A Dynamic Random Graph Model for Diameter Constrained Topologies in Networked Systems

Luigi Alfredo Grieco, Senior Member, IEEE, Mahdi Ben Alaya, Thierry Monteil, and Khalil Drira

Abstract

Random graphs have been widely investigated in literature because of their relevance to many scientific domains. In this brief, the attention is focused on diameter constrained random graphs, useful to analyze unstructured overlays for delay bounded network applications and systems. To this end, a general process of arrivals is considered that describes the sequence of vertex couples (i.e., node couples) among which a path composed of no more than $D$ edges (i.e., links) has to be established. Accordingly, a topology formation mechanism $M$ is formulated, expressing the rules that drive the addition of new edges, obeying to the constraint on the maximum diameter $D$. Third, using graph theoretic arguments, an original discrete time model is proposed that describes the evolution of the average network degree (i.e., the average number of edges per node) subject to $M$ and $D$. Fourth, the model is successfully validated using computer simulations in a wide range of scenarios (with up to $2^{16}$ nodes). Finally, concrete examples are provided to illustrate useful applications of the proposed approach, also in the presence of link failures.

Index Terms

Graph theory, Topology, Networks.

I. INTRODUCTION

Graph-based models are fundamental tools to assess, predict, and control the performance of complex networked systems, made of interacting dynamic units. In these systems, vertex are usually associated to the dynamic units whereas edges represent interactions. The application domains of graph-based models include coupled biological/chemical systems, social networks, software applications, and communication protocols [1]–[3].

In many real systems, unfortunately, the properties of their interacting units cannot be deterministically known in advance. In these cases, random graphs [4] can be fruitfully used to infer the characteristics of the topology, based on the probabilistic behavior of vertex and edges (see [5] for a comprehensive overview on the subject).

Till now, with reference to communication issues, random graphs have been mainly adopted to describe unstructured overlays [6]–[8], web properties [9], and Internet topology [10].

In this brief, we focus on diameter constrained overlays, i.e., virtual network topologies having a diameter no larger than a predefined threshold $D$. This kind of overlay is very useful to support delay sensitive applications, such as in Peer-to-Peer (P2P) TV [11] and emerging Machine-to-Machine (M2M) systems [12]. In fact, the higher the diameter $D$ the higher the end-to-end communication delay [13]. The problem of building diameter constrained graphs has been thoroughly afforded in [14] with reference to structured overlays, built upon distributed hash tables (DHT). Unfortunately, to the best of the authors’ knowledge, no theoretical contribution has been formulated yet, able to describe with closed form expressions the dynamics of an unstructured evolving overlay, subject to a constraint of the maximum diameter $D$. This kind of model could be very useful to enable closed loop autonomic management strategies as well as to characterize, in a tractable form, both transient and steady state properties of network topologies in M2M scenarios [12], [15] and beyond.

Starting from this premise, a theoretical model based on random graph is formulated herein, which considers a discrete time process of arrivals to describe the sequence of vertex couples among which a path composed of no more than $D$ edges has to be established.

Accordingly, a general topology formation mechanism $M$ is formulated, expressing the rules that drive the addition of new edges, obeying to the constraint on the maximum diameter $D$. Then, exploiting the properties of the binary adjacency matrix $A$ in graph theory [16], an original and tractable discrete time model is proposed that describes the evolution of the average network degree subject to $M$ and $D$.

The model is successfully validated using computer simulations in a wide range of scenarios (with up to $2^{16}$ nodes). Finally, concrete examples are provided to illustrate useful applications of the proposed approach. They include: (i) the derivation of an approximated upper bound $\sqrt{2\cdot N \cdot \ln N}$ on graph average degree (i.e., the average number of edges per vertex); (ii) the comparison with respect to delay optimal de Bruijn graphs [14]; (iii) the analysis of the graph robustness; (iv) the derivation of system dynamics also in presence of edge failures.

The rest of the brief is organized as follows: the main theoretical achievement is presented in Sec. II and validated in Sec. III. Useful examples of its applications are described in Sec. IV. The last Sec. V closes the brief and draws future research.

This work was supported by the INSA of Toulouse (FR) and by the PON project RES NOVAE funded by the Italian MIUR and by the European Union (European Social Fund).

