Modeling the Evolution of Degree Correlation in Scale-Free Topology Generators

Xiaoming Wang
Texas A&M University
College Station, TX 77843
xmwang@cs.tamu.edu

Xiliang Liu
The City University of New York
New York, NY 10016
liuxiliang@gmail.com

Dmitri Loguinov*
Texas A&M University
College Station, TX77843
dmitri@cs.tamu.edu

***Abstract*—In this paper, we examine the asymptotic behavior of degree correlation (i.e., the joint degree distribution of adjacent nodes) in several scale-free topology generators GED [13], PLRG [1], GLP [10], BA [3], AB [2]. We present a unifying analytical framework that allows tractable analysis of degree correlation in all studied models and derive asymptotic formulas of two degree correlation metrics – assortativity and clustering. Our results indicate that all studied generators become uncorrelated as graph size increases, which is inconsistent with time-invariance of these metrics in real networks such as the Internet [36], [48], [50]. Since the class of degree-based generators is incapable of reproducing evolving characteristics of the Internet, we study three other models that evolve graphs using different rules than preference of degree (e.g., based on random walks [50], optimization [17], and geometry [23]) and show using simulations that these models are much more viable alternatives for replicating the complex structure of Internet-like graphs.**

I. Introduction

Recent research suggests that many graphs found in the real world (such as social relationships, scientific collaborations, Internet autonomous system connectivity, web-page linkage, telephone call logs, various molecular structures, etc.) exhibit drastically different characteristics from those of classical Erdös-Rényi random graphs [16]. Besides the well-known heavy-tailed distribution of node degree [14], [18], real-world graphs demonstrate a strong correlation among the degree of adjacent nodes (e.g., large-degree nodes are more likely to be paired with small-degree nodes, or triangle formation is more likely between large-degree nodes). A significant research effort is currently under way to better understand the evolution of complex networks and design generators that can reproduce graph theoretic metrics found in these structures [1], [2], [3], [10], [12], [17], [22], [23], [27], [47], [50], [52].

Recent studies [28], [41] demonstrate that degree correlation (formally defined below) is sufficient for characterizing the structure of a random graph, where correlation among up to three neighboring nodes is enough to capture many commonly used graph properties. In fact, the authors of [28] propose a generator that randomly rewires links until it can match the desired degree correlation. The resulting graphs are then shown to replicate such global properties as coreness, spectrum, distance distribution, and betweenness of many existing topology generators. However, this approach is very computationally intensive and requires the knowledge of the joint degree distribution of neighbors in the target graph, which may be difficult to obtain in practice. Furthermore, the brute-force approach of [28] does not easily allow the graph to evolve and does not capture any qualitative characteristics of the system it creates.

To understand the fundamental properties of degree correlation in existing topology generators and how well they match those of existing networks, we propose a new analytical framework for modeling two- and three-node correlation in power-law random graphs with a specific focus on the *evolution* of the system. It is well-known that static properties of graphs (i.e., those observed at a fixed time t) can be captured with sufficient accuracy by many existing generators. However, a more interesting question [50] has recently emerged as to whether these generators can replicate the behavior of real graphs as time evolves. Specifically, for a fixed set of initial parameters, [50] shows that the trend of clustering coefficients of many evolving degree-based generators exhibits a significant discrepancy with that of the Internet as observed in [36], [48]. As clustering coefficients are a special case of three-node correlation, our proposed framework allows to understand what causes this discrepancy and whether it can be overcome within the studied class of topology generators.

The class of models we study in this work is called *degree-based* since link formation between individual nodes depends only on the current degree of potential neighbors, or equivalently, weights assigned to them. Many traditional generators such as GED [13], PLRG [1], GLP [10], BA [3], and AB [2] fall into this category. On the other hand, generators that implement neighbor selection based on additional information (besides node degree) constitute a completely different class of *link-based* models (e.g., WIT [50], HOT [17], and SWT [23]), which are not modeled in this paper, but briefly studied in simulations (see below).

A. Main Results

We first study GED as a foundation for understanding correlation in degree-based random graphs. We derive the asymptotic decay rate of the expected assortativity coefficient $r(G)$, which captures two-node correlation in the system, and the expected clustering coefficient $\gamma(G)$, which captures three-

*Supported by NSF grants CCR-0306246, ANI-0312461, CNS-0434940, CNS-0519442, and CNS-0720571.

node correlation, as graph size $n \to \infty$. Specifically, our analysis demonstrates that:

$$E[r(G)] = \begin{cases} \Theta\left(-n^{1-\alpha}\right) & 1 < \alpha < 2 \\ \Theta\left(-n^{-1}\log^2 n\right) & \alpha = 2 \\ 0 & \alpha > 2 \end{cases} \quad (1)$$

and

$$E[\gamma(G)] = \begin{cases} \Theta(n^{1-\alpha}\log n) & 1 < \alpha < 2 \\ \Theta(n^{-1}\log^2 n) & \alpha = 2 \\ \Theta(n^{-1}) & \alpha > 2 \end{cases}, \quad (2)$$

where $\alpha > 1$ is the shape parameter of the power-law degree distribution.[1] This analysis shows that GED is asymptotically *uncorrelated* for all values of $\alpha > 1$ and provides an explanation of the phenomena observed earlier in [50], where GED's clustering decayed to zero while that of the Internet remained constant.

Full derivation of (2) for $1 < \alpha < 2$ involves a great deal of tedious integration (see [49, Theorem 2]), which is unrealistic to perform for each studied method and sometimes even impossible depending on the type of function used to construct each link. To expand the GED result to a wider class of additional methods, we next propose a general framework for reducing degree-based generators to GED in order to obtain their asymptotic assortativity and clustering. We show that as long as the link-existence probability function π of a new method can be bounded by that of GED, its correlation asymptotically behaves the same as (1)-(2). We then show that this condition, which we call *asymptotic π-equivalence*, holds for PLRG, GLP, BA, and AB. We not only obtain completely novel results on the behavior of degree correlation in PLRG, GLP, and AB, but we also provide an analytical platform of π-equivalence that can be used to show similar results for any degree-based generator with a power-law degree distribution.

We finish the paper by discussing the implications of results obtained in this work. In particular, our results prove that all existing degree-based generators become uncorrelated as the graph grows in size. This is in stark contrast to networks observed in real life, in which both assortativity and clustering remain constant as the system evolves. This invariance of both the degree distribution and its correlation has an important impact on the design of future topology generators. Specifically, we conjecture that the entire class of degree-based generators is insufficient for capturing the evolving structure of small-world graphs such as the Internet and illustrate in simulations that link-based generators WIT [50], HOT [17], and SWT [23] are much more effective in keeping node correlation time-invariant. This suggests that the fundamental differences between degree-based and link-based models first observed in this work are indeed significant.

