6

A parametric model approach for structural reconstruction of scale-free networks

Most of the existing network generation models are based on either preferential attachment or random attachment with local information [Dorogovtsev *et al.*, 2000; Chakrabarti and Faloutsos, 2006]. Preferential linking utilizes a global view of the network since preferring a node over the other nodes needs global information of the network. Random attachment with local growth process is inspired by the philosophy that links are not always made by choice but also by chance. For example, making new connections/friendships in society often happens in attending a social gathering (which is by choice) but new links are generated by chance (random) during the meeting. A parametric model is proposed in this chapter using the convex combination of preferential attachment and random attachment with local growth. Here, the preferential attachment is defined by a sequence of independent Bernoulli trials whose success probabilities depend on the degree of the existing nodes. The random attachment is defined by a sequence of independent identical Bernoulli trials. The reconstruction is done by determining the optimal values of the model parameters that minimize the gap between the given degree distribution and the degree distribution of the proposed model network. The efficiency of this method is validated for SNR by reconstructing different real world networks using this approach. We observe that real world networks and its reconstructed copies do share similar statistical properties including distribution of clustering coefficient (CC), hops count, eigenvalues, network values (principal eigenvector of adjacency matrix), and the number of shared triangles. One may wonder why just the minimization of residual error of degree distributions contribute to reducing the residual error of other distributions. Recall that, these distributions are often correlated [Li *et al.*, 2011; Jamakovic and Uhlig, 2008; Hernández and Van Mieghem, 2011], and as a by-product of the given SNR technique and the proposed network model we obtain the desired results. Observing the usefulness of the proposed parametric growing network model into solving SNR, we call the proposed model as Network-Reconstruction-Model (NRM).

We analytically prove that NRM shows power-law in the tail of the degree distribution of the model networks in which the power law exponent ranges from 2 to ∞. We further show that the networks generated by NRM do replicate different established dynamic and structural phenomena during its formation that are observed in real-world networks [Leskovec *et al.*, 2007]. For example, we show that specific choice of model parameters can justify edge-densification, densification power law (DPL), shrinking diameter and modular structure in the corresponding model networks. We also provide an algebraic relation of the model parameters satisfying which NRM can generate networks which follow edge-densification and densification power law. Numerical results establish the property of shrinking diameter and existence of communities for a particular choice of the model parameters.

In order to verify the novelty of NRM in comparison to other existing parametric network models, we consider solving the SNR problem by using DMS model [Dorogovtsev *et al.*, 2000], context dependent preferential attachment model (CDPAM) [Pandey and Adhikari, 2015], forest fire model (FFM) [Leskovec *et al.*, 2007] and community guided attachment (CGA) model [Leskovec *et al.*, 2007]. The results show that performance of both DMS and CDPAM to solving SNR problem

is not satisfactory. FFM and CGA model are able to capture a few properties of the real-world networks for a specific choice of domains of its parameters but have more time complexity compared to NRM.

The chapter is organized as follows. SNR method is explained in Section 6.1 along with runtime and space complexity of the proposed method. In Section 6.2, accuracy (structural) and efficiency (in terms of execution time) of the defined method of structural reconstruction of the networks is compared with other considered network models in the same context. In Section 6.3, various properties of NRM are discussed and numerical simulations of structural properties including clustering coefficient, spectral radius, algebraic connectivity, assortativity, triangle counts are considered. The detail discussion of proposed NRM model and future directions are provided in Section 6.4 and chapter is concluded in Section 6.5.

6.1 NETWORK RECONSTRUCTION MODEL AND NETWORK RECONSTRUCTION METHOD

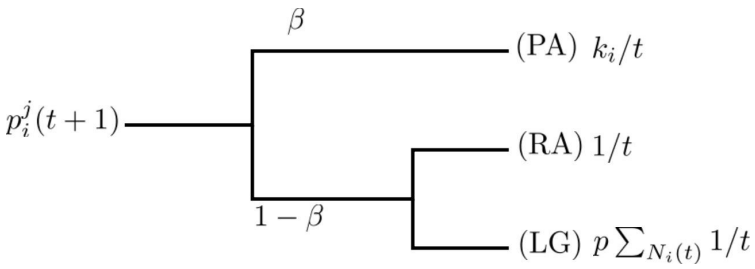


Figure 6.1 : Graphical representation for division of probability of connection between node i and j. In the figure, PA stands for preferential attachment, RA stands for random attachment and LG is local growth.