L. A. Grieco is with the Department of Electrical and Information Engineering, Politecnico di Bari, Bari, Italy, e-mail: alfredo.grieco@poliba.it

M. B. Alaya and T. Monteil are with CNRS-LAAS, INSA, and Univ. de Toulouse, Toulouse, France. K. Drira is with CNRS-LAAS and Univ. de Toulouse, Toulouse, France. e-mail: {ben.alaya, monteil, khalil}@laas.fr
II. Model

A. Target Scenario and Notation

The target scenario considered in this brief consists of a graph of \( N \) vertice, \( n_q \) being the \( q \)-th vertice \((q \in [1, N])\). Furthermore, an ordered sequence of equi-probable \(^1\) couples of vertice is considered, among which a path composed of no more than \( D \) edges has to be established. The \( t \)-th couiple in the ordered sequence couples of vertice is described by \((n_{i_t}, n_{j_t})\). For sake of simplicity, the variable \( t \) will be referred to as \textit{time} from now on. Knowing the \( t \)-th couple, a new edge is established in the graph if and only if the two vertice \((n_{i_t}, n_{j_t})\) are not reciprocally reachable in no more than \( D \) edges. To this end, \( P_{t-1} \) is defined as the probability that a couple of vertice at time \( t \) will not be reciprocally reachable in no more than \( D \) edges. Since we are assuming homogeneous conditions, \( P_{t-1} \) it is the same for all the possible couples \((n_{i_t}, n_{j_t})\).

In other terms, \( P_t \) expresses the expected number of links that will be added at time \( t \). Accordingly, our model is grounded on the following equation:

\[
l_{t+1} = l_t + P_t
\]

which, considering that the average degree \([4]\) (i.e., the number of edges per vertice) is \( k_t = \frac{2l_t}{N} \), can be also expressed as:

\[
k_{t+1} = k_t + \frac{2}{N} P_t
\]

The presence of an edge between any couple of vertice at time \( t \) will be expressed (as usual in graph theory) using the binary symmetric adjacency matrix \( A_{N \times N}^t \), so that \( A_t(i,j) = A_t(j,i) = 1 \) if and only if an edge between \( n_i \) and \( n_j \) exists at time \( t \) (otherwise \( A_t(i,j) = A_t(j,i) = 0 \)).

\[
\begin{array}{|c|l|}
\hline
\text{Symbol} & \text{Meaning} \\
\hline
N & \text{Number of vertice} \\
k_t & \text{Average degree at time } t \\
n_q & \text{\( q \)-th vertice} \\
D & \text{Maximum diameter} \\
(n_{i_t}, n_{j_t}) & \text{\( t \)-th couiple of vertice wishing to establish a path} \\
l_t & \text{Number of edges at time } t \\
w_t & \text{Number of path with less than } D + 1 \text{ edges between a couple of vertice at time } t \\
P_{t-1} & \text{Probability that no path exists shorter than } D + 1 \text{ edges between the vertice } (n_{i_t}, n_{j_t}) \\
A_{N \times N}^t & \text{Symmetric binary adjacency matrix at time } t \\
Pr\{x\} & \text{Probability of event } x \\
\hat{x} & \text{Upper bound on } x \\
\hline
\end{array}
\]

To provide an illustrative example of the networked system we are modeling as a random graph, Fig. 1 shows the evolution of a graph made of \( N = 5 \) vertice and constrained by \( D = 2 \) max path length. In this example, the ordered sequence of vertice \((n_{i_t}, n_{j_t})\) that ask for a path is: \( (n_2, n_4), (n_1, n_4), (n_1, n_2), (n_4, n_5), (n_2, n_3) \), and \( (n_1, n_3) \). Accordingly, for each of them, a new edge is added if and only if a path shorter than 3 edges is not already available among corresponding vertice. In the sequel of the contribution, we will derive a law that rule the evolution of this kind of graphs in a general case.

B. Main result

**Proposition 1.** For a sufficiently large \( N \), the following expression describes the dynamics of the average graph degree:

\[
k_{t+1} \approx k_t + \frac{2}{N} \cdot \exp\left(-\frac{1}{N} \cdot \frac{k_t^{D+1} - k_t}{k_t - 1}\right)
\]

**Proof.** The model considered here is based on Eq. (1), or equivalently on finding an accurate approximation for the probability \( P_t \). The latter expresses the probability to find a path at time \( t \) (no longer than \( D \) edges) between a generic couple of vertice \((n_{i_t+1}, n_{j_t+1})\), knowing that the number of already existing edges is \( l_t \).