[1]Note that prior results of this nature [4], [7], [8], [14], [24], [46] are available only for the "simple" case of $\alpha \geq 2$, which is not applicable to many existing networks (e.g., the Internet AS graph has $\alpha \approx 1.2$).

II. Background

In this section, we overview a small subset of related work, introduce the notation commonly used in this field, and mention several well-known models that we study later in the paper.

Assume an undirected connected graph $G = (\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$. Suppose that the graph has $|\mathcal{V}| = n$ nodes, whose degrees are given by $d_1, ..., d_n$. For asymptotically large $n \to \infty$, we treat node degree d as a continuous random variable with CDF $F(x) = P(d < x)$ and replace all degree-based summations with corresponding integrals. Since we solely focus on scale-free graphs, the CDF function is assumed to be Pareto $F(x) = 1 - (\beta/x)^\alpha$, where $\beta > 0$ is the scale parameter and $\alpha > 1$ is the shape parameter. Denote by $i \leftrightarrow j$ the event of link (i, j) being present in $\mathcal{E}$ and by $E[d^k]$ the k-th moment of node degree in graph G.

A. Degree Correlation

The *assortativity coefficient* $r(G)$ characterizes two-node degree correlation and measures the extent to which nodes connect preferentially to other nodes with similar degrees [31]. The expectation of $r(G)$ is given by:

$$E[r(G)] = \frac{E^2[d]\sum_x \sum_y xy\omega(x,y) - E^2[d^2]}{E[d]E[d^3] - E^2[d^2]}, \quad (3)$$

where $\omega(x, y)$ is defined as the probability that two arbitrary connected nodes have degree x and y, respectively [15]:

$$\omega(x, y) = P(d_i = x, d_j = y | i \leftrightarrow j). \quad (4)$$

Graphs with positive values of $r(G)$ are so-called *assortative*, graphs with negative $r(G)$ are called *disassortative*, and graphs with $r(G) = 0$ are called *uncorrelated*.

The most widely-used metric for three-node degree correlation is the *clustering coefficient* $\gamma(G)$, which quantifies how likely the neighbors of a node are to be connected to each other. Suppose that a given node i is contained in T_i triangles. Recall from [51] that the clustering coefficient γ_i of node i (as long as degree d_i is at least two) is defined as the ratio of T_i to the maximum number of such triangles:

$$\gamma_i = \frac{T_i}{d_i(d_i - 1)/2}, \; d_i \geq 2, \quad (5)$$

and the clustering coefficient $\gamma(G)$ of the graph is then defined to be the average of γ_i over all nodes i with degree $d_i \geq 2$:

$$\gamma(G) = \frac{\sum_{i \in \mathcal{V} - \mathcal{V}^{(1)}} \gamma_i}{|\mathcal{V}| - |\mathcal{V}^{(1)}|}, \quad (6)$$

where $\mathcal{V}^{(1)}$ is the set of both degree-zero and degree-one nodes in G, i.e., $\mathcal{V}^{(1)} = \{j \in \mathcal{V} : d_j \leq 1\}$. Note that other definitions of the clustering coefficient exist in [8], [45], [53], but they are not studied here.

The combination of assortativity and clustering has been used in characterizing real networks [21], [34]. For power-law

networks with shape parameter $\alpha > 2$, the following results have been established [7], [14], [32]:

$$E[r(G)] = 0, \quad E[\gamma(G)] = \frac{(E[d^2] - E[d])^2}{nE^3[d]}. \tag{7}$$

However, little is known about the case of $1 < \alpha \leq 2$ commonly found in real networks (e.g., the Internet). The closest analysis to tackling this issue is given in [14], which obtains $E[\gamma(G)]$ as a complex summation that allows only numerical computation and provides no qualitative asymptotic behavior in the limit of large n. In this paper, we solve this problem and establish the trend of $E[\gamma(G)]$ for all $\alpha > 1$.

B. Graph Models

In this section, we briefly mention several graph models that produce power-law degrees. For a complete literature survey, we refer readers to [15], [32], [37].

1) Non-Evolving Models: In this category, we mention several generators that do not grow (evolve) the network over time. One of the simplest power-law graph construction models is called *Given Expected Degree* (GED) [13], a specialized algorithm of *fitness models* [7], [11], [19], [44]. In GED, a sequence of *weights* $\{w_i\}$ is first generated according to a Pareto distribution and then each edge (i, j) is created with independent probability:

$$p_{ij} = \min\left(\frac{w_i w_j}{D}, 1\right), \tag{8}$$

where $D = \sum_{k=1}^{n} w_k$.

A similar graph construction method called *Power-Law Random Graph* (PLRG) [1], or more generally *configuration models*, which have been extensively studied in [5], [26], [29], [30], [33], [35], and later extended in [28], [41]. PLRG replicates each node i exactly w_i times and then places random edges between the replicated nodes with equal probability.

Additional non-evolving generators include random geometric graphs [23] and rewired small-world (Watts) networks that exhibit heavy-tailed degree distributions [38], [51].

2) Evolving Models: Next, we review several models that grow graph sizes over time. Since Barabási's revival of this direction in 1999, evolution models have taken the center stage of modern topology modeling; however, the earliest evolving model can be traced back to Simon's work in 1955 [9], [42], [43]. Several classes of generators can be described under the umbrella of *Generalized Linear Preference* (GLP) [10], in which the system grows by adding new links or joining new nodes at each time step. The linking decision is based on the the preferential-attachment function $p_i(t)$, which is the probability of node i to be chosen at time t:

$$p_i(t) = \frac{d_i(t) - \lambda}{\sum_{k=1}^{n(t)}(d_k(t) - \lambda)}, \tag{9}$$

where shift-parameter $\lambda \in [-\infty, 1]$ and the degree distribution is Pareto with shape $\alpha = 2 - \lambda \in [1, \infty)$. Similar attachment functions are used in BA [3] and AB [2] with certain changes to λ and the algorithm for creating new links. Variants of (9) are also used in non-linear preferential-attachment [25], [53] and multi-scaling fitness models [6].

III. Analysis of GED

Recall that GED [13] assigns random weights w_i drawn from the Pareto distribution and then creates each link with probability p_{ij} in (8). Next, we explore the asymptotic behavior of GED's degree correlation and answer the question of whether it can be maintained for large n. We then demonstrate that PLRG, BA, AB, GLP, and potentially many other graphs can be modeled under the same analytical umbrella.

A. Link Formation

We first formalize the relationship between weights and edge existence in G. Define $\pi(x, y)$ to be the probability of two nodes being connected given their weights x and y:

$$\pi(x, y) = P(i \leftrightarrow j | w_i = x, w_j = y). \tag{10}$$

For GED, this function is simply:

$$\pi(x, y) = \min\Big(\frac{xy}{D}, 1\Big), \tag{11}$$

where $D = \sum_{k=1}^{n} w_k$ is the total degree weight in the graph.[2]

Next, notice that the distribution of D is given by an n-fold convolution of Pareto distributions, which makes the analysis of degree correlation intractable. Therefore, most of the derivations below approximate D with its expectation, i.e., $D = nE[w_i]$.