We propose a 2-parameter growing network generation model which we call NRM as follows. The parameters of the NRM are:

- A probability β, where $0 \leq \beta \leq 1$
- A probability p where $0 \leq p \leq 1/2$

and both are not zero simultaneously. The network formation process starts with a link as an initial network G_0. Then at each time step $t \geq 3$ a new node appears and gets attached to the existing network $G(t)$. For $t = 1,2$ no new nodes and links are added. Thus the number of nodes in $G(t)$ is $t-1$ when $t \geq 3$. For $t \geq 3$ let j be the new node which joins the existing network $G(t)$. Then j forms links with the existing nodes based on the following steps.

- Bernoulli-step: The node j attempts to make a link with an existing node i with success probability $\frac{k_i(t)}{t}$, independently for all i, where $k_i(t)$ denotes the degree of node i in $G(t)$.
- Random attachment (RA) with local growth (LG) step: The node j attempts to form links with success probability $\frac{1}{t}$ to all the existing nodes and if it gets linked with a node i then it forms links with neighbors of i with constant probability p.

Then the NRM $N(\beta, p)$ is defined as follows. Begin with the initial link G_0. For $t \geq 2$, the network $G(t+1)$ is constituted from $G(t)$ by taking the Bernoulli step with probability β, otherwise by taking the random attachment with local growth step. Note that, for a given value of the

pair (β, p), the success probability for formation of a link with an existing node i in the Bernoulli step increases with its degree k_i. Hence, as if, the new coming node j prefers to join an existing node which has the high degree. Thus with a slight abuse of terms, we call the Bernoulli step as preferential attachment (PA) step. The entire process of link formation is depicted in Fig. 6.1.

Thus the probability that the new coming node j forms a link with an existing node i during the formation of $G(t+1)$ is given by

$$p_i^j(t+1) = \beta\,(\mathrm{PA}) + (1-\beta)\,(\mathrm{RA}+\mathrm{LG}),$$

which can be written as

$$p_i^j(t+1) = \beta\frac{k_i(t)}{t} + (1-\beta)\left(\frac{1}{t} + p\sum_{k_i(t)}\frac{1}{t}\right), t \geq 2. \tag{6.1}$$

Given the values of the parameters β and p, a network can be generated by using $p_i^j(t+1)$. We prove that the degree distribution of a network generated by NRM follows power-law in its tail, that is the probability of a node having degree k, which we denote by $P(k)$ is approximately $k^{-\gamma}$ when k is large, where $\gamma = 1 + \frac{1}{\beta+\theta p}$ and $\theta = 1-\beta$. Then degree of a node i after t time steps of its appearance in the network is given by

$$k_i(t) = \left[k_i^0 + \theta(\gamma-1)\right]\left(\frac{t}{t_i}\right)^{\frac{1}{\gamma-1}} - \theta(\gamma-1) \tag{6.2}$$

by Eq. (6.1). Here k_i^0 is the degree of the node i just after it appered at time t_i. Consequently, we obtain

$$p_i^j = \left(\frac{k_i^0}{\gamma-1} + \theta\right)\frac{1}{t_i^{\frac{1}{\gamma-1}} t_j^{1-\frac{1}{\gamma-1}}}. \tag{6.3}$$

The detailed derivation of these results are provided below.

Consider the link formation probability $p_i^j(t+1)$ from Eq. (6.1). Then assuming continuity of $k_i(t)$ with respect to time-step t, we obtain

$$\frac{\partial k_i}{\partial t} = \beta\frac{k_i}{t} + \theta\frac{k_i p + 1}{t} = \frac{1}{\gamma-1}\frac{k_i}{t} + \frac{\theta}{t} \tag{6.4}$$

using mean-field approach. Here $\gamma = 1 + \frac{1}{\beta+\theta p}$, $p \in [0,\ 1/2]$ and $\theta = 1-\beta$. Solving the Eq. (6.4) we obtain

$$k_i(t) = \left[k_i^0 + \theta(\gamma-1)\right]\left(\frac{t}{t_i}\right)^{\frac{1}{\gamma-1}} - \theta(\gamma-1). \tag{6.5}$$

where the initial condition is given by $k_i(t_i) = k_i^0$. This yields the probability of connection between node i and j given in Eq. (6.3)

$$p_j^i = \left(\frac{k_i^0}{\gamma-1} + \theta\right)\frac{1}{t_i^{\frac{1}{\gamma-1}} t_j^{1-\frac{1}{\gamma-1}}}.$$