\(^1\)It is worth to note that equi-probable arrivals (i.e., homogeneous conditions) are usually assumed in the current literature dealing with diameter constrained graphs \([14]\). In fact, if the vertice of the graph represent the gateway through which the service of a large number of nodes are made available (as in M2M systems \([15]\)), it is not unlikely that the overlay that inter-connect such gateways actually reflects this assumption.
Fig. 1. Evolution of a random graph ($N = 5, D = 2$): a) initial state; b) the first edge is added at $t = 1$ to connect vertex $n_2$ and $n_4$; c) the second edge is added at $t = 2$ to connect vertex $n_1$ and $n_3$, but no edge is added at $t = 3$ because the vertex $n_1$ and $n_2$ (asking for a connection) are already connected by a path of 2 hops $\leq D$; d) the third edge is added at $t = 4$ to connect vertex $n_4$ and $n_5$; e) the fourth edge is added at $t = 5$ to connect vertex $n_2$ and $n_3$; e) the fifth edge is added at $t = 6$ to connect vertex $n_1$ and $n_3$, for which the only existing path was longer than $D$ hops.

To fulfill this objective, we first leverage a well known property of the matrix $A_t$: $A_{ct}(i,j) = 0, c \in \mathbb{N}^+$, if and only if no path composed of $c$ edges exist between $n_i$ and $n_j$ at time $t$ [16].

In this way, without lack of generality, $P_t$ can be expressed as follows:

$$P_t = \prod_{c=1}^{D} Pr\{A_{ct}(i_{t+1}, j_{t+1}) = 0\}$$  \hspace{1cm} (4)

Now, given that $2 \cdot l_t$ elements are equal to one in $A_t$, it yields $Pr\{A_t(i,j) = 1\} = \frac{2 \cdot l_t}{N^2}$. Also, since any element of $A_{ct}$ is no other than the sum of $N^{c-1}$ elements, each one being a product of $c$ coefficients belonging to $A_t$, we can approximately write:

$$Pr\{A_{ct}(i_{t+1}, j_{t+1}) = 0\} \approx \left[1 - \left(\frac{2 \cdot l_t}{N^2}\right)\right]^{N^{c-1}}$$  \hspace{1cm} (5)

Now, for sake of mathematical tractability, we analyze Eq. (6) for a sufficiently large large $N$. To this end, we rewrite Eq. (6) as:

$$Pr\{A_{ct}(i_{t+1}, j_{t+1}) = 0\} \approx \left[1 - \left(\frac{2 \cdot l_t}{N^2}\right)\right]^{N^{c-1}} \left(\frac{2 \cdot l_t}{N^2}\right)^c$$  \hspace{1cm} (6)

Recalling that $(1 + \frac{1}{x})^x \to e$, when $x \to \infty$, we consider that, for a sufficiently large $N$, $\left[1 - \left(\frac{2 \cdot l_t}{N^2}\right)\right]^{\frac{2 \cdot l_t}{N^2}} \approx e$, thus obtaining the following new approximation for Eq. (6) as:

$$Pr\{A_{ct}(i_{t+1}, j_{t+1}) = 0\} \approx exp\left(-N^{c-1} \cdot \left(\frac{2 \cdot l_t}{N^2}\right)^c\right)$$  \hspace{1cm} (7)

Accordingly, by substituting (7) in (4), it is obtained:

$$P_t \approx exp\left(-\sum_{c=1}^{D} N^{c-1} \cdot \left(\frac{2 \cdot l_t}{N^2}\right)^c\right)$$  \hspace{1cm} (8)
which is equivalent to $P_t \approx \exp \left( - \sum_{c=1}^{D} \frac{\binom{N}{c-1}}{N^c} \cdot \left( \frac{2L}{N} \right)^c \right)$. This latter expression is no other than a geometric series with parameter $\left( \frac{2L}{N} \right)$, multiplied by $-1/N$, truncated at the first $D$ terms, and without the first addend 1, which, after a little algebra, can be expressed as:

$$P_t \approx \exp \left( - \frac{1}{N} \cdot \left( \frac{2L}{N} \right)^{D+1} - \frac{2L}{N} \right)$$  \hspace{1cm} (9)

From [4], the average degree can be expressed as $k_t = \frac{2L}{N}$, so that we obtain the proof by substituting (9) in (2).

### III. Simulations

To validate the model (3), we have considered a complex scenario composed of $N$ vertices (with $N$ up to $2^{16}$) and $D$ ranging from 3 to 10. Using an ad hoc simulator we developed in Matlab, the relative error between the real evolution of the degree $k(t)$ and the one estimated using (3) for all $t$ is evaluated. In any case, we found that the average relative error is below 10% if we consider the entire evolution of $k(t)$. Also, the relative error at steady state (once the graph is completely formed), for $D \leq 5$, is below 10%, whatever $N$. Finally, we notice a slight increase in the relative error as $D$ increases: (in any case) it remains smaller than 25% and it falls below 20% for $N > 2^{12}$ (see Fig. 2). These results are expected because the model is based on two main approximations, used to derive (6) and (7), respectively. The accuracy of the first (resp. second) approximation decreases by increasing (resp. decreasing) $D$ (resp. $N$). As a consequence, it is quite straightforward to observe that estimation errors increase by increasing $D$ whereas they decrease by increasing $N$.

