.

Finally, it would be natural to engage in Markov chain Monte Carlo and Gibbs sampling methods to improve the speed and perhaps accuracy of the computations, but also to try other IS schemes. As discussed in ref. 18, a recursion like ref. 1 opens more possibilities than the one presented here. These, including the one presented here, fall under the general principle of sequential IS in which one builds up the sampling distribution sequentially, see e.g., ref. 13 for general discussion and examples. Sequential IS might be particularly useful for random graphs because one can envisage the graph as being built up step by step. However, the shapes of the simulated likelihood curves in Fig. 4 also raise the question of whether, in large networks, individual paths provide a good estimate of the likelihood.

We thank Sylvia Richardson for helpful discussions. C.W. and M.P.H.S. are supported by the Carlsberg Foundation and the Royal Society. C.W. is supported by the Danish Cancer Society. M.P.H.S. is supported by a Wellcome Trust Fellowship and a European Molecular Biology Organization Young Investigator Award.

1. Thompson, C. J. (1979) *Mathematical Statistical Mechanics* (Princeton Univ. Press, Princeton).
2. Barabási, A. L. & Albert, R. (1999) *Science* **286,** 509–512.
3. Bollobás, B. (2001) *Random Graphs* (Cambridge Univ. Press, Cambridge, U.K.), 2nd Ed.
4. Dorogovtsev, S. N. & Mendes, J. F. F. (2003) *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Univ. Press, Oxford).
5. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskij, D. & Alon, U. (1997) *Science* **298,** 824–827.
6. Chung, F., Linyuan, L., Dewey, G. & Galas, D. (2003) *J. Comp. Biol.* **10,** 677–687.
7. Kumar, S. R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A. & Upfal, E. (2000) in *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science*, ed. Blum, H. (IEEE Computer Soc., Washington, DC), pp. 57–65.
8. Lauritzen, S. L. (1996) *Graphical Models* (Clarendon, Oxford).
9. Stumpf, M. P. H., Ingram, P. J., Nouvel, I. & Wiuf, C. (2005) in *Transactions in Computational Systems Biology*, ed. Priami, C. (Springer, New York), pp. 65–77.
10. Burda, Z., Diaz-Correia, J. & Krzywicki, A. (2001) *Phys. Rev. E* **64,** 046118.
11. Ewens, W. J. (2005) *Mathematical Population Genetics* (Springer, New York), 2nd Ed.
12. Ripley, B. (1987) *Stochastic Simulation* (Wiley, Sussex, U.K.).
13. Liu, J. S. (2001) *Monte Carlo Strategies in Scientific Computing* (Springer, New York).
14. Griffiths, R. C. & Tavaré, S. (1994) *Theor. Popul. Biol.* **46,** 131–159.
15. Griffiths, R. C. & Tavaré, S. (1994) *Stat. Sci.* **9,** 307–319.
16. Griffiths, R. C. & Marjoram, P. (1996) *J. Comp. Biol.* **3,** 479–502.
17. Kingman, J. F. C. (1982) *Stoch. Proc. Appl.* **13,** 235–248.
18. Stephens, M. & Donnelly, P. (2000) *J. R. Stat. Soc. B* **62,** 605–655.
19. Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.-O., Han, J.-D. J., Chesneau, A., Hao, T., *et al.* (2004) *Science* **303,** 540–543.
20. Agrafioti, I., Swire, J., Abbott, J., Huntley, D., Butcher, S. & Stumpf, M. P. H. (2005) *BMC Evol. Biol.* **5,** 23.
21. Lauritzen, S. L. & Richardson, T. S. (2002) *J. R. Stat. Soc. B* **64,** 321–348.
22. Schafer, J. & Strimmer, K. (2005) *Bioinformatics* **21,** 754–764.
23. Pournara, I. & Wernisch, L. (2004) *Bioinformatics* **20,** 2934–2942.
24. Snijders, T. A. B. (2002) *J. Soc. Struct.* **3,** 2.