Analyses of the Clustering Coefficient and the Pearson Degree Correlation Coefficient of Chung's Duplication Model

Duan-Shin Lee, Cheng-Shang Chang and Hao-Neng Chang

***Abstract*—Recent advances in gene expression profiling and proteomics techniques have spawn considerable interest in duplication models for modelling the evolution and growth of biological networks. In this paper, we consider the duplication model studied by Chung et al. It seems (to the best of our knowledge) that both the clustering coefficient and the Pearson degree correlation coefficient of this model have not been studied *analytically*. For such a model, we study the degree of a randomly selected vertex and derive first-order differential equations for its mean, second moment, and third moment. We also study the degrees of the two vertices that appear at both ends of a randomly selected edge and derive an approximation for the expected product of the degrees of these two vertices. Using these results, we obtain closed-form approximations for the clustering coefficient and the Pearson degree correlation coefficient of the duplication model. For the clustering coefficient, numerical results calculated from our approximation and the corresponding simulation results agree extremely well for the whole evolution process. For the Pearson degree correlation coefficient, there is some discrepancy at early times between the simulation results and the numerical results. However, as time goes on, the discrepancy diminishes. We present an asymptotic approximation by keeping only the dominant terms in the clustering coefficient and the degree correlation coefficient. Numerical study indicates that relative approximation error can decrease slowly with time, when the selection probability of the model is near some special values.**

keywords: duplication model, clustering coefficient, degree correlation

I. Introduction

Recent advances in gene expression profiling and proteomics techniques have fueled a series of studies on the structure and evolution of many biological networks [5], [6], [13]. Structures and properties of such networks have been studied. One property of biological networks that has received a lot of attention is the degree distribution of such networks. It is well known that biological networks typically have a scale-free degree distribution with exponents between 1 and 2 [4], non-biological networks typically have exponents between 2 and 4. Biological networks typically have large clustering coefficients, negative degree correlations, and small network diameters. Studies show that duplication of the information in the genome is a fundamental force to drive biological evolution [11], [5], [13]. This prompts the study of an extremely popular model called the duplication-divergence model [4], [3], [8], [15], [1], [12].

Extensive studies on duplication models are available in the literature. However, the duplication models studied by different research teams may not be identical. Chung and her colleagues studied a simple and elegant growth network model. At time t, randomly select a sample vertex. Add a new vertex and an edge between the sample vertex and the new vertex. This step is called the *vertex duplication*. Also at time t, for each neighbor of the sample vertex add an edge between the new vertex and the neighbor vertex with probability a. This step is called *edge duplication*. Chung *et al.* rigorously analyzed the scale-free degree distributions of duplication models [4], [3]. They also established bounds on the maximum degrees. A duplication model with edge rewiring was studied by Solè *et al.* [12]. Through computer simulation, Solè analyzed the clustering coefficient and the average path length. Bhan *et al.* [1] considered three duplication models. Their first model corresponds to Chung's model with $a = 1$. Their second model is the first model with random removal of edges. The third model is the first model with edge rewiring. Through simulations, Bhan *et al.* studied the clustering coefficients, average path lengths and exponents of scale-free degree distributions of the three models. Ispolatov *et al.* [8], [7] studied the average degree and the average number of cliques in a duplication-divergence model. Through computer simulations Zhao *et al.* studied the Pearson degree correlation coefficient for several duplication-divergence models [15]. Boccaletti *et al.* [2] considered a model similar to the duplication model. In Boccaletti's model a new vertex is added to the network at each time. A new vertex randomly selects a vertex from the network and the neighbors of the selected vertex. The new vertex establishes m edges randomly to the selected vertex and its neighbors. Rate equations were derived for the degree distribution and the conditional degree-degree probability of this model.

In this paper we consider the duplication model studied by Chung *et al.* [4], [3]. To the best of our knowledge, both the clustering coefficient and the Pearson degree correlation coefficient of this model have not been studied analytically. The goal of this paper is to derive the clustering coefficient and the Pearson degree correlation coefficient for this network. We now describe Chung's duplication model.

Chung's duplication model:

(i) At time zero there are m_0 vertices, forming an m_0-

D.-S. Lee, C.-S, Chang and H.-N. Chang are with the Institute of Communications Engineering, National Tsing Hua University, Hsinchu 300, Taiwan, R.O.C. email: lds@cs.nthu.edu.tw, cschang@ee.nthu.edu.tw, seenns@hotmail.com

Part of this paper was presented in [9] at *IEEE International Conference on Communications (ICC) 2014*, Sydney, Australia.

clique. That is, the m_0 vertices are fully connected. Any pair of two vertices out of the m_0 vertices is connected by an edge.

(ii) At each time t, where $t = 1, 2, \ldots$, a new vertex is added into the network.

(iii) The new vertex randomly selects an existing vertex in the network and attaches to the selected vertex with an undirected edge.

(iv) Then, each neighbor of the randomly selected vertex is attached to the new vertex with probability p through an undirected edge. This is independent of everything else.

(v) $t \leftarrow t + 1$. Repeat (ii).

Chung *et. al.* called the parameter p the selection probability [3]. We shall use the same terminology and symbol.

Since exactly one vertex is added into the network at each time, there are $m_0 + t$ vertices in the network at time t. For each time t, we label the vertices such that the vertex that is added into the network at time t is vertex m_0+t. For notational simplicity, we denote vertex $m_0 + t + 1$ by N at time $t + 1$. Denote by V the vertex randomly selected by N. Denote the degree of V by X. In addition, denote the neighbors of V by $V_1, V_2, \ldots, V_X$. Let the degree of V_i be denoted by Y_i. Denote the j^{th} neighbor of vertex V_i by V_{ij}. Let Z_{ij} be the degree of V_{ij}. We shall use notation $\mathcal{N}(v)$ to denote the set of neighbors of vertex v. Using this notation, we have $\mathcal{N}(V) = \{V_1, V_2, \ldots, V_X\}$. For vertex V_i, Bernoulli random variable $U_i = 1$ indicates that V_i is attached to vertex $m_0 + t + 1$. Otherwise, V_i is not attached and $U_i = 0$. Random variables $U_1, U_2, \ldots,$ are independent and identically distributed (i.i.d.). Finally, we shall use symbol A_{ij} to indicate the connectivity of vertices i and j. That is, $A_{ij} = 1$ if vertices i and j are connected. Otherwise, $A_{ij} = 0$. We refer the reader to Figure 1 for an illustration of these notations.

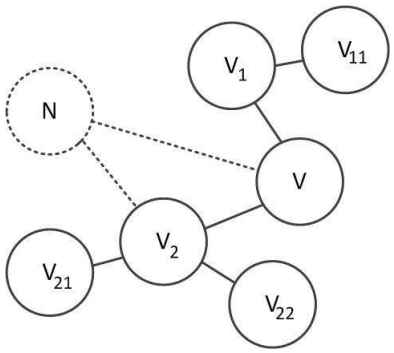

Fig. 1. A new vertex N joins the network and randomly attaches to a vertex, labeled V in the illustration. Vertex V has two existing neighbors V_1 and V_2. Vertex N attaches to V_2 but not to V_1 in this illustration. Newly created vertex and edges are marked with dashed lines in this illustration. Vertices V_1 and V_2 have two and three existing neighbors (not including N) respectively. Vertex V is a common neighbor of V_1 and V_2. Thus, V can be labeled as V_{12} and V_{23}. In this example, $X = 2$, $Y_1 = 2$, $Y_2 = 3$. In addition, $Z_{11} = Z_{21} = Z_{22} = 1$ and $Z_{12} = Z_{23} = 2$.

We summarize our main contributions for this duplication model as follows:

- In Section II, we first consider a simple undirected graph and derive some identities for a randomly selected vertex and the two vertices at the both ends of a randomly selected edge. These identities are needed in our analysis and appear to be of independent interest.
- In Section III, we perform an exact analysis for the clustering coefficient. We consider a randomly selected vertex in the duplication model and derive closed-form expressions for the expected degree in Section III-A, the number of triangles in Section III-B, and the second moment of the degree in Section III-C.
- In Section IV, we derive an approximation for the Pearson degree correlation coefficient. We consider a randomly selected vertex in the duplication model and derive a closed-form expression for the third moment of the degree in Section IV-A. We also consider the two vertices at both ends of a randomly selected edge in the duplication model and derive a closed-form approximation for the product of the degrees of these two vertices in Section IV-B.
- Our expressions for the clustering coefficient and the Pearson degree correlation coefficient are quite complicated. We present an asymptotic approximation by keeping only the dominant terms to simplify the expressions. We numerically study how relative approximation error decreases with time.
- In Section VI, we perform various simulations to verify our analytical results. For the clustering coefficient, numerical results calculated from our approximation and the corresponding simulation results agree extremely well for the whole evolution process. For the Pearson degree correlation coefficient, there is some discrepancy at early times between the simulation results and the numerical results. However, as time goes on, the discrepancy diminishes.

In Table I, we provide a list of notations used in this paper.

TABLE I
LIST OF NOTATIONS

N	The new vertex added to the network
V	The vertex selected by the new vertex N
X	The degree of vertex V
V_i	The i^{th} neighbor of vertex V
Y_i	The degree of vertex V_i
p	The selection probability to add a link between N and V_i
U_j's	i.i.d. Bernoulli r.v.s with mean a
V_{ij}	The j^{th} neighbor of vertex V_i
Z_{ij}	The degree of vertex V_{ij}
$k(t)$	The expected degree at time t
$\tau(t)$	The expected number of triangles at time t
$\alpha(t)$	The second moment of degree at time t
$\beta(t)$	The third moment of degree at time t
$w(t)$	$\mathsf{E}[X \sum_{i=1}^{X} Y_i]$
$\tilde{X}, \tilde{Y}$	The degrees of the two vertices that appear at the two ends of a randomly selected edge
n	The number of vertices in a graph
m	The number of edges in a graph
k_i	The degree of vertex i in a graph
$A = (A_{ij})$	The adjacency matrix of a graph
C	Clustering coefficient
ρ	Pearson degree correlation coefficient

II. Some Identities of Simple Undirected Graphs

In this section, we first prove some identities for simple undirected graphs. These identities are needed in our analysis later and appear to be of independent interest.