![Fig. 2](image-url)  \hspace{1cm} (a)

![Fig. 2](image-url)  \hspace{1cm} (b)

Fig. 2. Average (a) and Steady State (b) absolute relative errors of the model (3).

To provide an illustrative example, Fig. 3 plots the dynamic evolution of the number of edges versus the time $t$ (similar results have been obtained for different values of $N$ and $D$ so that the average relative error is below 10% in all cases).

We remark that knowing $k$ in advance allows to size the processing and communication capabilities of the physical nodes of the overlay. In fact, the higher $k$ is the higher the load due to packet processing and relaying. So that, an error by $x\%$ on
the estimated number of links per node means that physical nodes should be over provisioned $x\%$ more with respect to the outcomes of the model.

Fig. 3. Evolution of the number of edges over the time $t$, ($N = 1000$, $D = 5$).

It is worth to note that the number of edges linearly increases with $t$ till a saturation point is reached. From that moment on, $l$ exhibits a very slow rise. This can be explained by plotting also the values of the probability $P_t$. Fig. 4 shows that $P_t$ is almost equal to one for some time during the network formation, meaning that, since the number of edges is low, it is highly likely to add a new edge as soon as a new couple of nodes needs to establish a path. At the same time, the values of $P_t$ abruptly decreases after a certain $t$, meaning that the topology reached the steady state.

Fig. 4. Evolution of $P_t$, ($N = 1000$, $D = 5$).

To provide a further insight, Fig. 5 pictures the evolution of a random graph ($N = 1000$, $D = 5$): (a) at the beginning of the simulation, when no edge is present; (b) during the transient, when a few edges have been created and a new edge is added; and (c) at steady state, when all required paths (no longer than $D$ edges) have been already created.

Fig. 5. Evolution of a random graph ($N = 1000$, $D = 5$): a) initial state; b) during the transient a new edge is added; c) at steady state no edge is added because paths no longer than $D$ hops already exist.
IV. Example Applications

A. Rank bound and comparison

**Theorem 1.** Being $k$ the maximum degree $k$ in system (3), it can be bounded as follows:

$$
\hat{k} \leq \sqrt[2]{2 \cdot N \cdot \ln N}.
$$

*Proof.* The way we are building the overlay is so that an edge between a couple of vertex $(n_{i,}, n_{j,})$ is formed or not based only upon the first time that couple issues a request for a path. If a path shorter than $D + 1$ hops already exists the edge is not established otherwise it is established. From that moment on, the next requests for a path issued by the same couple of vertice will not sort any effect.

Based on this consideration, we extract from the sequence of equi-probable couples of vertice considered in the brief, the sequence of instants in which any couple of vertice appears for the first time. Of course, the length of such a sequence of time instants will be composed of at most $N(N - 1)/2$ elements.

In order to estimate an upper bound on the steady state average degree $\hat{k}$, it is necessary to consider that at time $t$, $1/P_t$ expresses the average time required to establish the next edge in the overlay.

Under this assumption, the expression of $P_t$ to consider is slightly different from that in Eq. (9) because the sequence of vertice couples we are considering to proof this Theorem is chosen in such a way that no one edge path exists at time $t$. Also in this case the properties of the adjacency matrix are exploited. In particular, it holds:

$$
\hat{k} \leq \sqrt[2]{2 \cdot N \cdot \ln N}.
$$

*Proof.* The way we are building the overlay is so that an edge between a couple of vertice appears for the first time. Of course, the number of any element if $A$ is $\frac{2l}{N}$ times larger at most, even if it is based on a much simpler construction mechanism.

B. Robustness

Knowing that the graph construction model adopted herein ensures, at steady state, at least one path shorter than $D + 1$ edges between any couple of nodes, it is worth investigating how many paths (composed of less than $D + 1$ edges) are present between any couple of nodes. This metric is intimately related to the topology robustness: the higher the number of paths the higher the number of alternative solutions to route messages in case of failures.