B. Two-Node Correlation – $E[r(G)]$

It has been shown in [7] that $\omega(x, y)$ in (3) can be described in terms of $\pi(x, y)$ as:

$$\omega(x, y) = \frac{nf(x)f(y)\pi(x, y)}{E[d]}, \tag{12}$$

where $f(x)$ is the PDF of weights. Substituting (12) into (3) and replacing the summations with integrals, we can express $E[r(G)]$ in terms of $\pi(x, y)$:

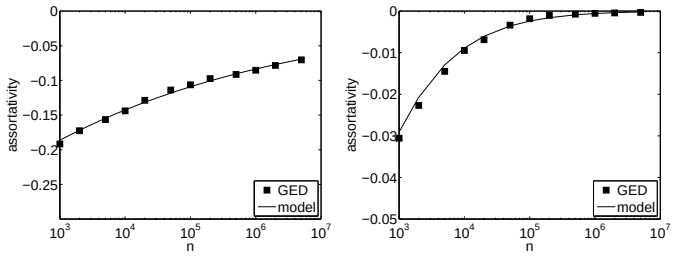
$$E[r(G)] = \frac{E[d]\rho - E^2[d^2]}{E[d]E[d^3] - E^2[d^2]}, \tag{13}$$

where ρ is given by:

$$\rho = n \iint xy\pi(x, y)f(x)f(y)dxdy. \tag{14}$$

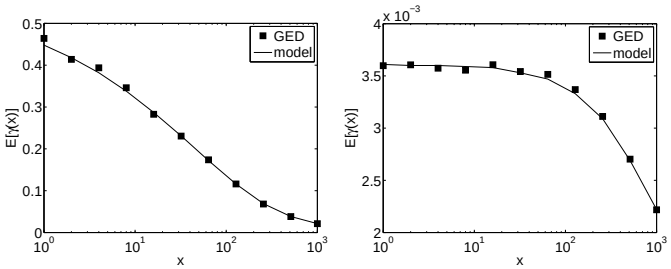
Note that for finite n, all node degree is no larger than $n-1$ and higher moments of d in (13) are finite. This leads to the next result that expands the integral in ρ and derives the asymptotic trend of $E[d^2]$ and $E[d^3]$ for Pareto distributions with $\alpha \leq 2$. We omit all proofs from the paper due to the page limit and refer interested readers to [49].

[2]The original model described in [13] does not have the min function as in (11), but instead requires that $w_i w_j < D$. Since this constraint does not hold in the Internet, we replace the inequality with the min function. A recent extension of the original model has been proposed in [7] to construct correlated graphs, however, their method is different from (11) and thus orthogonal to our discussion.



(a) $\alpha = 1.2$ (b) $\alpha = 2$

Fig. 1. Model (15) and $E[r(G)]$ in GED simulations. All cases use $10,000$ iterations.



(a) $\alpha = 1.2$ (b) $\alpha = 3$

Fig. 2. Verification of (20) in GED simulations with $1,000$ nodes and $10,000$ iterations).

Theorem 1: The expected assortativity coefficient of GED graphs is asymptotically:

$$E[r(G)] = \begin{cases} \Theta\left(-n^{1-\alpha}\right) & 1 < \alpha < 2 \\ \Theta\left(-n^{-1}\log^2 n\right) & \alpha = 2 \\ 0 & \alpha > 2 \end{cases}, \qquad (15)$$

where α is the shape parameter of the power-law weight distribution.

To verify the model, we constructed $10,000$ GED graphs with shape parameters $\alpha = 1.2$ and $\alpha = 2$ and extracted the corresponding average assortativity coefficients. In Fig. 1, we plot the curve of $E[r(G)]$ and the corresponding model (15), where we obtain the constant by fitting the model to the actual value for the smallest n used in the simulations. As the figure shows, the model matches simulations well. The result (15) shows that for finite n, GED graphs are expected to be disassortative, but as $n \to \infty$, two-node degree correlation disappears *regardless of shape* α. It is also worth noting that for different α, the assortativity coefficient decays to zero at different rates as specified in (15) – the heavier the tail, the slower the decay.

C. Three-Node Correlation – $E[\gamma(G)]$

In what follows, we reduce the problem of deriving $E[\gamma(G)]$ to finding the average clustering coefficient of nodes with a given weight. The first lemma indicates that the expected clustering of a single node i (averaged over all possible weight assignments in the graph) is sufficient for establishing $E[\gamma(G)]$.

Lemma 1: In any graph G, the expected clustering $E[\gamma(G)]$ is:

$$E[\gamma(G)] = E[\gamma_i | d_i \geq 2], \qquad (16)$$

where i is the index of any node in graph G.

Next, expanding $E[\gamma_i | d_i \geq 2]$, we yield a more convenient expression for $E[\gamma(G)]$ in the next lemma.

Lemma 2: The expected clustering of a GED graph is given by:

$$E[\gamma(G)] = \int E[\gamma(x)] f(x|d_i \geq 2) dx, \qquad (17)$$

where $E[\gamma(x)] = E[\gamma_i | w_i = x, d_i \geq 2]$ is the expected clustering of nodes with weight x and $f(x|d_i \geq 2)$ is the conditional PDF of weight w_i assigned to node i given that its degree $d_i \geq 2$.

To solve (17), we need to derive $f(x|d_i \geq 2)$. From Bayes' formula $P(A|B) = P(B|A)P(A)/P(B)$, we get:

$$f(x|d_i \geq 2) = \frac{P(d_i \geq 2 | w_i = x) f(x)}{P(d_i \geq 2)}, \qquad (18)$$

which can be easily computed knowing that the distribution of d_i is Poisson with mean $w_i = x$ [7]:

$$P(d_i = k | w_i = x) = \frac{x^k e^{-x}}{k!}. \qquad (19)$$

The last missing piece toward solving (17) is to develop a formula for $E[\gamma(x)]$.

Recall that $E[\gamma(x)] = E[\gamma_i | w_i = x, d_i \geq 2]$ is the conditional expectation of γ_i for a node with weight $w_i = x$ and degree at least two. Expanding this metric by additionally conditioning on weights of neighboring nodes j and k gives the following lemma.

Lemma 3: The expected clustering of a node with weight x is given by:

$$E[\gamma(x)] = \frac{\phi(x)}{\psi^2(x)}, \qquad (20)$$

where

$$\phi(x) = \iint \pi(u,v)\pi(x,u)\pi(x,v) f(u) f(v) du dv,$$
$$\psi(x) = \int \pi(x,u) f(u) du, \qquad (21)$$

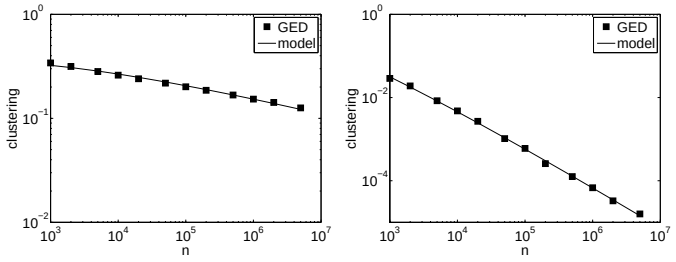
and $f(u)$ is the PDF of weights.

