2015 IEEE 54th Annual Conference on Decision and Control (CDC)
December 15-18, 2015. Osaka, Japan

Transitivity of Reciprocal Networks

Isabel Fernández[1] and Jorge Finke[2]

Abstract— **Network models are a useful tool to describe and predict dynamic relationships in large collections of data. Characterizing these relationships helps us to explain the emergence of structure as a systematic deviation from random connectivity. This paper introduces an event-driven model that captures the effects of three simple network formation mechanisms: random attachment (a generic abstraction of how a new incoming node connects to a network), triad formation (how the new node establishes transitive relationships), and network response (the way the overall network reacts to attachments). Our work focuses on the impact of the latter on clustering and degree distributions. We prove that any initial network will reach stationary local clustering coefficients, and obey an extended power law distribution for the in-degree and an exponential distribution for the out-degree. For the in-degree in particular, the response mechanism amplifies the scaling behavior that results from the other two mechanisms.**

I. Introduction

Dynamic network models are a useful tool to describe the evolution of relationships in large sets of data. Modeling the dynamics of structure provides the analytical basis for predicting global and local patterns of connectivity, characterizing sometimes subtle but systematic deviations that result from specific sensing and decision-making mechanisms (e.g., how reciprocal responses shape the degree distributions of human communication networks).

The mechanisms of preferential attachment and triad formation are believed to underlie a wide class of empirical networks [1]. Models combining these two mechanisms have been able to explain the emergence of extended power laws in the degree distribution of growing, highly-clustered networks (i.e., where nodes with a high degree obey a power law distribution, and nodes with a low degree an exponential distribution). Less attention has been paid to formation mechanisms where reciprocity, an essential feature in human communication networks, is thought to be the basis for establishing numerous links. A better, understanding of the impact of reciprocal responses on the in- and out-degree distributions and the local clustering coefficients of networks remains an open challenge [2], [3].

This work introduces a simple network response mechanism, comprised of random and reciprocal approaches, and explains how random attachment and triad-formation, together with the proposed response mechanism, impact the structure of growing networks. It is closely related to the work in [4]–[6], which proposes various formation models that characterize the degree distribution and clustering properties of growing networks, but which does not consider the effects of reciprocity. Our main results present analytical expressions for the asymptotic behavior of (i) the cumulative distribution functions of the in-degree and out-degree, and (ii) the local clustering coefficients for reciprocal networks.

II. A network formation model

Consider an ordered set of graphs $\mathcal{G} = \{\mathcal{G}(0), \mathcal{G}(1), \ldots\}$, where each graph $\mathcal{G}(t) = (\mathcal{H}(t), \mathcal{A}(t))$ describes a network at time index t with a set of nodes $\mathcal{H}(t) = \{1, \ldots, N_t\}$ and a set of directed edges $\mathcal{A}(t) = \{(i,j) : i,j \in \mathcal{H}(t)\}$. The pair $(j,i) \in \mathcal{A}(t)$ indicates that there exists an edge from node j to node i at time t, and $\mathcal{Q}_i(t) = \{j \in \mathcal{H}(t) : (j,i) \in \mathcal{A}(t)\}$ represents the set of incoming neighbors of node i. Let $k_i(t) = |\mathcal{Q}_i(t)|, k_i(t) \geq 0$, represent the in-degree of node i. Similarly, $\hat{\mathcal{Q}}_i(t) = \{j \in \mathcal{H}(t) : (i,j) \in \mathcal{A}(t)\}$ represents the set of outgoing neighbors of node i and $\hat{k}_i(t) = |\hat{\mathcal{Q}}_i(t)|, \hat{k}_i(t) > 0$, its out-degree. Assume the following mechanisms underlie the process of network evolution.

- M1 *Random attachment:* A new node links to $m \geq 1$, $m \in \mathbb{Z}^+$, different nodes, selected according to a uniformly random distribution over $\mathcal{H}(t-1)$.
- M2 *Triad formation:* For every link that a new node establishes during random attachment, it tries to establish an additional link within the new neighborhood. In particular, if node $j \notin \mathcal{H}(t-1)$ connects to some node $j' \in \mathcal{H}(t-1)$, it may also connect to an outgoing neighbor of node j' with probability $0 < \pi_f \leq 1$ (selected again according to a uniformly random distribution over $\hat{\mathcal{Q}}_{j'}(t-1)$).
- M3 *Network response:* There are two ways the network responds to the attachment of a new node. The first approach is based on reciprocity: Each of the m randomly selected nodes establishes a reciprocal link with probability $0 \leq \pi_r \leq 1$. The second approach shows no preference for establishing reciprocal links: A set of $n \geq 0$ randomly selected nodes connect to the new node.