A graph G with the $n \times n$ adjacency matrix $A = (A_{ij})$ is called a simple and undirected graph if (i) $A_{ii} = 0$ for all i, (ii) $A_{ij} = A_{ji}$ for all i and j, and (iii) $A_{ij} = 1$ if there is an edge between vertex i and vertex j and 0 otherwise. For such a graph, the degree of vertex i, denoted by k_i, is then

$$k_i = \sum_{i=1}^{n} A_{ij} = \sum_{j=1}^{n} A_{ij}. \tag{1}$$

Also, the total number of edges, denoted by m, in the simple undirected graph G is

$$m = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}. \tag{2}$$

Clearly, every realization of Chung's duplication model at any time t is a simple and undirected graph.

In the following lemma, we first show two identities for a randomly selected *vertex* in a simple undirected graph.

Lemma 1 *Suppose that V is a vertex uniformly selected from a simple undirected graph G. Let X be the degree of vertex V, Y_i, $i = 1, 2, \ldots, X$ be the degree of the i^{th} neighbor of vertex V, and Z_{ij} be the degree of j^{th} neighbor of the i^{th} neighbor of vertex V. Then*

$$\mathsf{E}[X^\nu] = \mathsf{E}\left[\sum_{i=1}^{X} Y_i^{\nu-1}\right], \quad \textit{for all } \nu = 2, 3, \ldots, \tag{3}$$

and

$$\mathsf{E}\left[X \sum_{i=1}^{X} Y_i\right] = \mathsf{E}\left[\sum_{i=1}^{X} \sum_{j=1}^{Y_i} Z_{ij}\right]. \tag{4}$$

Proof. Note from (1) and $A_{ij} = A_{ji}$ that

$$\begin{aligned}
\mathsf{E}[X^\nu] &= \frac{1}{n} \sum_{i=1}^{n} (k_i)^\nu \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} A_{ij}\right)^\nu \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} A_{ij}\right) \left(\sum_{k=1}^{n} A_{ik}\right)^{\nu-1} \\
&= \frac{1}{n} \sum_{j=1}^{n} \left(\sum_{i=1}^{n} A_{ji} \left(\sum_{k=1}^{n} A_{ik}\right)^{\nu-1}\right) \\
&= \frac{1}{n} \sum_{j=1}^{n} \left(\sum_{i=1}^{n} A_{ji} (k_i)^{\nu-1}\right) \\
&= \mathsf{E}\left[\sum_{i=1}^{X} Y_i^{\nu-1}\right].
\end{aligned}$$

Identity (4) follows simply from definitions and easy manipulation of summations. ■

Since every realization of Chung's duplication model at any time t is a simple and undirected graph, the two identities in Lemma 1 still hold for the random network from Chung's duplication model by taking the expectation (over all the realizations) on both sides of the identities. This leads to the following corollary.

Corollary 2 *The two identities in (3) and (4) hold for the random network from Chung's duplication model.*

Our second lemma shows two identities for a randomly selected *edge* in a simple undirected graph.

Lemma 3 *Consider a simple and undirected graph G with m edges. Suppose that we uniformly select an edge from the graph, i.e., with probability $1/m$, an edge is selected among these m edges in the graph. Let $\tilde{X}$ and $\tilde{Y}$ be the degrees of the two vertices at the two ends of the edge. Then for $\nu = 1, 2, \ldots$,*

$$\mathsf{E}[\tilde{X}^\nu] = \frac{\mathsf{E}[X^{\nu+1}]}{\mathsf{E}[X]}, \tag{5}$$

and

$$\mathsf{E}[\tilde{X}\tilde{Y}] = \frac{\mathsf{E}\left[X \sum_{i=1}^{X} Y_i\right]}{\mathsf{E}[X]}, \tag{6}$$

where X is the degree of a uniformly selected vertex from G, and Y_i, $i = 1, 2, \ldots, X$ is the degree of the i^{th} neighbor of that vertex.

Proof. Now we prove (5). Note that the probability that the edge between vertex i and vertex j is selected is A_{ij}/m. To compute $\mathsf{E}[\tilde{X}^\nu]$, we take an average of the ν-th power of the two degrees at the two sides of an edge. Thus, we have from the symmetric property of the adjacency matrix and $A_{ii} = 0$ that

$$\begin{aligned}
\mathsf{E}[\tilde{X}^\nu] &= \frac{1}{m} \sum_{i=1}^{n} \sum_{j>i} A_{ij} \left(\frac{(k_i)^\nu + (k_j)^\nu}{2}\right) \\
&= \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} (k_i)^\nu \\
&= \frac{1}{2m} \sum_{i=1}^{n} (k_i)^{\nu+1}.
\end{aligned}$$

In view of (1) and (2), it is easy to see that $\mathsf{E}[\tilde{X}^\nu]$ is equal to

$$\frac{\frac{1}{n}\sum_{i=1}^{n} (k_i)^{\nu+1}}{\frac{1}{n}\sum_{i=1}^{n} k_i},$$

which by definition is equal to the right side of (5).

Finally, we prove (6). Given graph G, by definition we have

$$\begin{aligned}
&\mathsf{E}\left[X \sum_{i=1}^{X} Y_i\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} A_{ij}\right) \sum_{k=1}^{n} A_{ik} \left(\sum_{\ell=1}^{n} A_{k\ell}\right).
\end{aligned} \tag{7}$$

Also,

$$\mathsf{E}\left[\tilde{X}\tilde{Y}\right] = \frac{1}{m}\sum_{i=1}^{n}\sum_{j>i} A_{ij}(k_i \cdot k_j) \quad (8)$$

$$= \frac{\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\left(\sum_{k=1}^{n} A_{ik}\right)\left(\sum_{\ell=1}^{n} A_{j\ell}\right)}{\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}}. \quad (9)$$

Since $\mathsf{E}[X] = (\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij})/n$,

$$\mathsf{E}[\tilde{X}\tilde{Y}]\mathsf{E}[X] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\left(\sum_{k=1}^{n} A_{ik}\right)\left(\sum_{\ell=1}^{n} A_{j\ell}\right). \quad (10)$$

To prove (6), we need to show that (7) and (10) are equal. We achieve this by simply exchanging the order of the second summation and the third summation in (7). ■

We note that $\mathsf{E}[\tilde{X}]$ is known as the average degree of a neighbor and the identity for $\nu = 1$ is also well-known for the configuration model (see e.g., the book [10]).

Though the two identities in Lemma 1 still hold for the random network from Chung's duplication model, the two identities in Lemma 3 cannot be extended in the same way to Chung's duplication model as they are represented in the ratio form. However, if the average degree, $\mathsf{E}[X]$, for every realization of a random network is a constant, then we can still take the expectation (over all the realizations) on both sides of the two identities in Lemma 3. One possible application is for random networks that are generated by a fixed degree sequence/distribution. As it has been shown that the degree distribution of Chung's duplication model converges to a scale-free distribution as t goes to infinity [4], intuitively one might expect that the average degree, $\mathsf{E}[X]$, would also converge to a constant and thus the two identities in Lemma 3 would (approximately) hold for the random network from Chung's duplication model as t goes to infinity. In view of this, we shall use the two identities in Lemma 3 as approximations for Chung's duplication model.

III. Clustering Coefficient

In this section we perform our analysis for the clustering coefficient. Recall that the clustering coefficient of a deterministic network is defined as the ratio [10]:

$$C = \frac{(\text{number of triangles}) \times 3}{(\text{number of connected triples})}. \quad (11)$$

The clustering coefficient of a random network is defined similarly with the numerator and the denominator in (11) replaced by their expectations [10]. Let $\tau(t)$ be the expected number of triangles that a randomly selected vertex has at time t. Let $n(t)$ be the number of vertices at time t. Since each triangle has three vertices, it follows that the expected number of triangles of the entire network is

$$n(t)\tau(t)/3. \quad (12)$$

Consider a randomly chosen vertex and let its degree be X. The expected total number of connected triples of the network is

$$n(t)\mathsf{E}\left[\binom{X}{2}\right] = n(t)\frac{\alpha(t) - k(t)}{2}, \quad (13)$$

where $\alpha(t)$ is defined as $\mathsf{E}[X^2]$ and $k(t) = \mathsf{E}[X]$. Substituting (12) and (13) into (11), we obtain

$$C = \frac{2\tau(t)}{\alpha(t) - k(t)}. \quad (14)$$

From (14), one needs to evaluate $k(t)$, $\tau(t)$ and $\alpha(t)$ in order to compute the clustering coefficient C. Differential equations for $k(t)$, $\tau(t)$ and $\alpha(t)$ will be shown later in Section III-A, Section III-B and Section III-C, respectively. By solving these differential equations, we obtain closed-form expressions for $k(t)$ in (18), $\tau(t)$ in (27), and $\alpha(t)$ in (48), respectively.

A. Expected Degree

Let $k(t)$ denote the expected degree at time t. We shall derive a differential equation for $k(t)$. We will first derive a difference equation by equating the total expected number of edges in the network right before and after time t. We then approximate the difference equation by a differential equation.

Since there are $m_0 + t$ vertices in the network at time t, clearly the total expected degree of the entire network is $(m_0 + t)k(t)$ at time t. Since each edge has two ends, the expected number of edges at time t is $(m_0+t)k(t)/2$. The new vertex that arrives at time t randomly selects and attaches to an existing vertex. This produces a new edge. With probability p a new edge is formed between each neighbor of the randomly selected vertex and the new vertex. Recall that the parameter p is called the selection probability. Since the expected degree of the randomly selected vertex is $k(t)$, additional $pk(t)$ edges are generated on average from step (iv) in the duplication model. Thus, on average $1 + pk(t)$ new edges are generated. Thus, we have

$$\frac{(m_0 + t + 1)k(t+1)}{2} = \frac{(m_0 + t)k(t)}{2} + 1 + pk(t). \quad (15)$$

Eq. (15) is a difference equation. We propose to approximate this difference equation by a differential equation. To achieve this, we rewrite (15) as

$$k(t+1) - k(t) = \frac{2 + (2p-1)k(t)}{m_0 + t + 1}. \quad (16)$$

We approximate the left hand side of (16) by derivative $k'(t)$ and obtain

$$k'(t) = \frac{2 + (2p-1)k(t)}{m_0 + t + 1} \quad (17)$$

with initial condition $k(0) = m_0 - 1$. If $p \neq 1/2$, (17) is a separable differential equation whose solution is

$$k(t) = k_1(m_0 + t + 1)^{2p-1} + k_2, \quad (18)$$

where constant k_1 is determined by the initial condition $k(0) = m_0 - 1$, *i.e.*

$$k_1 = \left(m_0 - 1 + \frac{2}{2p-1}\right)(m_0 + 1)^{1-2p}. \quad (19)$$

Constant k_2 in (18) is given by

$$k_2 = 2/(1-2p). \tag{20}$$

If $p = 1/2$, the coefficient of $k(t)$ in the right side of (17) is zero and this differential equation can be solved easily. The solution is

$$k(t) = 2\log(m_0 + t + 1) + c_1, \tag{21}$$

where

$$c_1 = m_0 - 1 - 2\log(m_0 + 1). \tag{22}$$

We remark here that all closed-form solutions of differential equations have been verified independently using Mathematica [14].