To verify (20), we created $10,000$ GED graphs with $n = 1,000$ in which one node always had weight $w_i = x$ and examined its average clustering coefficient $E[\gamma_i | w_i = x, d_i \geq 2]$. Simulation results are compared to (20) in Fig. 2, which shows that model (20) matches simulations very accurately.

Now, we are ready to derive an asymptotic formula for $E[\gamma(G)]$. Combining the results in (18)-(20), we can expand the integral in (17).

Theorem 2: With a power-law distributed weight sequence, the expected GED clustering is asymptotically:

$$E[\gamma(G)] = \begin{cases} \Theta(n^{1-\alpha}\log n) & 1 < \alpha < 2 \\ \Theta(n^{-1}\log^2 n) & \alpha = 2 \\ \Theta(n^{-1}) & \alpha > 2 \end{cases}, \qquad (22)$$



(a) $\alpha = 1.2$ (b) $\alpha = 2$

Fig. 3. Model (22) and $E[\gamma(G)]$ in GED simulations. All cases use $10,000$ iterations.

where α is the shape parameter of the power-law distribution.

We verified the model by constructing $10,000$ GED graphs with shape parameters $\alpha = 1.2$ as well as $\alpha = 2$ and extracting the corresponding average clustering coefficients, which is plotted in Fig. 3 along with the corresponding model (22). The figures shows that the model provides an accurate prediction of the asymptotic trend in clustering. Additional simulations show that (22) is accurate for different shape parameters α.

The result in (22) not only proves that GED's three-node correlation decays to zero for all values of $\alpha > 1$, but also provides the exact asymptotic decay rates.

IV. Generic Framework of Correlation Analysis

Notice that the results in (13), (17) do not specify $\pi(x, y)$ and thus can be used to derive the corresponding metric for any degree-based model with a known function $\pi(x, y)$. We next develop a generic framework that allows us to reuse the results derived from (13), (17) in application to PLRG, BA, AB, and GLP.

Consider a graph algorithm A that constructs random graphs G_A. Denote by $\pi_A(x, y)$ the probability of having a link between any pair of nodes with weights x, y in graph G_A, where weights may be explicitly pre-assigned as in GED or may represent the final degree of each node in the constructed graph. We start with the following definition.

Definition 1: Consider two graph algorithms A and B that operate with the same weight distribution $F(x)$. Define A to be *asymptotically π-equivalent* to B if the following holds for all $x \geq 0, y \geq 0$:

$$c_L \pi_B(x, y) \leq \pi_A(x, y) \leq c_U \pi_B(x, y) \tag{23}$$

where $0 < c_L < 1$ and $c_U \geq 1$ are some constants independent of x, y, or graph size n.

It is easy to verify that π-equivalence is both symmetric and transitive, which we formulate in the next lemma without proof.

Lemma 4: If A is π-equivalent to B, then B is π-equivalent to A. If A is π-equivalent to B and B is π-equivalent to C, then A is π-equivalent to C.

The next result follows after straightforward expansion of upper/lower bounds on $\pi(x, y)$ in formulas derived earlier and shows that asymptotic π-equivalence leads to easy ways of bounding the decay trend of two- and three-node correlation in any graph algorithm.

Theorem 3: If algorithm A is π-equivalent to B, then

$$\lim_{n\to\infty} \frac{E[r(G_A)]}{E[r(G_B)]} = c_r, \quad \lim_{n\to\infty} \frac{E[\gamma(G_A)]}{E[\gamma(G_B)]} = c_\gamma, \tag{24}$$

where constants c_r, c_γ are bounded by:

$$c_L \leq c_r \leq c_U, \quad \frac{c_L^3}{c_U} \leq c_\gamma \leq \frac{c_U^3}{c_L} \tag{25}$$

With the result in Theorem 3, the rest of the paper derives bounds on $\pi(x, y)$ in PLRG, GLP, BA, and AB. Specifically, we show that all of these methods can be reduced to GED through π-equivalence and thus exhibit the same asymptotic decay rate in $E[r(G)]$ and $E[\gamma(G)]$ in GED.

V. PLRG

In what follows, we first give a formula for the link-existence function $\pi(x, y)$ of PLRG and then study PLRG under the framework developed in the previous section.

Recall that PLRG first pre-assigns a random weight w_i drawn from a power-law distribution to each node i, generates w_i virtual copies of each node i, and then uniformly matches these virtual nodes to establish the actual links [1]. To make sure that the resulting degree d_i does not exceed the pre-assigned weights w_i, virtual nodes paired up during the process are immediately removed from the system.

To understand the statement of the following lemmas, define M_{ij} and L_{ij} to be random variables that specify the number of edges between nodes i and j *with* and *without* counting duplicate links, respectively, at the end of the graph formation process. It follows that:

$$L_{ij} = \begin{cases} 1 & M_{ij} \geq 1 \\ 0 & M_{ij} = 0 \end{cases}. \tag{26}$$

Let the node matching process in PLRG start at time 1 and finish at time D (recall that D is twice the number of edges), where the node chosen at time $2k - 1$ is paired up with the node chosen at time $2k$. Denote by $p_{ij}(t)$ the probability of forming a new edge between nodes i and j at time t. We first show in the next lemma that $p_{ij}(t)$ in PLRG does not depend on time t and then apply this result to deriving a more interesting expression for $E[L_{ij}] = P(i \leftrightarrow j) = \pi(w_i, w_j)$.

Lemma 5: The probability $p_{ij}(t)$ of forming a new link between i and j is independent of t:

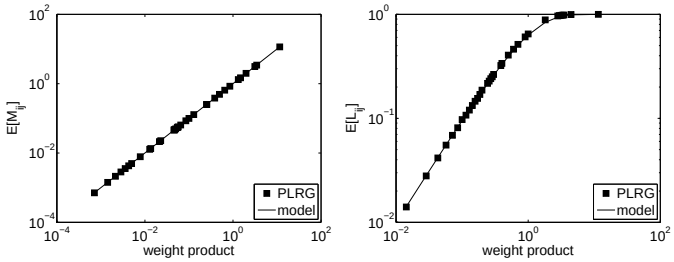
$$p_{ij}(t) = \frac{2w_i w_j}{D^2}, \tag{27}$$

where $D = \sum_{k=1}^{n} w_k$.

The result in (27) specifies the rate of accumulating links between i and j and leads to the following result stating the expectations of M_{ij} and L_{ij}.