Note that if both $\pi_r = 0$ and $n = 0$, that is, the existing nodes do not respond to node attachment, then mechanisms M1-M2 alone cannot induce directed cycles. Note also that the model does not allow self-loops.

To ensure that the formation mechanisms are properly defined, we require the following assumptions.

- A1 *The initial network*: The network $\mathcal{G}(0)$ is weakly connected and has more than $2m$ nodes, each with at least one outgoing neighbor.

[1]Department of Electrical and Computer Engineering, The Ohio State University, Columbus, Ohio.
[2]Department of Electrical Engineering and Computer Science, Pontificia Universidad Javeriana, Cali, Colombia.

A2 *The event-time incidence rate*: Mechanisms M1-M3 are triggered every index $t \in \mathcal{T}$, where $\mathcal{T}$ is a set of uniformly distributed time indices.

Here, we let a new node attach to the network every time index t, which implies that $\mathcal{H}(t) = \{1, \ldots, N_0 + t\}$ where $N_0 \geq 2m$ indicates the number of nodes at $t = 0$.

III. IN-DEGREE DISTRIBUTION

The following theorem presents sufficient conditions that guarantee that mechanisms M1-M3 yields a scaling behavior in the in-degree distribution of the network.

Theorem 1 (power law in-degree distribution): Suppose that assumptions A1-A2 hold. The asymptotic behavior of the in-degree distribution $p_k(t)$ follows a power law $p_k(t)|_{t\to\infty} \sim k^{-\alpha}$ with scaling exponent

$$\alpha = 2 + \frac{1}{\pi_f} + \frac{\pi_r}{\pi_f} + \frac{n}{m\pi_f}$$

for nodes with a degree $k \gg (\alpha - 1)m$.

Proof: Note that the expected in-degree of a randomly selected node at any time index t depends on all three mechanisms. However, to characterize the *rate of change* of the in-degree of a node (node i), we only need to take into account the formation of newly established incoming edges, which can only occur due to mechanisms M1-M2 (mechanism M3 can only establish outgoing edges at $t \neq t_i$). According to mechanism M1 and assumption A2, the instant node $j \notin \mathcal{H}(t-1)$ joins the network, $t = t_j \neq t_i$, it connects to m different nodes based on a uniform random attachment process. The probability that node j attaches to node $i \in \mathcal{H}(t-1)$ is $\frac{m}{N_0+t-1}$. If during attachment node j connects to an incoming neighbor of node i (some node $j' \in \mathcal{Q}_i(t)$), then it establishes an additional link to node i with probability π_f (based on mechanism M2). On average, triad formation increases the rate of change of the in-degree of node i by $\left(\frac{mk_i(t)}{N_0+t-1}\right)\left(\frac{\pi_f}{m(1+\pi_f)+n+m\pi_r}\right)$. The first term represents the probability of selecting, during random attachment, an incoming neighbor of node i. The second term represents the probability of choosing any outgoing neighbor of node j' (including node i) and forming a triad. In particular, note that the expected out-degree of node j' is the sum of the number of outgoing edges established at time $t_{j'}$ (i.e., m edges according to mechanism M1 plus an expected $m\pi_f$ additional edges according to mechanism M2) and the expect number of edges established based on mechanism M3. The overall rate of change of $k_i(t)$ is

$$\frac{dk_i(t)}{dt} = \frac{m(m(1+\pi_f+\pi_r)+n) + m\pi_f k_i(t)}{(N_0+t-1)(m(1+\pi_f+\pi_r)+n)} \quad (1)$$

with boundary condition $k_i(t_i) = n + m\pi_r$ (which corresponds to the average number of incoming edges established at time t_i through both approaches for mechanism M3). The solution to eq. (1) is

$$k_i(t) = (m(1+\pi_r)+n)\left(1+\frac{1}{\pi_f}\right)\left(\frac{N_0+t-1}{N_0+t_i-1}\right)^{\frac{1}{\alpha-1}} - m(\alpha-1) \quad (2)$$

where $\alpha = 1 + \left(1 + \frac{1}{\pi_f} + \frac{\pi_r}{\pi_f} + \frac{n}{m\pi_f}\right)$. The analytical expression for the cumulative distribution of the in-degree of node i equals

$$P[k_i(t) \leq k] = P\left[t_i \geq \left(\frac{(m(1+\pi_r)+n)(\pi_f+1)}{\pi_f(k+m(\alpha-1))}\right)^{\alpha-1} (N_0+t-1) - (N_0-1)\right]$$