B. Expected Number of Triangles

In this section we shall first derive a difference equation for $\tau(t)$ by equating the expected total number of triangles in the network right before and after time t. We then approximate the difference equation by a differential equation.

In the duplication model, the total number of vertices at time t is $n(t) = m_0 + t$. Since each triangle has three vertices, the expected total number of triangles at time t is $(m_0+t)\tau(t)/3$. Thus, we reach the following identity

$$\frac{(m_0+t+1)\tau(t+1)}{3} = \frac{(m_0+t)\tau(t)}{3} + pk(t) + p^2\tau(t). \tag{23}$$

We now explain the last two terms in the right hand side of (23). Suppose that at time t a new vertex N is attached to vertex V as shown in Figure 2. With probability a, an edge between vertex N and a neighbor of V is established. These two new edges introduce two new triangles NVV_1 and NVV_2. In addition, with probability p^2 triangle NV_1V_2 is formed. Thus, $k(t)$ new triangles (on average) could be introduced. Each is introduced with probability p independently. In addition, $\tau(t)$ new triangles (on average) could also be introduced. Each is introduced with probability p^2 independently.

Now we approximate $\tau(t+1) - \tau(t)$ by $\tau'(t)$. Eq. (23) can be approximated by the following differential equation

$$\tau'(t) = \frac{3pk(t) + (3p^2-1)\tau(t)}{m_0 + t + 1} \tag{24}$$

with initial condition

$$\tau(0) = (m_0 - 1)(m_0 - 2)/2. \tag{25}$$

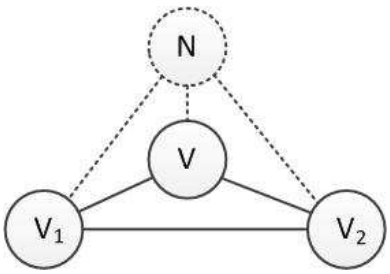

Fig. 2. A new vertex, denoted by vertex N, is attached to vertex V. Two new triangles NVV_1 and NVV_2 are formed, each with probability p. Triangle NV_1V_2 is formed with probability p^2.

In general, (24) is a first-order linear differential equation that can be solved by the technique of integrating factors. Specifically,

$$\tau(t) = c(m_0+t+1)^{3p^2-1} + (m_0+t+1)^{3p^2-1}\int \frac{3pk(t)}{(m_0+t+1)^{3p^2}}\,dt, \tag{26}$$

where c is a constant to be determined by the initial condition in (25). If $p \neq 0, 1/2$ or $2/3$, substitute (18) into (26) and integrate. We obtain

$$\tau(t) = \tau_1(m_0+t+1)^{3p^2-1} + \tau_2(m_0+t+1)^{2p-1} + \tau_3, \tag{27}$$

where constant τ_1 is determined by the initial condition in (25), *i.e.*

$$\tau_1 = \Big(\frac{(m_0-1)(m_0-2)}{2} - \frac{3pk_1(m_0+1)^{2p-1}}{2p-3p^2} + \frac{6p}{(2p-1)(1-3p^2)}\Big)(m_0+1)^{1-3p^2}. \tag{28}$$

Constants τ_2 and τ_3 in (27) are given by

$$\tau_2 = \frac{3pk_1}{2p-3p^2} \tag{29}$$

$$\tau_3 = \frac{6p}{(1-2p)(1-3p^2)}. \tag{30}$$

Solution (27) is not valid for $p = 0, 1/2, 2/3$, and $1/\sqrt{3}$. The constant τ_1 in (28) is not defined when p is 0 or $2/3$. If $p = 1/2$, substitute (21) into (26) and get the solution

$$\tau(t) = 6(2\log(m_0+t+1) - 8 + c_1) + c_2(m_0+t+1)^{-1/4}. \tag{31}$$

If $p = 1/\sqrt{3}$, the coefficient of $\tau(t)$ on the right side of (24) vanishes and (24) can simply be solved by integration. Its solution is

$$\tau(t) = \sqrt{3}(3+2\sqrt{3})\big(k_1(m_0+t+1)^{-1+2/\sqrt{3}} - 2\log(m_0+t+1)\big) + c_3. \tag{32}$$

If $p = 0$, the integral in (26) yields a logarithmic term and the solution is

$$\tau(t) = \frac{c_4}{m_0+t+1}. \tag{33}$$

We can similarly derive the solution for $p = 2/3$.

$$\tau(t) = (m_0+t+1)^{1/3}(c_5 + 2k_1\log(m_0+t+1)) + 36. \tag{34}$$

Constants c_2, c_3, c_4 and c_5 are determined through the initial condition in (25).

C. Second Moment of Degree

In this section we shall derive a differential equation for $\mathsf{E}[X^2]$ (denoted by $\alpha(t)$). We first study the difference between $(m_0+t+1)\alpha(t+1)$ and $(m_0+t)\alpha(t)$. Suppose that the degree of a vertex is changed from X to $X + \Delta X$, this vertex will contribute

$$(X+\Delta X)^2 - X^2 = (\Delta X)^2 + 2\Delta X \cdot X \tag{35}$$

toward to the difference. In view of the duplication model, there are three cases that the degree of a vertex is changed from time t to $t+1$.

- The new vertex: Vertex N is added into the network at time t. The second moment of its degree must be included in the difference, *i.e.,*
$$\left(1+\sum_{i=1}^{X} U_i\right)^2. \tag{36}$$
- The selected vertex: The degree of vertex V is changed from X to $X+1$. According to (35), vertex V contributes
$$1+2X \tag{37}$$
to the difference.
- The neighbors of the selected vertex: The i-th neighbor, denoted by V_i, may change its degree from Y_i to Y_i+1, if $U_i=1$. If $U_i=0$, the degree of V_i remains unchanged at Y_i. One can identify X with Y_i and ΔX with U_i in (35). According to (35), the contribution of vertex V_i is
$$U_i^2+2U_iY_i.$$
Summing up all the contributions from all the neighbors of V and using the identity $U_i^2=U_i$, we obtain the total contribution
$$\sum_{i=1}^{X}(1+2Y_i)U_i. \tag{38}$$

We first analyze the expectation of (36). Using conditional expectations, we have

$$\begin{aligned}\mathsf{E}\left[\left(\sum_{i=1}^{X} U_i\right)^2\right] &= p\mathsf{E}[X]+p^2\mathsf{E}[X(X-1)]\\ &= pk(t)+p^2(\alpha(t)-k(t)).\end{aligned} \tag{39}$$

Hence,

$$\mathsf{E}\left[\left(1+\sum_{i=1}^{X} U_i\right)^2\right]=1+3pk(t)+p^2(\alpha(t)-k(t)) \tag{40}$$

The expectation of (37) is simply

$$\mathsf{E}[2X+1]=2k(t)+1. \tag{41}$$

We now analyze the expectation of (38). Again using the technique of conditional expectations, we have

$$\begin{aligned}&\mathsf{E}\left[\sum_{i=1}^{X} Y_iU_i\right]=\mathsf{E}\left[\mathsf{E}\left[\sum_{i=1}^{X} Y_iU_i\right]\right]\\ &=\mathsf{E}\left[\sum_{i=1}^{X}\mathsf{E}[Y_iU_i\,|\,X]\right]=\mathsf{E}\left[\sum_{i=1}^{X}\mathsf{E}[Y_i|X]\mathsf{E}[U_i|X]\right]\\ &=\mathsf{E}\left[\sum_{i=1}^{X} p\mathsf{E}[Y_i|X]\right]=p\mathsf{E}\left[\mathsf{E}\left[\sum_{i=1}^{X} Y_i|X\right]\right]\\ &=p\mathsf{E}\left[\sum_{i=1}^{X} Y_i\right]\\ &=p\mathsf{E}[X^2]=p\alpha(t),\end{aligned} \tag{42}$$

where we use (3) (for $\nu=2$) in the last identity. Thus, the expectation of (38) is

$$\mathsf{E}\left[\sum_{i=1}^{X}(2Y_i+1)U_i\right]=2p\alpha(t)+pk(t). \tag{43}$$

It follows from (40), (41) and (43) that

$$\begin{aligned}&(m_0+t+1)\alpha(t+1)-(m_0+t)\alpha(t)=\\ &\quad(1+3pk(t)+p^2(\alpha(t)-k(t))+(2k(t)+1)\\ &\quad+(2p\alpha(t)+pk(t)).\end{aligned} \tag{44}$$

Equation (44) is a difference equation and it can be approximated by first-order linear differential equation

$$\begin{aligned}\alpha'(t) &= \frac{1}{m_0+t+1}\Big((-1+p^2+2p)\alpha(t)+2\\ &\quad+(2-p^2+4p)k(t)\Big)\end{aligned} \tag{45}$$

with initial condition

$$\alpha(0)=(m_0-1)^2. \tag{46}$$

Applying the technique of integrating factors, we obtain the solution as follows

$$\begin{aligned}\alpha(t) &= (m_0+t+1)^{p^2+2p-1}\Big(c+\\ &\quad\int\frac{2+(-p^2+4p+2)k(t)}{(m_0+t+1)^{p^2+2p}}\,dt\Big),\end{aligned} \tag{47}$$

where constant c is determined from the initial condition in (46). If $p\neq 0, 1/2$, or $\sqrt{2}-1$, one can substitute (18) into (47) and obtain a general solution

$$\alpha(t)=\alpha_1(m_0+t+1)^{p^2+2p-1}+\alpha_2(m_0+t+1)^{2p-1}+\alpha_3, \tag{48}$$

where constant α_1 is determined from the initial condition in (46). Constants α_2 and α_3 are given by

$$\alpha_2 = \frac{p^2-4p-2}{p^2}k_1 \tag{49}$$

$$\alpha_3 = \frac{2p^2-4p-6}{(2p-1)(-p^2-2p+1)}. \tag{50}$$

If $p=0$, the integral in (47) produces a logarithmic term and the solution is

$$\alpha(t)=\frac{2k_1\log(m_0+t+1)+6t+c_6}{m_0+t+1}. \tag{51}$$

If $p=1/2$, one substitutes (21) into (47) and obtain

$$\begin{aligned}\alpha(t) &= c_7(m_0+t+1)^{1/4}-30\log(m_0+t+1)-\\ &\quad 15c_1-128.\end{aligned} \tag{52}$$

If $p=\sqrt{2}-1$, the coefficient of $\alpha(t)$ on the right side of (45) vanishes and the differential equation can be solved by integration. The solution is

$$\begin{aligned}\alpha(t) &= 4(5+4\sqrt{2})\log(m_0+t+1)-\\ &\quad(9+8\sqrt{2})k_1(m_0+t+1)^{2\sqrt{2}-3}+c_8.\end{aligned} \tag{53}$$

Constants c_6, c_7 and c_8 are determined from the initial condition (46).