**Theorem 2.** Defined as $w_t$ the number of paths composed of less than $D + 1$ edges between any couple of nodes at time $t$, it holds:

$$
w_t = \frac{1}{N} \cdot \frac{k_t^{D+1} - k_t}{k_t - 1}.
$$

*Proof.* Also in this case the properties of the adjacency matrix are exploited. In particular, $A_t^c(i,j)$ indicates the number of paths composed of $c$ edges between the vertice $n_i$ and $n_j$. Thus, considering that the average value of any element if $A$ is $\frac{2l}{N}$, it holds:

$$
w_t = \frac{1}{N} \cdot \sum_{s=1}^{D} k_t^s = \frac{1}{N} \cdot \frac{k_t^{D+1} - k_t}{k_t - 1}.
$$

This end the proof.

Based on this Theorem, we can derive an approximate assessment of the level of redundancy, if we consider for $k_t$, the bound derived in Theorem 2, i.e., $k_t = \sqrt[2]{2 \cdot N \cdot \ln N}$. Under this assumption, we can obtain:
\[
\dot{w} = \frac{1}{N} \cdot \frac{\sqrt{2 \cdot N \cdot \ln N}}{\sqrt{2 \cdot N \cdot \ln N - 1}} \cdot (2 \cdot N \cdot \ln N - 1) 
\]  
(15)

This result indicates that, for \( N > e \), \( \dot{w} \geq 2 \cdot \ln N - 1/N \). Thus, for \( N > e^2 \), the overlay investigated in this brief is able to provide at steady state at least three paths among any couple of vertice.

C. Transient duration

From (3), \( k_t \) monotonically increases with \( t \). At the same time \( k_t \) is upper bounded, according to Theorem 1. As a consequence, we can conclude that the system (3) is convergent. Herein, the transient behavior of \( k_t \) is investigated to provide a more complete description of the network overlay dynamics.

**Proposition 2.** If we define the transient duration \( t_0 \) of the system (3) as the time instant for which \( P_{t_0} \leq \epsilon \), for any given \( \epsilon > 0 \), with \( \epsilon \) small enough, it yields:

\[
t_0 \leq \frac{\sqrt{N \cdot \ln(1/\epsilon)}}{\epsilon} 
\]  
(16)

**Proof.** During the transient of the system (3), \( k_t \approx t \) because the number of already existing edges is so low that any new couple of vertice that wish to establish a path will trigger the creation of a new link (see also Fig. 3 for an example). This is equivalent to write that, during the transient, \( P_t \approx 1 \). Now, to define the transient duration \( t_0 \), we impose that \( P_{t_0} \leq \epsilon \) for any given \( \epsilon > 0 \), with \( \epsilon \) smaller enough. But, during the transient, \( k_{t_0} \approx t_0 \), so that from (9) it results \( P_{t_0} \approx \exp\left(-\frac{1}{N} \frac{t_0^{D+1} - t_0}{t_0^{D-1} - 1}\right) \). Accordingly \( P_{t_0} \leq \epsilon \) translates to \( \exp\left(-\frac{1}{N} \frac{t_0^{D+1} - t_0}{t_0^{D-1} - 1}\right) \leq \epsilon \), which can be also written as \( t_0^{D} + t_0^{D-1} \ldots + t_0 \geq N \cdot \ln(1/\epsilon) \). The latter inequality is satisfied for \( t_0 = \sqrt{N \cdot \ln(1/\epsilon)} \), from which (16) can be derived.

D. Link failures

To include also possible link failures and dynamics in the model, it is necessary to modify Eq. (1) as follows:

\[
l_{t+1} = l_t + P_t - \lambda_o \cdot l_t 
\]  
(17)

where \( \lambda_o \) is the probability that an edge is removed during one time step. The resulting equations could be very useful to design topology management algorithms using control theoretic arguments. Its utility in finding the uniqueness of the equilibrium point is shown in the following Theorem.

**Theorem 3.** The system (17) admits one and only one equilibrium point \( l = l_\infty \).

**Proof.** In order to find the equilibrium point of system (17), we impose \( l_{t+1} = l_t = l_\infty \) in (17). Accordingly, the following equality is obtained:

\[
P_\infty = \lambda_o \cdot l_\infty 
\]  
(18)

Eq. (18) admits only one solution because its leftmost member monotonically decreases with \( l \), starting from the value one at \( l = 0 \) whereas the rightmost member monotonically increases, starting from zero at \( l = 0 \). This ends the proof.

V. Conclusion

A novel tractable model for describing the dynamics of a diameter constrained random graph is proposed, validated, and analyzed in this brief. Useful examples of its adoption have been also provided in order to demonstrate its real utility. Its pros are the closed form formulation, which eases the math tractability and paves the way to topology control mechanisms, and bounded estimation errors that help sizing the physical resources of overlay nodes. Its cons are the only partial coverage of use cases with link failures and heterogeneous conditions, which will be faced in future research.

REFERENCES