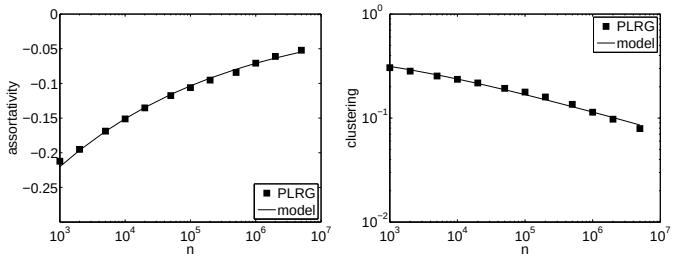
Theorem 4: The expected number of links between i and j is given by:

$$E[M_{ij}] = \frac{w_i w_j}{D}, \qquad E[L_{ij}] = 1 - e^{-w_i w_j / D}. \tag{28}$$



(a) $E[M_{ij}]$ (b) $E[L_{ij}]$

Fig. 4. Simulations of PLRG ($\alpha = 1.2$). All cases use $30,000$ iterations.



(a) assortativity (b) clustering

Fig. 5. Simulations of PLRG ($\alpha = 1.2$). All cases use $10,000$ iterations.

To verify the correctness of (28), we constructed 30,000 graphs PLRG and computed the average M_{ij} and L_{ij} of all pairs of nodes that have the same weight products. Fig. 4 plots $E[M_{ij}]$ and $E[L_{ij}]$ against the normalized weight product $w_i w_j/D$, indicating that model (28) follows the actual values accurately.

Note that the distribution of L_{ij} specifies the link-existence probability function $\pi(x, y)$, which is formulated in the next corollary.

Corollary 1: For PLRG, the link-existence probability function $\pi(x, y)$ is given by:

$$\pi(x, y) = 1 - e^{-xy/D}. \tag{29}$$

Note that for $xy/D \approx 0$, PLRG is identical to GED. However, as xy increases, the two algorithms exhibit diverging link formation probabilities and thus produce graphs with numerically different assortativity and clustering.

With the result in (29), we next show that PLRG's correlation decays to zero at the same rate as that in GED.

Theorem 5: PLRG is π-equivalent to GED.

Corollary 2: As graph size $n \to \infty$, two- and three-node correlation in PLRG decays to zero according to (15) and (22).

Fig. 5 plots the average degree correlation coefficients obtained from simulations of $10,000$ PLRG graphs with $\alpha = 1.2$ and indicates that the asymptotic models (15) and (22) match the actual very well for PLRG. Numerous additional simulations (omitted for brevity) have been conducted to verify (15) and (22) for other values of α in PLRG.

VI. GLP

In the category of evolving models, we start with the most general case of GLP and later reduce the obtained result to BA and AB. Recall from [10] that at each time step t, with probability p GLP adds m new links or with probability $1 - p$ joins a new node with m links. The way of selecting existing nodes for building new links or attaching new nodes is specified by the shifted preferential function $p_i(t)$:

$$p_i(t) = \frac{d_i(t) - \lambda}{D_\lambda(t)}, \tag{30}$$

where $d_i(t)$ is the degree of node i at time t, $\lambda \leq 1$ is the shift parameter fed into the algorithm, and $D_\lambda(t) = \sum_{k=1}^{n(t)} (d_k(t) - \lambda)$ is the total shifted degree.

A. Derivation of $\pi(x, y)$

In what follows, we assume that all nodes join the system sequentially and that each node i arrives at time t_i. It has been shown in [10] that degree $d_i(t)$ of node i follows:

$$d_i(t) = \left(\frac{\alpha t - b}{\alpha t_i - b}\right)^{1/\alpha} (m - \lambda) + \lambda, \tag{31}$$

where α is the shape of the Pareto degree distribution and b is a constant:

$$\alpha = \frac{2m - \lambda(1 - p)}{m(1 + p)}, \quad b = \frac{m_0 \lambda}{m(1 + p)}. \tag{32}$$

The total shifted degree $D_\lambda(t)$ at time t in GLP graphs is then given by:

$$D_\lambda(t) = 2mt - (m_0 + (1 - p)t)\lambda = m(1 + p)(\alpha t - b). \tag{33}$$

Now, consider two nodes i, j in the graph and suppose that j joins after i, i.e., $t_i < t_j$. Denote by $M_{ij}(t)$ the number of duplicate links between i and j accumulated up to time t and by $L_{ij}(t)$ the indicator variable that represents if any link between i, j exists by time t:

$$L_{ij}(t) = \begin{cases} 1 & M_{ij}(t) \geq 1 \\ 0 & M_{ij}(t) = 0 \end{cases}. \tag{34}$$

Using similar technique as in PLRG, we obtain formulas for $M_{ij}(t)$ and $L_{ij}(t)$. For simplicity, we only present in the next lemma the result for $\alpha \neq 2$ and leave the analysis of the case $\alpha = 2$ to the discussion of BA.

Theorem 6: For $t_i < t_j$, the expected duplicated links between i and j in GLP with $\alpha \neq 2$ is given by:

$$E[M_{ij}(t)] = c_1 \frac{(d_i(t) - \lambda)(d_j(t) - \lambda)}{D_\lambda(t)} \left(p + c_2 (d_j(t) - \lambda)^{\alpha - 2}\right), \tag{35}$$

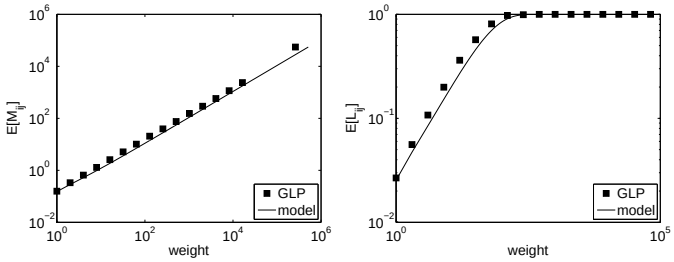
and $L_{ij}(t)$ is asymptotically Bernoulli with expectation:

$$E[L_{ij}(t)] = 1 - e^{-E[M_{ij}(t)]}, \tag{36}$$

where c_1 and c_2 are constants:

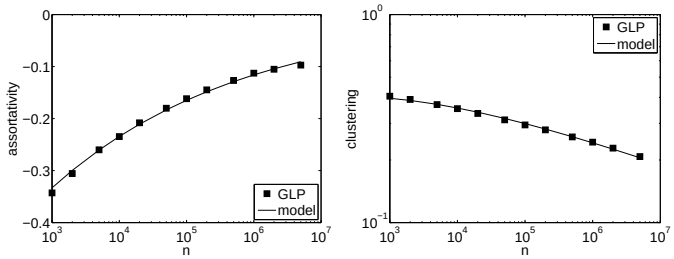
$$c_1 = \frac{2}{(1 + p)(2 - \alpha)}, \quad c_2 = \frac{m\lambda(1 + p)}{2(m - \lambda)^{\alpha - 1}}. \tag{37}$$