IV. Pearson Degree Correlation Coefficient

In this section, we derive an approximation for the Pearson degree correlation coefficient in the duplication model.

We first give the definition of the Pearson degree correlation coefficient of a random network. Randomly select an edge from the network. Let $\tilde{X}$ and $\tilde{Y}$ be the degrees of the two vertices at the two ends of the edge. The Pearson degree correlation coefficient of the network is defined as the correlation coefficient of $\tilde{X}$ and $\tilde{Y}$, *i.e.,*

$$\rho(\tilde{X},\tilde{Y}) = \frac{\mathsf{E}[\tilde{X}\tilde{Y}] - \mathsf{E}[\tilde{X}]\mathsf{E}[\tilde{Y}]}{\sigma_{\tilde{X}}\sigma_{\tilde{Y}}}, \tag{54}$$

where $\sigma_{\tilde{X}}$ and $\sigma_{\tilde{Y}}$ are the standard deviation of random variables $\tilde{X}$ and $\tilde{Y}$ respectively [10]. Since $\tilde{X}$ and $\tilde{Y}$ are statistically indistinguishable and thus identically distributed, (54) can be rewritten as

$$\rho(\tilde{X},\tilde{Y}) = \frac{\mathsf{E}[\tilde{X}\tilde{Y}] - (\mathsf{E}[\tilde{X}])^2}{\mathsf{E}[\tilde{X}^2] - (\mathsf{E}[\tilde{X}])^2}. \tag{55}$$

The expectations in (55) were studied in Lemma 3 for a deterministic undirected graph. However, the duplication model is a random network and the identities in (5) and (6) can not be extended to the averages of all the realizations of a random network. To evaluate (55), we shall approximate (5) and (6) by

$$\mathsf{E}[\tilde{X}^\nu] \approx \frac{\mathsf{E}[X^{\nu+1}]}{\mathsf{E}[X]} \tag{56}$$

and

$$\mathsf{E}[\tilde{X}\tilde{Y}] \approx \frac{\mathsf{E}\left[X\sum_{i=1}^{X} Y_i\right]}{\mathsf{E}[X]} \tag{57}$$

respectively. Recall that (5) and (6) are valid for a deterministic graph. In a ratio form, (5) and (6) are in general not true for random graphs. However, from Chung's study and our simulation study, $\mathsf{E}[X]$ converges to a constant as time becomes large. The approximation in (56) and (57) becomes more accurate as t goes to infinity.

Let $\beta(t) = \mathsf{E}[X^3]$ and $w(t) = \mathsf{E}[X\sum_{i=1}^{X} Y_i]$. Then we can approximate the Pearson degree correlation coefficient in the duplication model as follows:

$$\rho(\tilde{X},\tilde{Y}) \approx \frac{\frac{w(t)}{k(t)} - (\frac{\alpha(t)}{k(t)})^2}{\frac{\beta(t)}{k(t)} - (\frac{\alpha(t)}{k(t)})^2}, \tag{58}$$

where $k(t) = \mathsf{E}[X]$ and $\alpha(t) = \mathsf{E}[X^2]$ are derived in the previous section. In Section IV-A, we shall derive a differential equation for $\beta(t)$. Then in Section IV-B, we shall use the identity in (4) to derive a differential equation for $w(t)$. By solving these differential equations, we obtain closed-form expressions for $\beta(t)$ in (72) and $w(t)$ in (103), respectively.

A. Third Moment of Degree

In this section we shall derive a differential equation for $\mathsf{E}[X^3]$ (denoted by $\beta(t)$). The argument is parallel to that in Section III-C for $\mathsf{E}[X^2]$. Consider the difference between $(m_0+t+1)\beta(t+1)$ and $(m_0+t)\beta(t)$. Suppose that the degree of a vertex is changed from X to $X+\Delta X$, this vertex will contribute

$$(X+\Delta X)^3 - X^3 = (\Delta X)^3 + 3(\Delta X)^2\cdot X + 3\Delta X\cdot X^2 \tag{59}$$

toward to the difference. In view of the duplication model, there are three cases that the degree of a vertex is changed from time t to $t+1$.

- The new vertex: Vertex N is added into the network at time t. The third moment of its degree must be included in the difference, *i.e.,*
$$\left(1+\sum_{i=1}^{X} U_i\right)^3. \tag{60}$$
- The selected vertex: The degree of vertex V is changed from X to $X+1$. According to (59), vertex V contributes
$$1 + 3X + 3X^2 \tag{61}$$
to the difference.
- The neighbors of the selected vertex: Vertex V_i may change its degree from Y_i to Y_i+1, if $U_i = 1$. According to (59), the total contribution in this case is
$$\sum_{i=1}^{X}(1 + 3Y_i + 3Y_i^2)U_i \tag{62}$$

We first analyze the expectation of (60). Using conditional expectations, we have

$$\begin{aligned}
&\mathsf{E}\left[\left(\sum_{i=1}^{X} U_i\right)^3\right] \\
&= p\mathsf{E}[X] + p^2\mathsf{E}[3X(X-1)] + p^3\mathsf{E}[X(X-1)(X-2)] \\
&= p\cdot k(t) + 3p^2(\alpha(t) - k(t)) + p^3(\beta(t) - 3\alpha(t) + 2k(t)) \\
&= p^3\beta(t) + (3p^2 - 3p^3)\alpha(t) + (p - 3p^2 + 2p^3)k(t).
\end{aligned} \tag{63}$$

Hence, using (40) and (63) yields

$$\begin{aligned}
&\mathsf{E}\left[\left(1+\sum_{i=1}^{X} U_i\right)^3\right] \\
&= 1 + 3pk(t) + 3\mathsf{E}\left[\left(\sum_{i=1}^{X} U_i\right)^2\right] \\
&\quad + \mathsf{E}\left[\left(\sum_{i=1}^{X} U_i\right)^3\right] \\
&= 1 + p^3\beta(t) + (6p^2 - 3p^3)\alpha(t) \\
&\quad + (7p - 6p^2 + 2p^3)k(t).
\end{aligned} \tag{64}$$

The expectation of (61) is simply

$$\mathsf{E}[1 + 3X + 3X^2] = 1 + 3k(t) + 3\alpha(t). \tag{65}$$

We now analyze the expectation of (62). Again using the technique of conditional expectations as in (42), we have

$$\begin{aligned}
&\mathsf{E}\left[\sum_{i=1}^{X} U_i(1 + 3Y_i + 3Y_i^2)\right] \\
&= \mathsf{E}\left[a\sum_{i=1}^{X}(1 + 3Y_i + 3Y_i^2)\right].
\end{aligned} \tag{66}$$

By (3), (66) is equal to

$$a(k(t)+3\alpha(t)+3\beta(t)). \tag{67}$$

From (64), (65) and (67), we reach the following balance equation

$$\begin{aligned}(m_0+t+1)\beta(t+1) &= (m_0+t)\beta(t)+\\ &\quad 1+p^3\beta(t)+(6p^2-3p^3)\alpha(t)+\\ &\quad +(7p-6p^2+2p^3)k(t)+\\ &\quad 1+3k(t)+3\alpha(t)+\\ &\quad a(k(t)+3\alpha(t)+3\beta(t)). \end{aligned}\tag{68}$$

Equation (68) is a difference equation and it can be approximated by first-order linear differential equation

$$\begin{aligned}\beta'(t) &= \frac{1}{m_0+t+1}\Big((p^3+3p-1)\beta(t)+2\\ &\quad +(3p+6p^2-3p^3+3)\alpha(t)\\ &\quad +(8p-6p^2+2p^3+3)k(t)\Big)\end{aligned}\tag{69}$$

with initial condition

$$\beta(0)=(m_0-1)^3. \tag{70}$$

Using the technique of integrating factors, we obtain the following solution

$$\begin{aligned}\beta(t) &= (m_0+t+1)^{p^3+3p-1}\Bigg(c+\\ &\quad \int \frac{2+(-3p^3+6p^2+3p+3)\alpha(t)}{(m_0+t+1)^{p^3+3p}}\,dt+\\ &\quad \int \frac{(2p^3-6p^2+8p+3)k(t)}{(m_0+t+1)^{p^3+3p}}\,dt\Bigg),\end{aligned}\tag{71}$$

where constant c is determined from the initial condition in (70). For general values of p, we substitute $k(t)$ in (18) and $\alpha(t)$ in (48) into (71) and obtain

$$\begin{aligned}\beta(t) &= \beta_1(m_0+t+1)^{p^3+3p-1}+\beta_2(m_0+t+1)^{p^2+2p-1}\\ &\quad +\beta_3(m_0+t+1)^{2p-1}+\beta_4,\end{aligned}\tag{72}$$

where constant β_1 is determined from the initial condition in (70). Constants β_2, β_3 and β_4 are given by

$$\beta_2 = \frac{3\alpha_1(p^3-2p^2-p-1)}{p(p^2-p+1)} \tag{73}$$

$$\beta_3 = \frac{k_1(p^5-12p^4+7p^3+18p^2+18p+6)}{p^5+p^3} \tag{74}$$

$$\beta_4 = \frac{2(p^5-10p^4+8p^3-p^2+13p+13)}{(1-2p)(p^2+2p-1)(p^3+3p-1)}. \tag{75}$$

The solution expressed in (72) is not valid for special cases in which p is $0, 1/2, \sqrt{2}-1$ and the unique root of $p^3+3p=1$ in the interval $(0,1)$. It can be shown that the unique root in the interval $(0,1)$ is

$$r=\left(\frac{\sqrt{5}+1}{2}\right)^{1/3}-\left(\frac{\sqrt{5}-1}{2}\right)^{1/3}. \tag{76}$$