Because new nodes are added at a constant rate over time (according to assumption A2), letting $P_k(t) = P[k_i(t) \leq k]$, we know

$$P_k(t) = \frac{1}{t}\left[t - \left(\frac{(m(1+\pi_r)+n)(\pi_f+1)}{k+m(\alpha-1)}\right)^{\alpha-1} (N_0+t-1) + (N_0-1)\right] \quad (3)$$

As $t \to \infty$

$$P_k(t)|_{t\to\infty} = 1 - \left(\frac{(m(1+\pi_r)+n)\left(1+\frac{1}{\pi_f}\right)}{k+m(\alpha-1)}\right)^{\alpha-1} \quad (4)$$

Finally,

$$p_k(t)|_{t\to\infty} = \left.\frac{dP_k(t)}{dk}\right|_{t\to\infty} = d_1\,(k+m(\alpha-1))^{-\alpha} \quad (5)$$

where $d_1 = (\alpha-1)\left((m(1+\pi_r)+n)\left(1+\frac{1}{\pi_f}\right)\right)^{\alpha-1}$ and $\alpha = 2 + \frac{1}{\pi_f}\left(1+\pi_r+\frac{n}{m}\right)$. When $k \gg (\alpha-1)m$, eq. (5) represents a power law that satisfies $p_k(t)|_{t\to\infty} \sim k^{-\alpha}$. ■

Theorem 1 implies that the probability distribution of the in-degree becomes stationary as the network grows in size. This distribution follows a power law for nodes with a degree much greater than $(\alpha-1)m$. Note that the scaling exponent depends on the parameters of all three mechanisms, and that both the ratio $\frac{n}{m}$ and the probability π_r have similar effects on the exponent α. Figure 1 shows the resulting scaling exponent α for different values of n and varying m, π_r, and π_f. Note that small values of π_f have the largest impact on α. Empirical networks usually have a scaling exponent $\alpha \leq 5$, suggesting that the parameter of the triad formation should be bounded be bounded by $\pi_f \geq 0.6$ [7]. Note also that both approaches for the response mechanism amplify the scaling behavior. For nodes with a low degree, we obtain the following result.

Corollary 1 (exponential in-degree distribution): Suppose that assumptions A1-A2 hold. The asymptotic behavior of the in-degree distribution $p_k(t)$ follows an exponential form $p_k(t)|_{t\to\infty} \sim \exp(-\lambda k)$ with exponent

$$\lambda = \frac{\pi_f}{m(\pi_f+1+\pi_r)+n} + \frac{1}{m} \quad (6)$$

for nodes with a degree $k \ll \frac{m}{m\lambda-1}$.

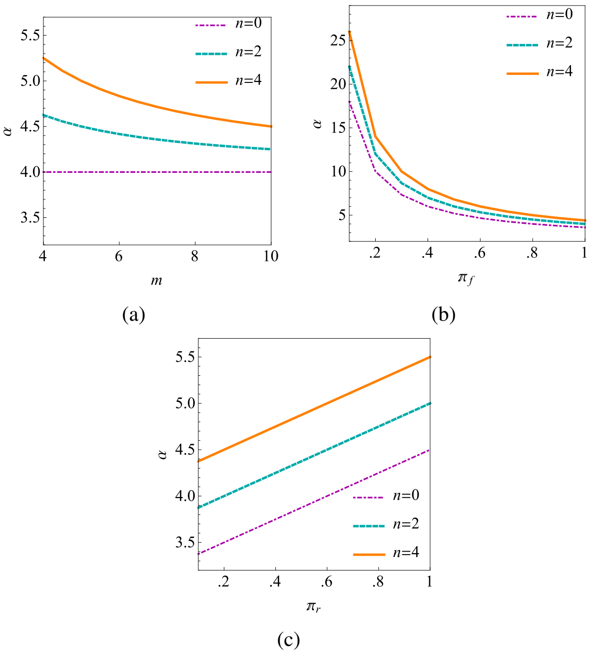


Fig. 1. Scaling exponent for different values of n and fixed values $m = 5$, $\pi_f = 0.8$, and $\pi_r = 0.6$; (a) varying m; (b) varying π_f; (c) varying π_r.