To verify models (35) and (36), we generated a GLP graph with 10^6 nodes using parameters $m_0 = 2, m = 1.13, p = 0.4695, \lambda = 0.6$. We fixed $i = 1$ and uniformly



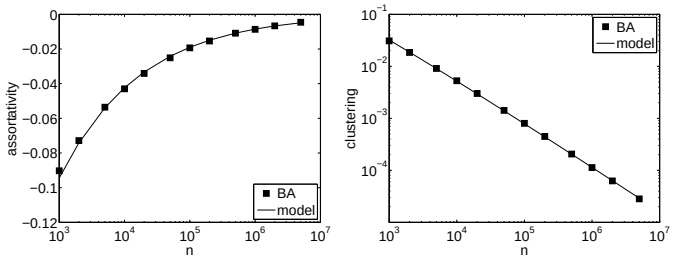
(a) $E[M_{ij}]$ (b) $E[L_{ij}]$

Fig. 6. Simulations of GLP ($\alpha = 1.2$). All cases use $10,000$ iterations.



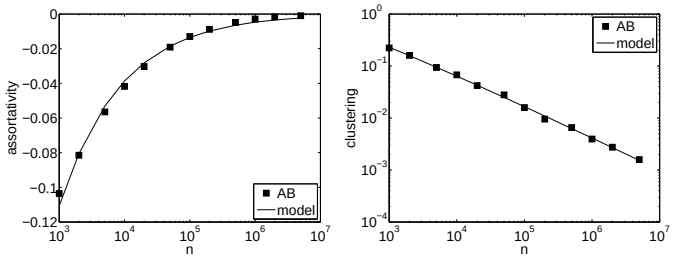
(a) assortativity (b) clustering

Fig. 7. Simulations of GLP ($\alpha = 1.2$). All cases use $10,000$ iterations.



(a) assortativity (b) clustering

Fig. 8. Simulations of BA. All cases use $10,000$ iterations.



(a) assortativity (b) clustering

Fig. 9. Simulations of AB ($\alpha = 1.2$). All cases use $10,000$ iterations.

randomly select node j from the rest nodes and computed the average number of duplicate/non-duplicate links. Fig. 6 plots the obtained $E[M_{ij}(t)]$ and $E[L_{ij}(t)]$ against d_j from the simulations along with the model values. The figure shows that models (35) and (36) follow the simulation results pretty accurately.

Notice that the result in (36) assumes that node j arrives after i. However, for any pair of randomly chosen nodes i, j, their join times satisfy $P(t_i < t_j) = P(t_i > t_j) = 1/2$. Further note that we set the *weight* of each node i to be its shifted degree, i.e., $w_i = d_i - \lambda$. This leads to the weight distribution $F(x)$ being the same Pareto distribution $1 - (x/\beta)^{-\alpha}$ as in GED and allows application of π-equivalence later.

Corollary 3: In GLP with $\alpha \neq 2$, link-existence probability $\pi(x, y)$ is given by:

$$\pi(x,y) = 1 - 1/2\exp\Big(-c_1\frac{xy}{D}\Big(p + c_2 x^{\alpha-2}\Big)\Big) - 1/2\exp\Big(-c_1\frac{xy}{D}\Big(p + c_2 y^{\alpha-2}\Big)\Big), \quad (38)$$

where x, y are node weights, D is the total weight, and c_1, c_2 are given in (37).

B. Bounds on $\pi(x, y)$

Now we use simple exponential functions to bound $\pi(x, y)$. In the next lemma, we only show the results for the case of $\lambda > 0$ (i.e., $1 < \alpha < 2$). However, the following results also hold for $\lambda < 0$ after reversing the direction of each inequality.

Lemma 6: In GLP with $\lambda > 0$, the link-existence probability function $\pi(x, y)$ is bounded as follows:

$$1 - \exp\Big(-\frac{xy}{c_4 D}\Big) \leq \pi(x,y) \leq 1 - \exp\Big(-\frac{xy}{c_3 D}\Big), \quad (39)$$

where D is the total degree and c_3, c_4 are constants of D.

The following results show that we can infer the asymptotics of $E[r(G)]$ and $E[\gamma(G)]$ of GLP from those of PLRG and, through transitivity, those of GED.

Theorem 7: GLP is π-equivalent to PLRG.

Corollary 4: As graph size $n \to \infty$, two- and three-node correlation in GLP decays to zero according to (15) and (22).

Again, we verify Corollary 4 by constructing $10,000$ GLP graphs and extracting $r(G)$ and $\gamma(G)$. We plot in Fig. 7 the values from simulations as well as from models (15) and (22). The figure displays that the model curves follow the actual values very accurately.

It is interesting to note that while GLP has significantly higher clustering than GED/PLRG according to [10], [50], the results in Corollary 4 for the first time analytically show that GLP is asymptotically the same as GED/PLRG in the limit of large n.

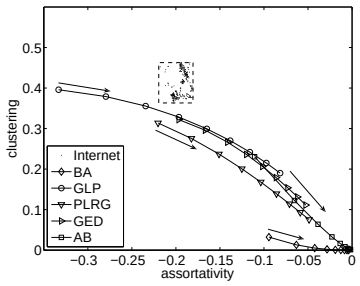
C. Application to BA and AB

Note that BA is a special case of GLP with $p = 0$ and $\lambda = 0$, from which it follows that $\alpha = 2$ and $b = 0$. Notice that (38) does not admit $\alpha = 2$ since $c_1 = \infty$. We repeat the derivation of (35)-(36) with $\alpha = 2, p = 0$, and $\lambda = 0$ to yield the following results.

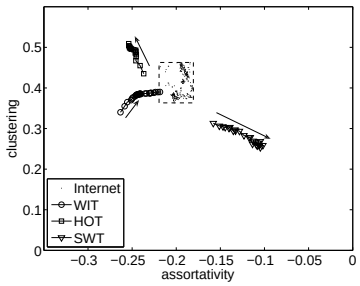
Theorem 8: The link existence probability $\pi(x, y)$ in BA is identical to that of PLRG given in (29).

Corollary 5: BA is π-equivalent to PLRG and the following asymptotic decay rates hold for BA: $E[r(G)] = \Theta(n^{-1}\log^2 n)$ and $E[\gamma(G)] = \Theta(n^{-1}\log^2 n)$.

Considering that AB without edge rewiring is just another special case of GLP, we get the following results by setting $\lambda = -1$ in (39).



(a) degree-based



(b) link-based

Fig. 10. Comparison of degree-based and link-based models to the Internet. All markers are obtained from asymptotic models (15), (22) in (a) and from simulations in (b). Arrows indicate the trends in corresponding curves.

Corollary 6: AB (without rewiring) is π-equivalent to PLRG and its correlation satisfies (15), (22) as $n \to \infty$.