If $p=0$, substitute $\alpha(t)$ from (51) and $k(t)$ from (18) into (71) and we have

$$\begin{aligned}\beta(t) &= \frac{c_9+26t}{m_0+t+1}\\ &\quad +\frac{3(c_9+k_1-6(m_0+1))\log(m_0+t+1)}{m_0+t+1}\\ &\quad +\frac{3k_1(\log(m_0+t+1))^2}{m_0+t+1}.\end{aligned}\tag{77}$$

If $p=\sqrt{2}-1$, we substitute $\alpha(t)$ from (52) and $k(t)$ from (18) into (71) and we obtain

$$\begin{aligned}\beta(t) &= c_{10}(m_0+t+1)^{-11+8\sqrt{2}}+\frac{16(94-69\sqrt{2})}{-2052+1451\sqrt{2}}\\ &\quad -\frac{(11118-7865\sqrt{2})k_1}{-2052+1451\sqrt{2}}(m_0+t+1)^{-3+2\sqrt{2}}\\ &\quad +\frac{12(1486+1093\sqrt{2})}{-2052+1451\sqrt{2}}\log(m_0+t+1).\end{aligned}\tag{78}$$

If $p=1/2$, substituting $\alpha(t)$ from (52) and $k(t)$ from (21) into (71) we obtain

$$\begin{aligned}\beta(t) &= c_{11}(m_0+t+1)^{5/8}+\\ &\quad \frac{1258}{5}\log(t)-15c_7(m_0+t+1)^{1/4}+\\ &\quad \frac{38784+3145c_1}{25}.\end{aligned}\tag{79}$$

Constants c_9, c_{10} and c_{11} are determined by the initial condition in (70). If $p=r$, the solution is too complicated and messy to be expressed in closed form. We suggest that (69) be solved numerically when $p=r$. We remark here that all closed-form solutions of differential equations have been verified independently using Mathematica [14].

B. Product of Degrees

In this section, we shall derive a differential equation for $w(t)$. This is done by deriving a balance equation that relates $w(t)$ and $w(t+1)$. By definition, $w(t)\cdot(m_0+t)$ equals to

$$\sum_{i=1}^{m_0+t} X_i \sum_{j=1}^{X_i} Y_{ij}, \tag{80}$$

where X_i is the degree of vertex i before N is added into the network and Y_{ij} is the degree of the j-th neighbor of vertex i. Similarly, $w(t+1)\cdot(m_0+t+1)$ equals to

$$\sum_{i=1}^{m_0+t+1} \hat{X}_i \sum_{j=1}^{\hat{X}_i} \hat{Y}_{ij}, \tag{81}$$

where $\hat{X}_i$ is the new degree of vertex i after vertex N has been introduced into the network. Similarly, $\hat{Y}_{ij}$ is the new degree of the j^{th} neighbor of vertex i. We need to express the quantity in (81) as the sum of the quantity in (80) and an increment. To find the increment, we rewrite (81) as

$$\hat{X}_{m_0+t+1}\cdot\sum_{j=1}^{\hat{X}_{m_0+t+1}}\hat{Y}_{m_0+t+1,j}+\sum_{i=1}^{m_0+t}\hat{X}_i\sum_{j=1}^{\hat{X}_i}\hat{Y}_{ij}. \tag{82}$$

We first analyze the first term in (82). Recall that vertex N is the new vertex that is introduced into the network. Thus, vertex N is also vertex m_0+t+1. Clearly,

$$\hat{X}_{m_0+t+1} = 1 + \sum_{i=1}^{X} U_i$$

and

$$\hat{Y}_{m_0+t+1,j} = \begin{cases} X+1, & \text{if the } j^{th} \text{ neighbor is vertex } V, \\ (Y_j+1)U_j, & \text{otherwise.} \end{cases}$$

Thus, the first term in (82) is equal to

$$\left(1 + \sum_{i=1}^{X} U_i\right)\left((X+1) + \sum_{j=1}^{X}(Y_j+1)U_j\right). \tag{83}$$

Now we analyze the second term in (82). We let $\hat{X}_i = X_i + \Delta X_i$ and $\hat{Y}_{ij} = Y_{ij} + \Delta Y_{ij}$, where ΔX_i and ΔY_{ij} are the increments of X_i and Y_{ij} respectively due to the introduction of vertex N. The second term in (82) corresponds to vertices that are already present in the network before vertex N is introduced and attached. Since vertex i, where $1 \le i \le m_0+t$, is an existing vertex, the change of i's degree can be at most one. That is, ΔX_i is either one or zero. Thus, the second term in (82) can be rewritten as

$$\begin{aligned}
&\sum_{i=1}^{m_0+t} \hat{X}_i \sum_{j=1}^{\hat{X}_i} \hat{Y}_{ij} \\
&= \sum_{i=1}^{m_0+t} (X_i + \Delta X_i) \sum_{j=1}^{X_i+\Delta X_i} (Y_{ij} + \Delta Y_{ij}) \\
&= \sum_{i=1}^{m_0+t} \Bigg\{ X_i \sum_{j=1}^{X_i} Y_{ij} + && (84) \\
&\quad X_i \sum_{j=1}^{X_i} \Delta Y_{ij} + && (85) \\
&\quad \Delta X_i \sum_{j=1}^{X_i} Y_{ij} + && (86) \\
&\quad \Delta X_i \sum_{j=1}^{X_i} \Delta Y_{ij} + && (87) \\
&\quad (X_i + \Delta X_i)(Y_{i,X_i+\Delta X_i} + \Delta Y_{i,X_i+\Delta X_i}) \Bigg\}. && (88)
\end{aligned}$$

Note that the term in (84) is equal to $(m_0+t)w(t)$.

In view of the above analysis, to derive the increment from (80) to (81), one considers all vertices in the network at time $t+1$. The contribution of the new vertex N is expressed in (83). Next, one considers each existing vertex whose degree has changed. One also needs to consider existing vertices whose degrees have not changed, but the degrees of at least one of their neighbors have changed. Consider vertex i. If i's degree has changed, this change causes an increment corresponding to the term in (86), and possibly the term in (87) if at least one of i's neighbors has also changed its degree. If i's degree has not changed but one of its neighbors has changed its degree, this causes an increment corresponding to the term in (85). Note that the above two cases assume that i and its neighbors are connected by edges that exist before vertex N is attached. The term in (88) corresponds to the case in which vertex i acquires a new edge. For edges that are not present before N is attached, the products of degrees at the two sides of such edges must be included in the calculation of the difference between (81) and (80). To make subsequent discussions easier to understand, we shall refer to vertex i in (85) to (88) as a *focused* vertex.

In the following we shall enumerate and identify each vertex in the network as a focused vertex. For each focused vertex, we shall discuss their contributions to the increment. We shall identify their contributions to the increment with the terms in (85) up to (88). Note that vertices whose degrees have not changed and none of their neighbors have changed their degrees are simply ignored in the analysis, as they make no contribution to the increment.

- Identify vertex N as a focused vertex. Its corresponding increment is given in (83).
- Identify vertex V as a focused vertex. The degree of V before and after vertex N is added is X and $X+1$, respectively. The degrees of V's existing neighbors may change from Y_j to $Y_j + U_j$. In addition, V has a new neighbor. The new edge between N and V contributes $(X+1)\left(1+\sum_{i=1}^{X} U_i\right)$ to the increment. This increment corresponds to (88). The degree change of V contributes $1 \cdot \sum_{i=1}^{X}(Y_i+U_i)$. This corresponds to (86) and (87). The degree changes of V's neighbors contribute $X\sum_{i=1}^{X} U_i$. This corresponds to (85). Combining the above, the total increment due to the identification of V as a focused vertex is

$$(X+1)\left(1+\sum_{i=1}^{X} U_i\right) + 1 \cdot \sum_{i=1}^{X}(Y_i+U_i) + X \cdot \sum_{i=1}^{X} U_i. \tag{89}$$

- Identify an existing neighbor of V, say vertex V_i, as a focused vertex. If V_i is attached to N, the new edge between these two vertices produces a contribution of $(Y_i+1)U_i \cdot \left(1+\sum_{j=1}^{X} U_j\right)$. This corresponds to (88). The change of degree of V_i produces a contribution of $U_i \sum_{j=1}^{Y_i} Z_{ij}$ through the edges between V_i and its existing neighbors. This increment corresponds to (86). Now consider V_i's neighbors. Consider the possibility that V_i's neighbors may change their degrees. One of V_i's neighbor is V, which surely changes its degree. This causes an increment $\sum_{i=1}^{X}(Y_i+U_i) \cdot 1$. This corresponds to the sum of terms in (85) and (87). In addition, suppose that there exists a vertex V_j such that V_i, V_j and V form a triangle and V_j also establishes a new edge with N. This case causes increments of amount

$$\sum_{i=1}^{X}\sum_{j=1}^{X} A_{V_iV_j} U_j Y_i$$

and amount

$$\sum_{i=1}^{X}\sum_{j=1}^{X} A_{V_iV_j}U_jU_i.$$

These increments correspond to (85) and (87), respectively. Overall, the total increment due to the neighbors of V is

$$\sum_{i=1}^{X}(Y_i+1)U_i\cdot\left(1+\sum_{j=1}^{X}U_j\right)+\sum_{i=1}^{X}U_i\sum_{j=1}^{Y_i}Z_{ij} +\sum_{i=1}^{X}\sum_{j=1}^{X}A_{V_iV_j}U_j(Y_i+U_i)+1\cdot\sum_{i=1}^{X}(Y_i+U_i). \tag{90}$$

- Consider the neighbors of V_i. One of V_i's neighbors clearly is V. The neighbors of V_i can be another neighbor of V besides V_i, say V_j, if V_i, V and V_j form a triangle. Finally the neighbors of V_i can be a vertex that is neither V nor a neighbor of V. Since the first two cases have been analyzed already, we consider the last case. Consider the j^{th} neighbor of V_i. Denote this vertex by V_{ij}. Now identify vertex V_{ij} as a focused vertex. Since V_{ij} is not a neighbor of V, the degree of V_{ij} must be unchanged after N is introduced to the network. However, V_{ij}'s neighbor, vertex V_i, has a degree change of U_i. Thus, the edge between V_{ij} and V_i produces an increment of $Z_{ij}\cdot U_i$. This increment corresponds to the term in (85). The increment is

$$\sum_{i=1}^{X}U_i\sum_{\substack{j=1\\V_j\neq V}}^{Y_i}Z_{ij}(1-A_{V_{ij}V}). \tag{91}$$

To see the correspondence between (91) and (85), one must identify Z_{ij} and U_i in (91) with X_i and ΔY_{ij} in (85), respectively.