Proof: Using eq. (5), note that when $k \ll (\alpha-1)m$ we know that

$$\begin{aligned}\ln p_k(t)|_{t\to\infty} &\sim -\alpha \ln(k + (\alpha-1)m)\\ &= -\alpha\left[\ln\left(1+\frac{k}{(\alpha-1)m}\right) + \ln((\alpha-1)m)\right]\\ &\sim -\alpha\left[\frac{k}{(\alpha-1)m} + \ln((\alpha-1)m)\right]\end{aligned}$$

and applying the exponential function on both sides, we obtain

$$p_k(t)|_{t\to\infty} \sim ((\alpha-1)m)^{-\alpha}\exp\left(-\alpha\frac{k}{(\alpha-1)m}\right) \quad (7)$$

When $k \ll (\alpha-1)m = \frac{m}{m\lambda-1}$, eq. (7) is proportional to the general form $\exp(-\lambda k)$ with $\lambda = \frac{\alpha}{(\alpha-1)m} = \frac{m(2\pi_f+1+\pi_r)+n}{m(m(\pi_f+1+\pi_r)+n)} = \frac{\pi_f}{m(\pi_f+1+\pi_r)+n} + \frac{1}{m}$. ■

Corollary 1 means that for nodes with a degree much smaller than $\frac{m}{m\lambda-1}$, the in-degree distribution has an exponential exponent that depends on the parameters of all three mechanisms.

Combined, Theorem 1 and Corollary 1 show that the network follows what is called an extended power law distribution with threshold $\varepsilon = (\alpha-1)m = \frac{m}{m\lambda-1}$. In general, network response has a noticeable effect on both the scaling and exponential exponents. Figure 2 quantifies how variations in any of the three mechanisms affect the complementary cumulative in-degree distribution of the network. The plots at the top illustrate the impact of random attachment and triad formation, and the plots at the bottom of the response mechanism. The vertical lines represent the transitions from an exponential to a power law distribution. Based on eq. (6), it is easy to see that the number of links established during random attachment has a significant effect on the head of the distribution. All parameters have a significant effect on the tail.

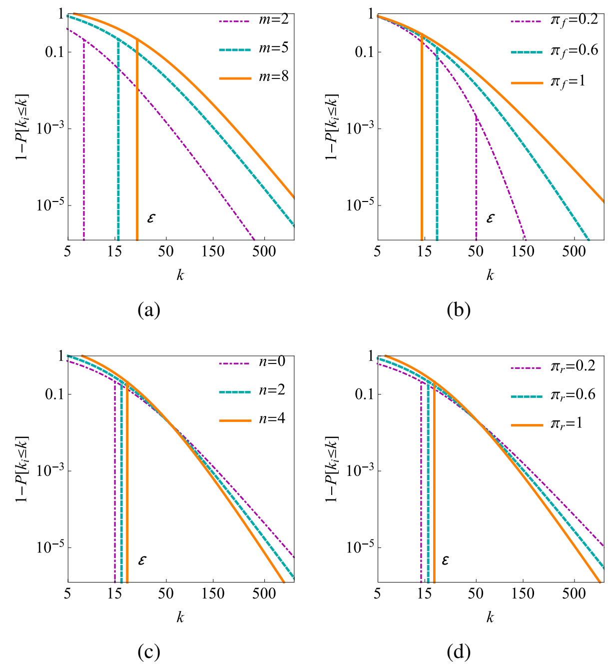


Fig. 2. Complementary cumulative in-degree distribution on a logarithmic scale; (a) variation in the number of edges formed due to random attachment; (b) variation in the probability of forming triads; (c) variation in the number of edges established due to the random response; (d) variation in the probability of forming reciprocal edges.

IV. Out-Degree Distribution

The following theorem presents sufficient conditions that guarantee that the proposed mechanisms yield an exponential out-degree distribution.

Theorem 2 (exponential out-degree distribution): Suppose that assumptions A1-A2 hold and mechanism M3 satisfies $n > 0$ or $\pi_r > 0$. The asymptotic behavior of the out-degree distribution $p_{\hat{k}}(t)$ follows an exponential form $p_{\hat{k}}(t)|_{t\to\infty} \sim \exp(-\beta\hat{k})$ with exponent

$$\beta = \frac{1}{n + m\pi_r}$$

for any node with out-degree $\hat{k} \gg m(1+\pi_f)$.

Proof: Note that the out-degree of a randomly selected node (node i) at time index t depends, again in general, on all mechanisms. However, to characterize the *rate of change* of the out-degree, we only need to take account of the additional edges established due to network response. According to mechanism M1 and assumption A2, when node $j \notin \mathcal{H}(t-1)$ attaches to the network at time $t = t_j$, the probability that it connects to node $i \in \mathcal{H}(t-1)$, and that node i establishes a reciprocal edge (based on mechanism M3) is $\frac{m\pi_r}{N_0+t-1}$. Random response increases the rate of change of the out-degree of node i by $\frac{n}{N_0+t-1}$. Thus, the overall rate of change of the out-degree of node i, denoted by $\hat{k}_i(t)$, is

$$\frac{d\hat{k}_i(t)}{dt} = \frac{m\pi_r}{N_0+t-1} + \frac{n}{N_0+t-1} \quad (8)$$

with boundary condition $\hat{k}_i(t_i) = m + m\pi_f$ (which corresponds to the number of outgoing edges established by node i at time t_i due to mechanisms M1 and M2). The solution to eq. (8) is

$$\hat{k}_i(t) = m(1+\pi_f) + (n+m\pi_r)\ln\left(\frac{N_0+t-1}{N_0+t_i-1}\right) \quad (9)$$