Not surprisingly, both BA and AB have decaying assortativity and clustering, just as GLP. Simulation results of BA and AB are given in Fig. 8-9, which show that the asymptotic degree correlation in BA and AB graphs can be accurately predicted using (15), (22).

VII. Discussion

In this section, we compare the evolution of degree correlation given in (15), (22) to that of the Internet AS-level structure. We specifically focus on the subgraphs of the Internet induced by *provider-customer* links, which have been shown in [20] to exhibit a power-law degree distribution. According to [20], the combination of BGP routing tables from [39], [40] provides a reasonably complete view of the provider-customer links in the Internet.

We first obtain more than 300 snapshots of the Internet topology from BGP routing tables [39], [40], which cover the last 7 years as the size of the Internet has increased from $4,000$ to $23,000$ nodes. We then extract degree correlation from these snapshots and plot them in Fig. 10 with $r(G)$ on the x-axis and $\gamma(G)$ on the y-axis. The figure indicates that both degree correlation coefficients of the Internet, shown as dots inside a small dashed rectangle, do not significantly change over the years, which is consistent with observations in [50].

We also plot in Fig. 10(a) the corresponding metrics computed from (15) and (22) for the studied degree-based algorithms with graph size ranging from 10^3 to 10^7. The figure shows that for small graph sizes, degree correlation of GLP/GED/PLRG is close to that of the Internet; however, all studied degree-based models converge both assortativity and clustering to zero as system size increases. This contradicts the behavior of these metrics in real networks and suggests that attachment decisions of users in small-world graphs may be based on additional factors besides the degree of potential neighbors.

To support this conjecture, we next study three link-based graph algorithms: *Small-World Topologies* (SWT) [23], *Heuristically Optimized Trade-offs* (HOT) [17], and *Wealth-based Internet Topology* (WIT) [50]. Recall that SWT forms p percent of links using geographic preference and creates the rest of the links using random pairing as in PLRG; HOT models each attachment decision as an optimization problem with the goal of minimizing both the geographical length of potential links and the average number of hops to other nodes in the graph; and WIT adjusts the number of links based on an underlying stochastic wealth process and selects neighbors based on random walks.

We configure SWT/HOT/WIT in simulations to produce power-law degree distributions that are the similar to those observed in the Internet. In SWT, we set $p = 0.31$ as in [23]; in HOT, we add two new links to each new node as in [50]; for WIT, we used the parameters of [50]. For each of these three algorithms, we construct $10,000$ graphs with n increasing from 10^3 to 10^7 and extract the averaged degree correlation.

Observe from Fig. 10(b) that SWT exhibits decreasing $r(G)$ and $\gamma(G)$ and displays a similar trend as degree-based algorithms, which can be explained by the fact that SWT contains PLRG-like random pairing. In contrast, notice that HOT increases its clustering coefficient as the graph becomes more disassortative, an opposite trend to that of degree-based methods. Both approaches, however, exhibit much less sensitivity to increasing n than degree-based generators and reach a saturation point as $n \to \infty$, which closely replicates the qualitative behavior of the Internet. Finally, $r(G)$ and $\gamma(G)$ of WIT graphs stabilize around those of the Internet, even reaching and staying inside the dashed rectangle for $n \geq 50,000$.

These results show the significant difference between degree-based and link-based methods, as well as underline the importance of modeling higher-order degree correlation in constructed graphs.

VIII. Conclusion

In this paper, we examined several degree-based topology generators and derived the asymptotic shape of their assortativity and clustering as $n \to \infty$. We found that all studied degree-based generators were virtually uncorrelated for sufficiently large graph size. Our simulations also showed that link-based algorithms [17], [23], [50], which grew the system based on the structure of the current graph rather than just the degree, were much more capable of keeping two and three-node correlation close to time-invariant.

Future work involves analysis of link-based methods and impact of higher-order degree correlation on graph structure.

IX. Acknowledgements

We would like to thank Dr. Walter Willinger, Xiafeng Li, and anonymous INFOCOM reviewers for helpful input on earlier drafts of the paper.

References

[1] W. Aiello, F. R. K. Chung, and L. Lu, "A Random Graph Model for Massive Graphs," in *Proc. ACM STOC*, May 2000, pp. 171–180.

[2] R. Albert and A. Barabási, "Topology of Evolving Networks: Local Events and Universality," *Phys. Rev. Lett.*, vol. 85, no. 24, pp. 5234–5237, Dec. 2000.

[3] A.-L. Barabasi and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.

[4] A. Barrat and R. Pastor-Satorras, "Rate Equation Approach for Correlations in Growing Network Models," *Phys. Rev. E*, vol. 71, no. 3, Mar. 2005.

[5] E. A. Bender and E. R. Canfield, "The Asymptotic Number of Labeled Graphs with Given Degree Sequences," *J. Combin. Theory Ser. A*, vol. 24, pp. 296–307, 1978.

[6] G. Bianconi and A. Barabási, "Competition and Multiscaling in Evolving Networks," *Europhys. Lett.*, vol. 54, no. 4, pp. 436–442, May 2001.

[7] M. Boguna and R. Pastor-Satorras, "Class of Correlated Random Networks with Hidden Variables," *Phys. Rev. E*, vol. 68, no. 3, p. 036112, Sep. 2003.

[8] B. Bollobás and O. Riordan, "Mathematical Results on Scale-free Random Graphs," in *Handbook of Graphs and Networks*, S. Bornholdt and H. G. Schuster, Eds. Wiley-VCH, Berlin, 2002.

[9] S. Bornholdt and H. Ebel, "World Wide Web Scaling Exponent from Simon's 1955 Model," *Phys. Rev. E*, vol. 64, no. 3, p. 035104, Aug. 2001.

[10] T. Bu and D. Towsley, "On Distinguishing between Internet Power Law Topology Generators," in *Proc. IEEE INFOCOM*, vol. 2, Jun. 2002, pp. 638–647.

[11] G. Caldarelli, A. Capocci, P. D. L. Rios, and M. A. Munoz, "Scale-Free Networks from Varying Vertex Intrinsic Fitness," *Phys. Rev. Lett.*, vol. 89, no. 25, p. 258702, Dec. 2002.

[12] H. Chang, S. Jamin, and W. Willinger, "To Peer or not to Peer: Modeling the Evolution of the Internets AS-level Topology," in *Proc. IEEE INFOCOM*, Apr. 2006.

[13] F. R. K. Chung and L. Lu, "Connected Components in Random Graphs with Given Expected Degree Sequences," *Annals of Combinatorics*, vol. 6, no. 2, pp. 125–145, Nov. 2002.

[14] S. N. Dorogovtsev, "Clustering of Correlated Networks," *Phys. Rev. E*, vol. 69, no. 2, p. 027104, Feb. 2004.

[15] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford Univ. Press, 2003.

[16] P. Erdös and A. Rényi, "On Random Graphs I," *Publication Math. Debrecen*, vol. 6, pp. 290–297, 1959.