To derive a balance equation, one needs to derive expectations for (83), (89), (90) and (91). Before deriving the expectations, we first simplify the sum of (83), (89), (90) and (91). It is a routine task to show that the sum is

$$\begin{aligned}&2(X+1)\left(1+\sum_{i=1}^{X}U_i\right)\\&+2\left(1+\sum_{i=1}^{X}U_i\right)\left(\sum_{i=1}^{X}(Y_i+1)U_i\right)\\&+2\sum_{i=1}^{X}U_i\sum_{j=1}^{Y_i}Z_{ij}+2\left(\sum_{i=1}^{X}(Y_i+U_i)\right)\\&+\sum_{i=1}^{X}\sum_{j=1}^{X}A_{V_iV_j}U_j(Y_i+U_i)\\&-\sum_{i=1}^{X}U_i\sum_{j=1}^{X}Y_jA_{V_iV_j}.\end{aligned} \tag{92}$$

One rewrites the difference between the last two terms in (92) as

$$\begin{aligned}&\sum_{i=1}^{X}\sum_{j=1}^{X}A_{V_iV_j}U_j(Y_i+U_i)-\sum_{i=1}^{X}U_i\sum_{j=1}^{X}Y_jA_{V_iV_j}\\&=2\sum_{i=1}^{X}\sum_{j=1}^{X}\frac{U_iU_j}{2}A_{V_iV_j}.\end{aligned}$$

Substituting the above identity into (92), we obtain the total increment

$$\begin{aligned}&2(X+1)\left(1+\sum_{i=1}^{X}U_i\right)\\&+2\left(1+\sum_{i=1}^{X}U_i\right)\left(\sum_{i=1}^{X}(Y_i+1)U_i\right)\\&+2\sum_{i=1}^{X}U_i\sum_{j=1}^{Y_i}Z_{ij}+2\left(\sum_{i=1}^{X}(Y_i+U_i)\right)\\&+2\sum_{i=1}^{X}\sum_{j=1}^{X}\frac{U_iU_j}{2}A_{V_iV_j}.\end{aligned} \tag{93}$$

We now derive the expectation of (93). Using conditional expectations, it is easy to show that

$$\mathsf{E}\left[\sum_{i=1}^{X}Y_iU_i\right]=p\mathsf{E}\left[\sum_{i=1}^{X}Y_i\right]=p\alpha(t). \tag{94}$$

Similarly, using conditional expectations, it is easy to show that

$$\mathsf{E}\left[\sum_{i=1}^{X}U_i\sum_{j=1}^{Y_i}Z_{ij}\right]=p\mathsf{E}\left[\sum_{i=1}^{X}\sum_{j=1}^{Y_i}Z_{ij}\right]=pw(t). \tag{95}$$

The second equalities in (94) and (95) are due to (3) and (4) respectively. In addition, using conditional expectations, we have

$$\begin{aligned}\mathsf{E}\left[X\sum_{i=1}^{X}U_i\right]&=\mathsf{E}\left[\mathsf{E}\left[X\sum_{i=1}^{X}U_i\right]\middle|X\right]\\&=\mathsf{E}\left[\mathsf{E}[X\cdot a\cdot X|X]\right]\\&=p\alpha(t).\end{aligned} \tag{96}$$

Finally, using conditional expectations, we have

$$\begin{aligned}&\mathsf{E}\left[\sum_{i=1}^{X}\sum_{j=1}^{X}U_iU_jY_j\right]\\&=\mathsf{E}\left[\left[\sum_{i=1}^{X}\sum_{j=1}^{X}U_iU_jY_j\middle|X\right]\right]\\&=\mathsf{E}\left[\sum_{i=1}^{X}\sum_{j=1}^{X}\mathsf{E}[Y_j|X]\mathsf{E}[U_iU_j|X]\right]\\&=a\mathsf{E}\left[\sum_{i=1}^{X}Y_i\right]+\mathsf{E}\left[\sum_{i=1}^{X}\sum_{j=1,j\neq i}^{X}\mathsf{E}[Y_i|X]\cdot p^2\right]\\&=p\mathsf{E}\left[\sum_{i=1}^{X}Y_i\right]+p^2\mathsf{E}\left[(X-1)\sum_{i=1}^{X}\mathsf{E}[Y_i|X]\right]\\&=p\alpha(t)+p^2(w(t)-\alpha(t)),\end{aligned} \tag{97}$$

where we use (3) and (4) in the last identity.

With (39), (94), (95), (96) and (97), it is not difficult to show that the expectation of (93) is

$$2\cdot\big(p^2w(t)+pw(t)+(1+4p-p^2)k(t) +(1+3p)\alpha(t)+p^2\tau(t)+1\big). \quad (98)$$

From (98), we obtain the following balance equation

$$(m_0+t+1)w(t+1)=(m_0+t)w(t) +2\big(p^2w(t)+pw(t)+(1+4p-p^2)k(t) +(1+3p)\alpha(t)+p^2\tau(t)+1\big). \quad (99)$$

This difference equation can be approximated by the following first-order linear differential equation

$$w'(t) = \frac{2}{m_0+t+1}\left(\left(p^2+p-\frac{1}{2}\right)w(t) +(1+4p-p^2)k(t)+(1+3p)\alpha(t) +p^2\tau(t)+1\right) \quad (100)$$

with initial condition

$$w(0)=(m_0-1)^3. \quad (101)$$

Applying the technique of integrating factors, the solution of (100) can be written as

$$w(t)=(m_0+t+1)^{2p^2+2p-1}\Big(c+ 2\int(m_0+t+1)^{-2p^2-2p}(1+(3p+1)\alpha(t)+ (1+4p-p^2)k(t)+p^2\tau(t))\,dt, \quad (102)$$

where c is a constant to be determined from the initial condition (101). Substituting $k(t)$ from (18), $\tau(t)$ from (27) and $\alpha(t)$ from (48) into (102), we obtain

$$w(t)=w_1(m_0+t+1)^{2p^2+2p-1}+ w_2(m_0+t+1)^{p^2+2p-1}+w_3(m_0+t+1)^{3p^2-1}+ w_4(m_0+t+1)^{2p-1}+w_5, \quad (103)$$

where constant w_1 is determined from the initial condition in (101). Constants w_2 through w_5 are given by

$$\begin{aligned} w_2 &= -\alpha_1(6p+2)p^{-2}\\ w_3 &= 2\tau_1 p(p-2)^{-1}\\ w_4 &= k_1(p^2-4p-1)p^{-2}- k_1(3p+1)(p^2-4p-2)p^{-4}-3k_1(2-3p)^{-1}\\ w_5 &= \frac{4p^2-12p-6}{(2p-1)(-2p^2-2p+1)}+ \frac{(6p+2)(2p^2-4p-6)}{(2p-1)(-p^2-2p+1)(-2p^2-2p+1)}- \frac{12p^3}{(2p-1)(1-3p^2)(-2p^2-2p+1)}. \end{aligned}$$

The solution in (103) is not valid for $p=0,1/2,2/3,1/\sqrt{3},\sqrt{2}-1$ and $(\sqrt{3}-1)/2$, which is the positive root of $2p^2+2p-1=0$. For $p=0$, substituting $k(t)$, $\tau(t)$ and $\alpha(t)$ from (18), (33), and (51) respectively into (102), we obtain

$$w(t)=(m_0+t+1)^{-1}\Big(c_{17}+18t+2k_1(\log(m_0+t+1))^2 +2(-6+k_1+c_6-6m_0)\log(m_0+t+1)\Big). \quad (104)$$

For $p=1/2$, substituting $k(t)$, $\tau(t)$ and $\alpha(t)$ from (21), (31), and (52) respectively into (102), we obtain

$$w(t)=1856+133c_1-\frac{2c_2}{3(m_0+t+1)^{1/4}} -20c_7(m_0+t+1)^{1/4}+c_{18}\sqrt{m_0+t+1} +266\log(m_0+t+1). \quad (105)$$

For $p=2/3$, substituting $k(t)$, $\tau(t)$ and $\alpha(t)$ from (18), (34), and (48) respectively into (102), we obtain

$$w(t)=\frac{-1578}{11}+\frac{4807k_1-88}{88}(m_0+t+1)^{1/3} -297\alpha_1(m_0+t+1)^{7/9}+c_{19}(m_0+t+1)^{11/9} -2k_1(m_0+t+1)^{1/3}\log(m_0+t+1). \quad (106)$$

For $p=1/\sqrt{3}$, substituting $k(t)$, $\tau(t)$ and $\alpha(t)$ from (18), (32), and (48) respectively into (102), we obtain

$$w(t)=(m_0+t+1)^{(-3+2\sqrt{3})/3}\Bigg(c_{20}+ \frac{(m_0+t+1)^{-2(1+\sqrt{3})/3}}{-38+21\sqrt{3}}\times\Bigg((492+242\sqrt{3}+2(-8+5\sqrt{3})c_3)(m_0+t+1)- (374-143\sqrt{3})k_1(m_0+t+1)^{2/\sqrt{3}}- (150-102\sqrt{3})\alpha_1(m_0+t+1)^{(1+2\sqrt{3})/3}+ +12(1-2\sqrt{3})(m_0+t+1)\log(m_0+t+1)\Bigg)\Bigg). \quad (107)$$

For $p=\sqrt{2}-1$, substituting $k(t)$, $\tau(t)$ and $\alpha(t)$ from (18), (27), and (53) respectively into (102), we obtain

$$w(t)=(m_0+t+1)^{3-2\sqrt{2}}\Bigg(c_{21}+2\times \Bigg(\frac{(-65+46\sqrt{2})\tau_1(m_0+t+1)^{5-4\sqrt{2}}}{-119+84\sqrt{2}}+ \frac{(-295+218\sqrt{2})k_1(m_0+t+1)^{-6+4\sqrt{2}}}{2(-65+46\sqrt{2})}+ \frac{(68-3\sqrt{2}+(812-574\sqrt{2})c_8(m_0+t+1)^{-3+2\sqrt{2}}}{2(379-268\sqrt{2})}+ \frac{56(-38+27\sqrt{2})(m_0+t+1)^{-3+2\sqrt{2}}\log(m_0+t+1)}{2(379-268\sqrt{2})}\Bigg)\Bigg). \quad (108)$$

When p is $(\sqrt{3}-1)/2$, it follows that $2p^2+2p-1=0$ and the coefficient of $w(t)$ on the right side of (100) vanishes. In

this case, the differential equation can be solved by integration. That is,

$$w(t) = c_{22} + \frac{(m_0+t+1)^{-(4+3\sqrt{3})/2}}{9812-5665\sqrt{3}} \times \Big(-22(-362+209\sqrt{3})\tau_1(m_0+t+1)^4 + 22(1481-855\sqrt{3})\alpha_1(m_0+t+1)^{1+2\sqrt{3}} + (74798-43303\sqrt{3})k_1(m_0+t+1)^{5\sqrt{3}/2} - 176(9-5\sqrt{3})(m_0+t+1)^{(4+3\sqrt{3})/2}\log(m_0+t+1)\Big). \quad (109)$$

Constants c_{17} through c_{22} in Eqs. (104) through (109) are all determined by the initial condition in (101).