Using eq. (9), the analytical expression for the cumulative distribution of the out-degree of node i equals

$$P[\hat{k}_i(t) \le \hat{k}] = P\left[t_i \ge \exp\left(\frac{m(1+\pi_f)-\hat{k}}{n+m\pi_r}\right)(N_0+t-1) - (N_0-1)\right]$$

As in the proof of Theorem 1, because new nodes are added at a constant rate over time (according to assumption A2), and letting $P_{\hat{k}}(t) = P\left[\hat{k}_i(t) \le \hat{k}\right]$ we know

$$P_{\hat{k}}(t) = \frac{1}{t}\left[t - \exp\left(\frac{m(1+\pi_f)-\hat{k}}{n+m\pi_r}\right)(N_0+t-1) + (N_0-1)\right] \quad (10)$$

and as $t \to \infty$

$$P_{\hat{k}}(t)\big|_{t\to\infty} = 1 - \exp\left(\frac{m(1+\pi_f)-\hat{k}}{n+m\pi_r}\right) \quad (11)$$

Finally,

$$p_{\hat{k}}(t)\big|_{t\to\infty} = \frac{dP_{\hat{k}}(t)}{d\hat{k}}\bigg|_{t\to\infty} = d_2\exp\left(\frac{-\hat{k}}{n+m\pi_r}\right) \quad (12)$$

where $d_2 = \frac{1}{n+m\pi_r}\exp\left(\frac{m(1+\pi_f)}{n+m\pi_r}\right)$. Note that eq. (12) obeys an exponential distribution of the form $p_{\hat{k}}(\infty) \sim \exp(-\beta\hat{k})$, where $\beta = \frac{1}{n+m\pi_r}$ for any node with out-degree $\hat{k} \gg m(1+\pi_f)$. ■

Theorem 2 implies that the resulting probability distribution of the out-degree has an exponential exponent that depends on the parameters of both random attachment and network response, but not on triad formation. Note that according to eq. (11), as a consequence of the parameters that specify mechanisms M1 and M2, there does not exits any node with an out-degree $\hat{k} \le m(1+\pi_f)$.

Figure 3 quantifies how the mechanisms affect the complementary cumulative out-degree distribution (eq. (11)). The plots at the top examine the mechanisms of random attachment and triad formation; the plots at the bottom the two approaches for network response. Note that the values of m, n, and π_r can significantly impact the out-degree distribution. However, triad formation does not have a significant impact on the slope of the tail of the distribution

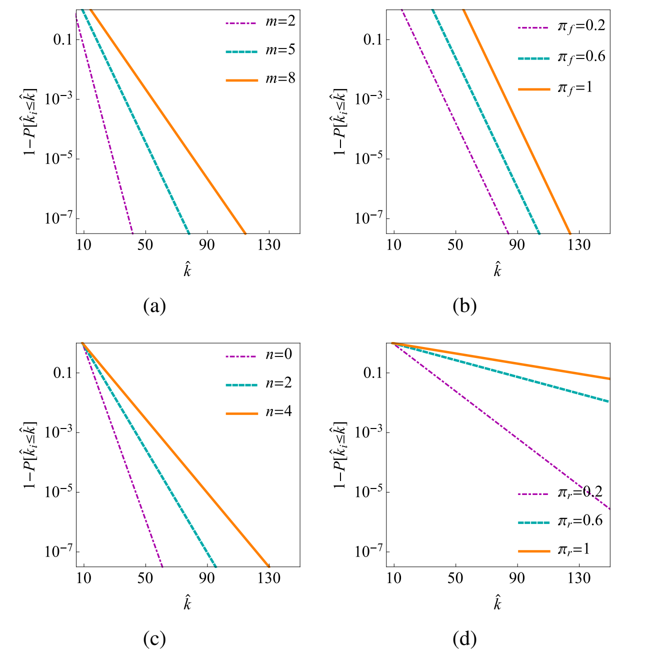


Fig. 3. Complementary cumulative out-degree distribution on a semi-logarithmic scale; (a) variation in the number of edges formed due to random attachment; (b) variation in the probability of forming triads; (c) variation in the number of edges established through random response; (d) variation in the probability of forming edges due to reciprocal response.

for the out-degree.