[17] A. Fabrikant, E. Koutsoupias, and C. Papadimitriou, "Heuristically Optimized Tradeoffs," in *Proc. ICALP*, Jul. 2002, pp. 110–122.

[18] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-law Relationships of the Internet Topology," in *Proc. ACM SIGCOMM*, Aug. 1999, pp. 251–262.

[19] K. I. Goh, B. Kahng, and D. Kim, "Universal Behavior of Load Distribution in Scale-Free Networks," *Phys. Rev. Lett.*, vol. 87, no. 27, p. 278701, Dec. 2001.

[20] Y. He, G. Siganos, M. Faloutsos, and S. V. Krishnamurthy, "A Systematic Framework for Unearthing the Missing Links: Measurements and Impact," in *Proc. USENIX/SIGCOMM NSDI*, Apr. 2007, pp. 3632–3637.

[21] P. Holme and J. Zhao, "Exploring the Assortativity-Clustering Space of a Network's Degree Sequence," *Phys. Rev. E*, vol. 75, no. 4, p. 046111, Apr. 2007.

[22] C. Jin, Q. Chen, and S. Jamin, "Inet: Internet Topology," Univ. of Michigan, Tech. Rep. CSE-TR-456-02, 2002.

[23] S. Jin and A. Bestavros, "Small-world Internet Topologies: Possible Causes and Implications on Scalability of End-system Multicast," in *Proc. IEEE/ACM International Symposium on Modeling, Analysis and Simulation (MASCOTS)*, Oct. 2003.

[24] K. Klemm and V. M. Eguiluz, "Growing Scale-Free Networks with Small World Behavior," *Phys. Rev. E*, vol. 65, no. 5, p. 057102, May 2002.

[25] P. L. Krapivsky, S. Redner, and F. Leyvraz, "Connectivity of Growing Random Networks," *Phys. Rev. Lett.*, vol. 85, no. 21, pp. 4629–4632, Nov. 2000.

[26] T. Luczak, "Sparse Random Graphs with a Given Degree Sequence," *Random Graphs*, vol. 2, pp. 165–182, 1992.

[27] D. Magoni and J. Pansiot, "Analysis of the Autonomous System Network Topology," *ACM SIGCOMM Comp. Comm. Rev.*, vol. 31, pp. 26–37, 2001.

[28] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, "Systematic Topology Analisys and Generation Using Degree Correlations," in *Proc. ACM SIGCOMM*, Sep. 2006, pp. 135–146.

[29] M. Mihail and N. Visnoi, "On Generating Graphs with Prescribed Degree Sequences for Complex Network Modeling Applications," in *Proc. ARACNE*, Sep. 2002.

[30] M. Molloy and B. Reed, "A Critical Point for Random Graphs with a Given Degree Sequence," *Random Structures and Algorithms*, vol. 6, pp. 161–179, 1995.

[31] M. E. J. Newman, "Assortative Mixing in Networks," *Phys. Rev. Lett.*, vol. 89, no. 20, p. 208701, Oct. 2002.

[32] M. E. J. Newman, "Random Graphs as Models of Networks," in *Handbook of Graphs and Networks*, S. Bornholdt and H. G. Schuster, Eds. Wiley-VCH, Berlin, 2002.

[33] M. E. J. Newman, "Mixing Patterns in Networks," *Phys. Rev. E*, vol. 67, no. 2, p. 026126, Feb. 2003.

[34] M. E. J. Newman and J. Park, "Why Social Networks are Different from Other Types of Networks," *Phys. Rev. E*, vol. 68, no. 3, p. 036122, Sep. 2003.

[35] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random Graphs with Arbitrary Degree Distributions and their Applications," *Phys. Rev. E*, vol. 64, no. 2, p. 026118, Jul. 2001.

[36] R. Pastor-Satorras, A. Vazquez, and A. Vespignani, "Dynamical and Correlation Properties of the Internet," *Phys. Rev. Lett.*, vol. 87, no. 25, p. 258701, Dec. 2001.

[37] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press, 2004.

[38] A. R. Puniyani, R. M. Lukose, and B. A. Huberman, "Intentional Walks on Scale Free Small Worlds," *cond-mat/0107212*, Jul. 2001.

[39] RIPE. [Online]. Available: http://www.ripe.net/.

[40] Route Views Project. [Online]. Available: http://www.antc.uoregon.edu/route-views.

[41] M. A. Serrano and M. Boguna, "Tuning Clustering in Random Networks with Arbitrary Degree Distributions," *Phys. Rev. E*, vol. 72, no. 3, p. 036133, Sep. 2005.

[42] H. A. Simon, "On a Class of Skew Distribution Functions," *Biometrika*, vol. 42, pp. 425–440, 1955.

[43] H. A. Simon, "Some Further Notes on a Class of Skew Distribution Functions," *Information and Control*, vol. 3, pp. 80–88, 1960.

[44] B. Soderberg, "General Formalism for Inhomogeneous Random Graphs," *Phys. Rev. E*, vol. 66, no. 6, p. 066121, Dec. 2002.

[45] S. N. Soffer and A. Vazquez, "Network Clustering Coefficient without Degree-Correlation Biases," *Phys. Rev. E*, vol. 71, no. 5, p. 057101, May 2005.

[46] G. Szabo, M. Alava, and J. Kertesz, "Structural Transitions in Scale-Free Networks," *Phys. Rev. E*, vol. 67, no. 5, p. 056102, May 2003.

[47] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Network Topology Generators: Degree-Based vs. Structural," in *Proc. ACM SIGCOMM*, Aug. 2002, pp. 147–159.

[48] A. Vazquez, R. Pastor-Satorras, and A. Vespignani, "Large-Scale Topological and Dynamical Properties of the Internet," *Phys. Rev. E*, vol. 65, no. 6, p. 066130, Jun. 2002.

[49] X. Wang, X. Liu, and D. Loguinov, "Modeling the Evolution of Degree Correlation in Scale-Free Topology Generators," Texas A&M University, Tech. Rep. 2007-12-1, Dec. 2007. [Online]. Available: http://irl.cs.tamu.edu/publications/.

[50] X. Wang and D. Loguinov, "Wealth-Based Evolution Model for the Internet AS-Level Topology," in *Proc. IEEE INFOCOM*, Apr. 2006.

[51] D. J. Watts and S. Strogatz, "Collective Dynamics of Small World Networks," *Nature*, vol. 393, pp. 440–442, Jun. 1998.

[52] S. Yook, H. Jeong, and A. Barabási, "Modeling the Internet's Large-scale Topology," *Natl. Acad. Sci.*, vol. 99, no. 21, pp. 13382–13386, Oct. 2002.

[53] S. Zhou and R. J. Mondragon, "Accurately Modeling the Internet Topology," *Phys. Rev. E*, vol. 70, no. 6, p. 066108, Dec. 2004.