V. Asymptotic Analysis

In this section we derive an asymptotic analysis for the clustering coefficient and the degree correlation coefficient. We begin with the clustering coefficient in (14). We approximate the functions $k(t)$, $\tau(t)$ and $\alpha(t)$ by keeping the dominant terms. Specifically, from (18), (27) and (48), we have

$$k(t) \approx \begin{cases} k_1 t^{2p-1} & p > 1/2 \\ k_2 & p < 1/2, \end{cases} \quad (110)$$

$$\tau(t) \approx \begin{cases} \tau_1 t^{3p^2-1} & p > 2/3 \\ \tau_2 t^{2p-1} & 1/2 < p < 2/3 \\ \tau_3 & p < 1/2, \end{cases} \quad (111)$$

$$\alpha(t) \approx \begin{cases} \alpha_1 t^{p^2+2p-1} & p > \sqrt{2}-1 \\ \alpha_3 & p < \sqrt{2}-1. \end{cases} \quad (112)$$

In the above approximations we have approximated m_0+t+1 by t, since we consider very large t. We also note that although $\tau(t)$ and $\alpha(t)$ have different expressions in (32), (33) and (51) when p is 0 or $1/\sqrt{3}$, the approximations in (111) and (112) remain valid for these special values of p. We substitute approximations (110), (111) and (112) into (14) and approximate C by C_a, where

$$C_a \approx \begin{cases} (2\tau_1/\alpha_1)t^{3p^2-1-(p^2+2p-1)} & p > 2/3 \\ (2\tau_2/\alpha_1)t^{(2p-1)-(p^2+2p-1)} & 1/2 < p < 2/3 \\ (2\tau_3/\alpha_1)t^{-(p^2+2p-1)} & \sqrt{2}-1 < p < 1/2 \\ 2\tau_3/(\alpha_3-k_2) & p < \sqrt{2}-1. \end{cases} \quad (113)$$

The clustering coefficient at the special points of p can be approximated using (21), (31), (34), (52), (53). We have

$$C_a \approx \begin{cases} (4k_1/\alpha_1)t^{-4/9}\log(t) & p = 2/3 \\ \frac{12\log(t)}{c_7 t^{1/4}} & p = 1/2 \\ \frac{2\tau_3}{4(5+4\sqrt{2})\log(t)} & p = \sqrt{2}-1. \end{cases} \quad (114)$$

From (113) and (114), it follows that as t approaches infinity, the limiting clustering coefficient is

$$C \approx \begin{cases} 1 & p = 1 \\ \frac{3p(p^2+2p-1)}{(2p+1)(3p^2-1)} & 0 < p < \sqrt{2}-1 \\ 0 & \text{otherwise.} \end{cases}$$

Numerical calculations show that the limiting clustering coefficient is between zero and one. For p between 0 and $\sqrt{2}-1$, the maximum is 0.2743 and occurs at $p = 0.2181$.

Next we derive an approximation for the degree correlation coefficient in (58). From (110) and (112) we approximate $\alpha(t)/k(t)$, *i.e.*

$$\frac{\alpha(t)}{k(t)} \approx \begin{cases} \frac{\alpha_1}{k_1}t^{p^2} & p > 1/2 \\ \frac{\alpha_1}{k_2}t^{p^2+2p-1} & \sqrt{2}-1 < p < 1/2 \\ \frac{\alpha_2}{k_2} & p < \sqrt{2}-1. \end{cases} \quad (115)$$

From (72) and (103) we approximate $\beta(t)$ and $w(t)$ as follows

$$\frac{\beta(t)}{k(t)} \approx \begin{cases} \beta_1 t^{p^3+p}/k_1 & p > 1/2 \\ \beta_1 t^{(p^3+3p-1)}/k_2 & r < p < 1/2 \\ \beta_4/k_2 & p < r, \end{cases} \quad (116)$$

where r is defined in (76), and

$$\frac{w(t)}{k(t)} \approx \begin{cases} w_1 t^{2p^2}/k_1 & p > 1/2 \\ w_1 t^{(2p^2+2p-1)}/k_2 & (\sqrt{3}-1)/2 < p < 1/2 \\ w_5/k_2 & p < (\sqrt{3}-1)/2. \end{cases} \quad (117)$$

We note that although $\beta(t)$ and $w(t)$ have different expressions when p is 0 or $1/\sqrt{3}$, the approximations above remain valid for these two special values of p. Substituting (115), (116) and (117) into (58), we approximate ρ by ρ_a, where

$$\rho_a \approx \begin{cases} \frac{(w_1/k_1)t^{2p^2}-(\alpha_1/k_1)^2 t^{2p^2}}{(\beta_1/k_1)t^{p^3+p}-(\alpha_1/k_1)^2 t^{2p^2}} & p > 1/2 \\ \frac{(w_1/k_2)t^{2p^2+2p-1}-(\alpha_1/k_2)^2 t^{2(p^2+2p-1)}}{(\beta_1/k_2)t^{p^3+3p-1}-(\alpha_1/k_2)^2 t^{2(p^2+2p-1)}} & \sqrt{2}-1 < p < 1/2 \\ \frac{(w_1/k_2)t^{2p^2+2p-1}-(\alpha_3/k_2)^2}{(\beta_1/k_2)t^{p^3+3p-1}-(\alpha_3/k_2)^2} & (\sqrt{3}-1)/2 < p < \sqrt{2}-1 \\ \frac{(w_5/k_2)-(\alpha_3/k_2)^2}{\beta_1 t^{p^3+3p-1}/k_2-(\alpha_3/k_2)^2} & r < p < (\sqrt{3}-1)/2 \\ \frac{(w_5/k_2)-(\alpha_3/k_2)^2}{(\beta_4/k_2)-(\alpha_3/k_2)^2} & p < r. \end{cases} \quad (118)$$

At other special points of p, the degree correlation is approximately equal to

$$\rho_a \approx \begin{cases} \frac{c_{19}-\alpha_1^2/k_1}{\beta_1}t^{-2/27} & p = 2/3 \\ (c_{11}/c_{18})t^{-1/8} & p = 1/2 \\ (c_{21}/c_{10})t^{14-10\sqrt{2}} & p = \sqrt{2}-1. \end{cases} \quad (119)$$

It follows from (118) and (119) that

$$\rho_a \approx \begin{cases} \frac{(w_1/k_1)-(\alpha_1/k_1)^2}{-(\alpha_1/k_1)^2} & p > 1/2 \\ 0 & r < p < 1/2 \\ \frac{(w_5/k_2)-(\alpha_3/k_2)^2}{\beta_4/k_2-(\alpha_3/k_2)^2} & p < r \end{cases} \quad (120)$$

as t goes to infinity.

VI. Numerical and Simulation Results

In this section we present numerical and simulation results to verify our analysis. First, we simulate the duplication model one hundred times and calculate the mean degree and the expected number of triangles per vertex. Ninety-five percent

confidence intervals were collected based on the repeated simulation of one hundred times. We choose $m_0 = 4$.

We first compare the numerical evaluation of the clustering coefficient with the simulation result. The results with $p = 0.3$ and $p = 0.6$ are shown in Figure 3. As one can see, the analytical result is very accurate. This accuracy holds for other values of p as well.

We compare the numerical evaluation of $\beta(t)$ and $w(t)$ shown in (72) and (103) with simulation. The results are shown in Figure 4 and Figure 5, respectively. We study the accuracy of the approximation of $\mathsf{E}[\tilde{X}^\nu]$ in (56) for $\nu = 1, 2$. The result is shown in Figure 6. These figures show that the numerical results and approximation are very close to the simulation results. Finally, we show the Pearson degree correlation coefficients for $p = 0.3$ and 0.6 in Figure 7. There is an obvious discrepancy at early times. However, as time goes on, the discrepancy diminishes. As argued in Section II, one possible explanation for this is that the two identities in Lemma 3 would (approximately) hold for the random network from Chung's duplication model as t goes to infinity.

Finally we study the accuracy of approximating clustering coefficients and degree correlation coefficients by their asymptotic formulae. For a specific value p and t, we define the relative approximation error for the clustering coefficient to be

$$\left| \frac{C - C_a}{C} \right|,$$

where C is the analytic solution using differential equations and C_a is obtained from (113) and (114). Our numerical experience indicates that for some values of p, the most dominant term in C can be quite close to the second dominant term. It takes very large t for the asymptotic approximation to be close to the analytic result. For these large values of t, the computation time and storage requirements can be enormous and impractical. Our study has established that analytical results are very close to simulation results when t is large. Thus, we have not compared asymptotic results with simulation results. Instead, we compare the asymptotic results with analytical results. The result is shown in Figure 8 and Figure 9. As shown in these figures, as t becomes large, the asymptotic approximation improves. However, for p in the neighborhood of one of its special values, the accuracy improves very slowly. We numerically compute the first time that the relative error of the clustering coefficient and the degree correlation coefficient reaches ten percents, five percents and one percent. The result is shown in Figure 10. From this figure, we see that the time to reach small relative errors can be very large, especially when p is close to one of its special values. The accuracy of the approximation depends not only on the most dominant terms, but also on the second dominant terms. We illustrate this point using the clustering coefficient as an example. Consider the approximation of $\tau(t)$ in the neighborhood of $p = 2/3$. From (111) the first two terms on the right side are the the most dominant term and the second dominant term, respectively, of $\tau(t)$ depending on whether p is greater than or less than $2/3$. From (28) and (29), $p = 2/3$ is a singlar point of the coefficients of the two terms. Thus, the most dominant term and the second dominant term both have a large coefficient in the neighborhood of $p = 2/3$. In addition, the exponents of the dominant term and the second dominant term are equal at $p = 2/3$. These two factors imply that it takes a long time to compute $\tau(t)$ accurately using only the most dominant term when p is near $2/3$.