V. Local clustering coefficient

To characterize the evolution of the local clustering, we use the following definition of a triad.

Definition 1 (triad): Consider three nodes $i, j, j' \in \mathcal{H}(t)$ be such that $(j', j) \in \mathcal{A}(t)$. A subgraph is said to be a triad involving node j' if

1) $(j', i) \in \mathcal{A}(t)$ and either (j, i) or $(i, j) \in \mathcal{A}(t)$; or
2) $(i, j) \in \mathcal{A}(t)$ and $(i, j') \in \mathcal{A}(t)$.

In other words, triads represent all possible interconnections between three nodes, except the two closed three-way cycles. The local clustering coefficient of node i describes the ratio between the number of triads involving node i and the total number of possible triads that could involve this particular node. We are interested in understanding the evolution of the *average* local clustering coefficient that involves nodes with a particular in-degree.

To characterize the dynamics of clustering, we first need to express the (expected) out-degree of node i as a function of its in-degree.

Lemma 1 (expected out-degree of a node): Suppose that assumptions A1-A2 hold. The expected out-degree of node i is given by

$$\hat{k}_i(t) = m(1+\pi_f) + (n+m\pi_r)(\alpha-1)\ln\left(\frac{\pi_f(k_i(t)+(\alpha-1)m)}{(n+m(1+\pi_r))(1+\pi_f)}\right) \quad (13)$$

The proof of Lemma 1 follows directly from eqs. (2) and (9). The following theorem presents sufficient conditions to guarantee that mechanisms M1-M3 yield stationary local clustering coefficients.

Theorem 3 (clustering): Suppose that assumptions A1-A2 hold. The asymptotic behavior of the average local clustering coefficient over all nodes with in-degree k satisfies

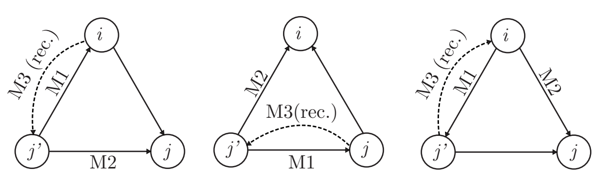
$$\begin{aligned} c(k) =& \frac{2(\pi_r+1)}{\pi_f(k+f(k))(k+f(k)-1)}\Bigg((n+m(1+\pi_f+\pi_r)) \\ & (\pi_f-1)\ln\left(\frac{\pi_f k+(1+\pi_f+\pi_r)m+n}{(n+m(1+\pi_r))(1+\pi_f)}\right) \\ & +\ \pi_f(k-n+m(\pi_f-\pi_r))\Bigg) \end{aligned}$$

where $f(k) = m(1+\pi_f) + (n+m\pi_r)(\alpha-1) \ln\left(\frac{\pi_f(k_i(t)+(\alpha-1)m)}{(n+m(1+\pi_r))(1+\pi_f)}\right)$.

Proof: To capture the expected local clustering coefficient of a randomly selected node (node i), we calculate, based on its total degree (in-degree plus out-degree), the expected number of triads and the maximum number of possible triads that involve node i. First, to estimate the maximum number of possible triads, we consider the number of possible pairs of incoming and outgoing edges of node i. Based on eq. (13), let $\hat{k}_i(t) = f(k_i(t))$. Then, for a fixed time index t, the total degree of node i is $k_i + f(k_i)$. The maximum number of possible pairs of incoming and outgoing edges of node i is

$$\Omega_i = \binom{k_i+f(k_i)}{2} = \frac{(k_i+f(k_i))(k_i+f(k_i)-1)}{2} \quad (14)$$

Now, to capture the expected number of triads that involve a randomly selected node with in-degree $k_i(t)$, we estimate the rate of change of the number of triads involving node i. Figure 4 illustrates all the possible link configurations that, according to Definition 1, may precede the formation of a triad.



(a) (b) (c)

Fig. 4. Possible link configurations involving node i; (a) node i has one incoming edge established through M1 and one outgoing edge; (b) node i has two incoming edges, one of them established through M2; (c) node i has two outgoing edges, one established through M1 and one through M2. M1 indicates an edge established through random attachment, M2 through triad formation, and M3(rec) through the reciprocal response approach of network response.