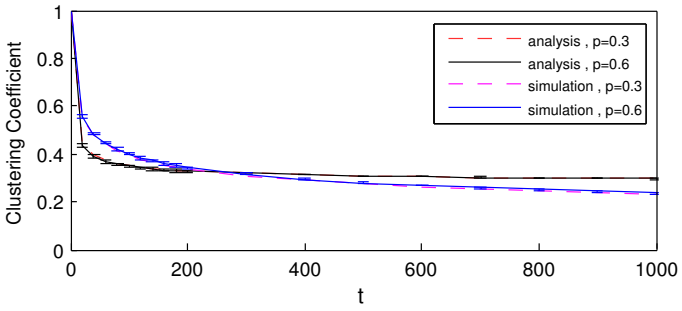

Fig. 3. Clustering coefficients as a function of time with $p = 0.3$ and 0.6.

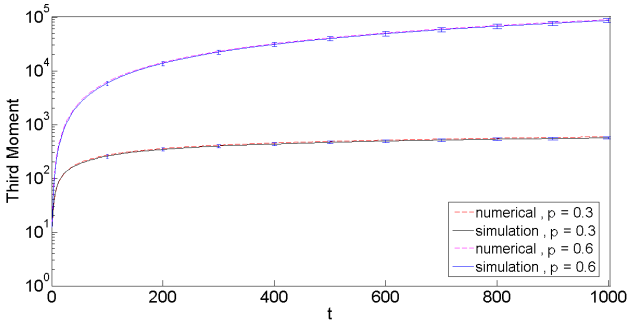

Fig. 4. $\beta(t)$ as a function of time with $p = 0.3$ and 0.6.

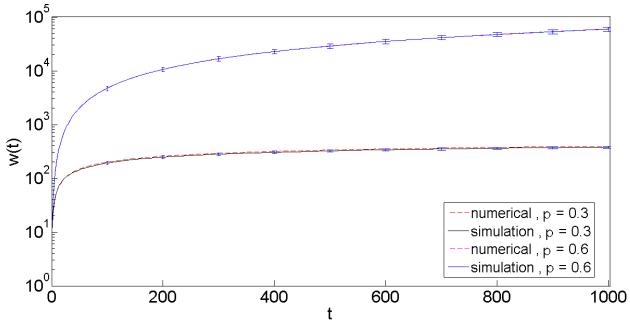

Fig. 5. $w(t)$ with $p = 0.3$ and 0.6.

VII. Conclusions

Duplication models have been used successfully to study biological networks and social networks. In this paper, we first derived some identities that relate the degree of a randomly selected vertex and those of its neighbors in a simple undirected network. These identities were used to derived closed-form expressions for the clustering coefficient and Pearson degree correlation coefficient of the duplication model. These identities can potentially be applied to other problems and are of independent interest. The two identities in (5) and (6) that were used to the Pearson degree correlation coefficient involve

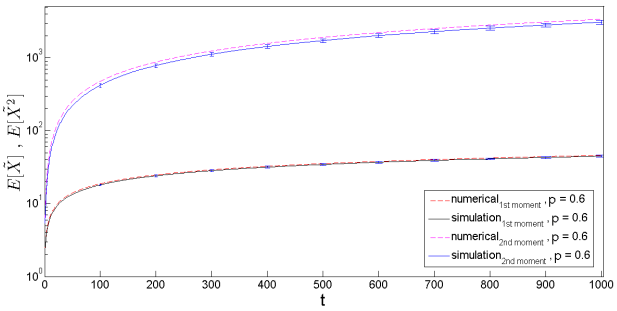

Fig. 6. $\mathsf{E}[\tilde{X}]$ and $\mathsf{E}[\tilde{X}^2]$. $p = 0.6$.

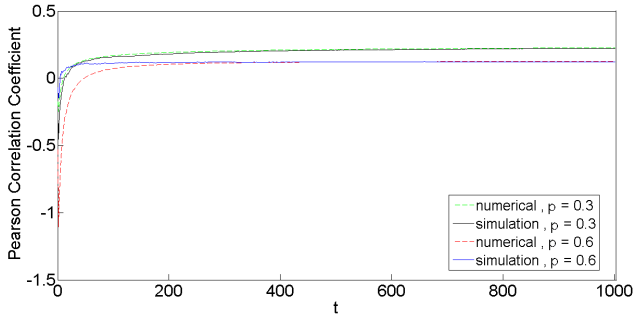

Fig. 7. Pearson degree correlation coefficients as a function of time with $p = 0.3$ and 0.6.

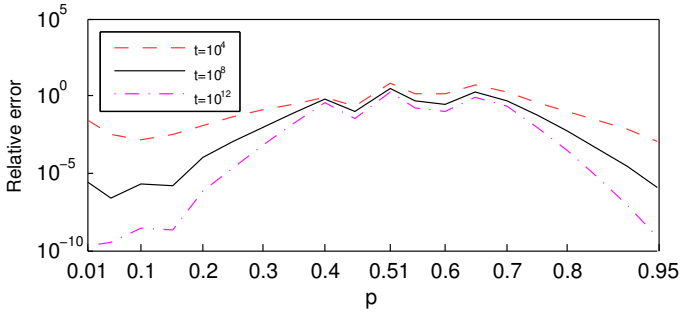

Fig. 8. Asymptotic analysis of the clustering coefficient.

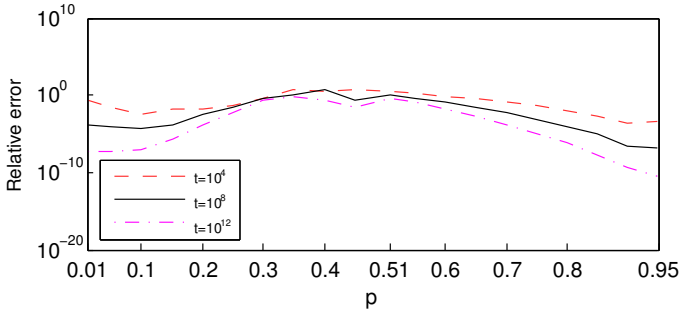

Fig. 9. Asymptotic analysis of the Pearson degree correlation coefficient.

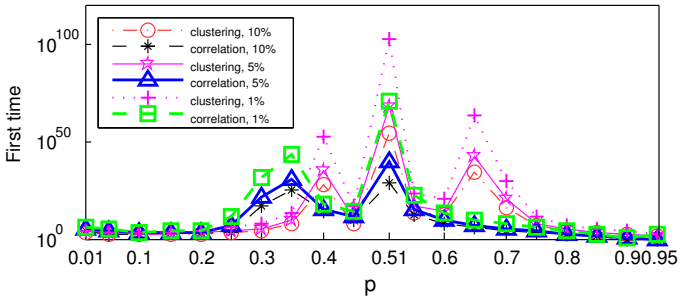

Fig. 10. First time that the relative error of the clustering coefficient and the degree correlation coefficient reaches ten percents, five percents and one percent.

with expected ratios of two random quantities. We approximate the expected ratio by the ratio of expected numerator and expected denominator. This approximation works well when the expected denominator, *i.e.* the mean degree, converges as the time becomes large. The closed-form solutions for the clustering coefficient and the Pearson degree correlation coefficient are quite complicated and lengthy. We proposed an asymptotical approximation by keeping only a dominant term and ignoring the rest terms.

Through numerical and simulation study, we showed that the numerical calculation of the clustering coefficient and the corresponding simulation agree extremely well. For the Pearson degree correlation coefficient there is an obvious discrepancy at early times between the simulation results and the numerical results. However, as time goes on, the discrepancy diminishes. We numerically studied the accuracy of the asymptotic approximation. Our numerical result indicated that this approximation works reasonably well for only moderate value of t. However, when the selection probability p is near one of its special values, it can take a very large t to reach a small relative approximation error.

Acknowledgement

This research was supported in part by the Ministry of Science and Technology, Taiwan, R.O.C., under Contract NSC-99-2221-E-007-079-MY3.

References

[1] A. Bhan, D. J. Galas, and T. G. Deway. A duplication growth model of gene expression networks. *Bioinformatics*, 18(11):1486–1493, 2002.

[2] S. Boccaletti, D.-U. Hwang, and V. Latora. Growing hierarchical scale-free networks by means of nonhierarchical processes. *International Journal of Bifurcation and Chaos*, 17(7):2447–2452, 2007.

[3] F. Chung and L. Lu. Complex graphs and networks. In *Regional Conference Series in Mathematics*, number 107. American Mathematical Society, 2004.

[4] F. Chung, L. Lu, and T. G. Dewey. Duplication models for biological networks. *Journal of Computational Biology*, 10(5):677–687, 2003.

[5] R. Friedman and A. Hughes. Gene duplications and the structure of eukaryotic genomes. *Genome Res.*, 11:373–381, 2001.

[6] Z. Gu and A. Cavalcanti and F.-C. Chen and P. Bouman and w.-H. Li. Extent of gene duplication in the genomes of drosophila, nematode, and yeast. *Mol. Biol. Evol.*, 19:256-262, 2002.

[7] I. Ispolatov, P. L. Krapivsky, I. Mazo, and A. Yuryev. Cliques and duplication-divergence network growth. *New Journal of Physics*, june 2005.

[8] I. Ispolatov, P. L. Krapivsky, and A. Yuryev. Duplication-divergence model of protein interaction network. *Phys. Rev. E*, 71:061911, Jun 2005.

[9] Duan-Shin Lee, Cheng-Shang Chang, Wen-Gui Ye, and Min-Chien Cheng. Analysis of clustering coefficients of online social networks by duplication models. In *IEEE ICC 2014*, Sydney, Australia.
[10] M. Newman. *Networks: An Introduction*. Oxford, 2010.
[11] S. Ohno. *Evolution by Gene Duplication*. Springer, New York, 1970.
[12] R. V. Solé, R. Pastor-Satorras, E. Smith, and T. B. Kepler. A model of large-scale proteome evolution. *Advances in Complex Systems*, July 2002.
[13] L. Stubbs, Genome comparison techniques, Genomic Technologies: Present and Future, D. Galas and S. McCormack, Caister Academic Press, 2002.
[14] Wolfram Research, Inc., Mathematica, Version 8.0, Champaign, IL (2001).
[15] D. Zhao, Z.-R. Liu, and J.-Z. Wang. Duplication: a mechanism producing disassortative mixing networks in biology. *Chin. Phys. Lett.*, 24(10):2766–2768, 2007.