According to Figure 4(a), node i has an incoming edge that was established through random attachment from some node j', and an outgoing edge (established through any other mechanism). The triad is complete when node j' connects to one of the neighbors of node i. The probability of completing the triad and that another triad is formed because of reciprocal response is $\frac{m\pi_f}{N_0+t-1}(1+\pi_r)$. According to Figure 4(b), node i has an incoming edge from some node j (established through any mechanism) and the new node connects to node j through the process of random attachment. The triad will be closed if node j' connects to node i, which happens with probability π_f. Based on the probability of forming a triad with an edge that contributes to the in-degree of node i and due to the reciprocal response mechanism, on average, the rate of change of the maximum number of triads involving node i increases by $\frac{mk_i(t)}{N_0+t-1}\frac{\pi_f}{m(1+\pi_f+\pi_r)+n}(1+\pi_r)$. The overall rate of change of the expected number of triads involving node i is

$$\frac{d\omega_i(t)}{dt} = \frac{m\pi_f}{N_0+t-1}(1+\pi_r) + \left(\frac{k_i(t)(1+\pi_r)}{(\alpha-1)(N_0+t-1)}\right)$$

and using eq. (2), we obtain

$$\begin{aligned} \frac{d\omega_i(t)}{dt} = \frac{1+\pi_r}{N_0+t-1}\Bigg(& m(\pi_f-1) + \frac{1}{\alpha-1}\left(1+\frac{1}{\pi_f}\right) \\ & (n+m(1+\pi_r))\left(\frac{N_0+t-1}{N_0+t_i-1}\right)^{\frac{1}{\alpha-1}}\Bigg) \quad (15) \end{aligned}$$

with boundary condition $\omega_i(t_i) = m\pi_f + m\pi_f\pi_r$ (which corresponds to Figure 4(c) and indicates the number of triads that involve node i at $t = t_i$; note that triads may be formed through mechanisms M2-M3). The solution to eq. (15) is

$$\begin{aligned} \omega_i(t) =& \frac{\pi_r+1}{\pi_f}\Bigg((\pi_f-1)\, m\pi_f \ln\left(\frac{N_0+t-1}{N_0+t_i-1}\right) \\ & + (\pi_f+1)(m(1+\pi_r)+n)\left(\frac{N_0+t-1}{N_0+t_i-1}\right)^{\frac{1}{\alpha-1}} \\ & + m(\pi_f(\pi_f-\pi_r-1)-\pi_r-1)-n(\pi_f+1)\Bigg) \end{aligned}$$

Moreover, using eq. (2), we know that

$$\begin{aligned} \omega_i(t) =& \frac{\pi_r+1}{\pi_f}\Bigg((\pi_f-1)(n+m(1+\pi_f+\pi_r)) \\ & \ln\left(\frac{\pi_f k_i(t)+(1+\pi_f+\pi_r)m+n}{(n+m(1+\pi_r))(1+\pi_f)}\right) \\ & + (k_i(t)-n+m(\pi_f-\pi_r))\,\pi_f\Bigg) \quad (16) \end{aligned}$$

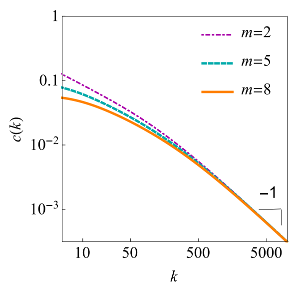
Finally, for a fixed time index t, dividing the expected number of triads (eq. (16)) by the maximum number of possible triads (eq. (14)), we know that the clustering coefficient of node i with in-degree $k_i = k$ satisfies

$$\begin{aligned} c(k) =& \frac{2(\pi_r+1)}{\pi_f(k+f(k))(k+f(k)-1)}\Bigg((n+m(1+\pi_f+\pi_r)) \\ & (\pi_f-1)\ln\left(\frac{\pi_f k+(1+\pi_f+\pi_r)m+n}{(n+m(1+\pi_r))(1+\pi_f)}\right) \\ & +\ \pi_f(k-n+m(\pi_f-\pi_r))\Bigg) \quad (17) \end{aligned}$$

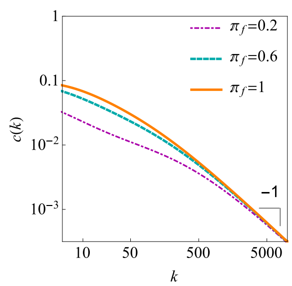
■

Theorem 3 implies that for a large enough network, its clustering properties do not depend on the initial network

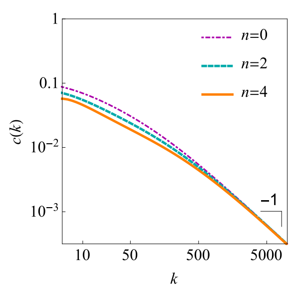
$\mathcal{G}(0)$ or its current size (i.e., the local clustering coefficients of $\mathcal{G}(t)$ do not vanish as t tends to infinity). Figure 5 quantifies how the different mechanisms affect the clustering properties. The curves represent the theoretical predictions according to eq. (17) for different values of m, π_f, n, and π_r. Note that three of the parameters (m, π_f, and n) do not affect the local clustering coefficient for large values of k. It is important to highlight that the reciprocal response is the only mechanism that impacts the coefficient for nodes with a high degree. In particular, small values of π_r, yield smaller clustering coefficient for nodes with a high degree. The combination of the three mechanism results in a well-defined relationship between nodes with a high degree and their local clustering properties. In particular, as the network grows in size, the clustering coefficient c_i of a node with in-degree k satisfies $c_i \sim dk^{-1}$ for some $d > 0$.



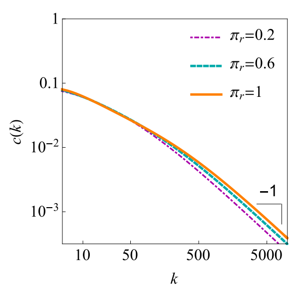
(a)



(b)



(c)



(d)

Fig. 5. Local clustering coefficient; (a) variation in the number of edges formed due to random attachment; (b) variation in the probability of forming edges due to triad formation; (c) variation in the number of edges established through random response; (d) variation in the probability of forming reciprocal edges.

VI. Simulations

To gain further insight into the network formation model, let $N_0 = 10$, $m = 5$, $\pi_f = 0.8$, $n = 2$, and $\pi_r = 0.6$. Figure 6 shows the clustering coefficient at different time instances. The dots represent the simulation of the process and the solid curves indicate the analytical prediction. Note that at $t = 100 \times 10^3$, the clustering coefficient reaches a stationary value. However, depending on the combination of the model parameters, simulations may require $t > 300 \times 10^3$.

VII. Conclusions

This work explains how random attachment and triad-formation, together with network response, impact the dynamics of the structure of growing networks. The proposed model generates networks with an extended power law in-degree, and exponential out-degree distribution, and with a stationary clustering properties (i.e., the local clustering coefficients of the network do not vanish as it grows). We show that the response mechanism, in particular, plays an important role in the formation of clustered networks with high scaling coefficients. Evaluating whether these properties are stable is an important direction for future research.

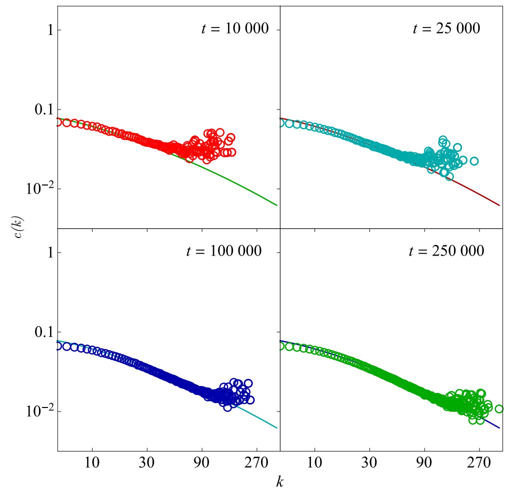


Fig. 6. Local clustering coefficient at different time instances.

References

[1] S. L. Johnson, S. Faraj, and S. Kudaravalli, “Emergence of power laws in online communities: The role of social mechanisms and preferential attachment.” *MIS Quarterly*, vol. 38, no. 3, pp. 795–808, 2014.

[2] N. Durak, G. Tamara, A. Pinar, G. Kolda, and C. Seshadhri, “A scalable directed graph model with reciprocal edges,” in *the IEEE 2nd Workshop on Network Science*, 2013.

[3] T. Lou, J. Tang, J. Hopcroft, Z. Fang, and D. Xiaowen, “Learning to predict reciprocity and triadic closure in social networks.” *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 7, no. 2, pp. 5:1–25, 2013.

[4] R. Toivonen, J.-P. Onnela, J. Saramäki, J. Hyvönen, and K. Kaski, “A model for social networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 371, no. 2, pp. 851 – 860, 2006.

[5] P. Moriano and J. Finke, “On the formation of structure in growing networks,” *Journal of Statistical Mechanisms: Theory and Experiment*, vol. 2013, no. 06, p. 06010, 2013.

[6] O. Hagberg and C. Wiuf, “Convergence properties of the degree distribution of some growing network models,” *Bulletin of Mathematical Biology*, vol. 68, no. 6, pp. 1275–1291, 2006.

[7] M. Newman, “The structure and function of complex networks,” *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.