Now we show that the networks generated by NRM follow power-law with exponent $\gamma = 1 + \frac{1}{\beta+\theta p}$ in the tail of its degree distribution. We adopt a similar approach as described in Chapter 3, [Chung and Lu, 2006]. Recall that the link formation probability of NRM is given by

$$\begin{aligned} p_i^j(t+1) &= \beta \frac{k_i(t)}{t} + (1-\beta)\left(\frac{1}{t} + p\frac{k_i(t)}{t}\right), t \geq 1 \\ &= (\beta + \theta p)\frac{k_i(t)}{t} + \frac{\theta}{t} \end{aligned} \tag{6.6}$$

where $0 \leq \beta \leq 1$, and $0 \leq p \leq 1/2, \theta = 1-\beta$ and β, p are not zero simultaneously. Let $m_{k,t}$ be the number of nodes of degree k in $N(t)$. Then

$$\begin{aligned} E(m_{k,t}|m_{k,t-1}) &= m_{k,t-1}\left(1 - (\beta+\theta p)\frac{k}{t} - \frac{\theta}{t}\right) \\ &\quad + m_{k-1,t-1}\left((\beta+\theta p)\frac{k-1}{t} + \frac{\theta}{t}\right). \end{aligned}$$

Then taking expectation both sides we obtain

$$\begin{aligned} E(m_{k,t}) &= E(m_{k,t-1})\left(1 - (\beta+\theta p)\frac{k}{t} - \frac{\theta}{t}\right) \\ &\quad + E(m_{k-1,t-1})\left((\beta+\theta p)\frac{k-1}{t} + \frac{\theta}{t}\right). \end{aligned} \tag{6.7}$$

Let

$$M_k = \lim_{t\to\infty} \frac{E(m_{k,t})}{t}.$$

Now we recall the following lemma from [Chung and Lu, 2006].

Lemma 6.1.1. *Suppose that a sequence $\{a_t\}$ satisfies the recurrence relation*

$$a_{t+1} = \left(1 - \frac{b_t}{t+t_1}\right)a_t + c_t \text{ for } t \geq t_0. \tag{6.8}$$

Furthermore, suppose $\lim_{t\to\infty} b_t = b > 0$ and $\lim_{t\to\infty} c_t = c$. Then $\lim_{t\to\infty} \frac{a_t}{t}$ exists and

$$\lim_{t\to\infty} \frac{a_t}{t} = \frac{c}{1+b}. \tag{6.9}$$

Consider $b_t = (\beta+\theta p)k + \theta, t_1 = 0, t_0 = 1$ and

$$c_t = \frac{E(m_{k-1,t-1})}{t}[(\beta+\theta p)(k-1) + \theta].$$

Then $\lim_{t\to\infty} b_t = (\beta+\theta p)k + \theta$ and

$$\lim_{t\to\infty} c_t = M_{k-1}[(\beta+\theta p)(k-1) + \theta].$$

Hence from Eq. (6.7)

$$M_k = M_{k-1}\frac{[(\beta+\theta p)(k-1) + \theta}{(\beta+\theta p)k + 1 + \theta}. \tag{6.10}$$

Now if $M_k = k^{-\gamma}$ for some $\gamma > 0$ then

$$\frac{M_k}{M_{k-1}} = \frac{k^{-\gamma}}{(k-1)^{-\gamma}} = 1 - \frac{\gamma}{k} + O\frac{1}{k^2}. \tag{6.11}$$

From Eq. (6.10),

$$\begin{aligned}
\frac{M_k}{M_{k-1}} &= \frac{[(\beta+\theta p)(k-1)+\theta}{(\beta+\theta p)k+1+\theta} \\
&= \frac{(\beta+\theta p)k-(\beta+\theta p)+\theta}{(\beta+\theta p)k+1+\theta} \\
&= \frac{(\beta+\theta p)k+1-1-(\beta+\theta p)+\theta}{(\beta+\theta p)k+1+\theta} \\
&= 1-\frac{1+(\beta+\theta p)}{(\beta+\theta p)k+1+\theta} \\
&= 1-\frac{(\beta+\theta p)+1}{(\beta+\theta p)k+1+\theta} \\
&= 1-\frac{1+\frac{1}{\beta+\theta p}}{k+\frac{1+\theta}{\beta+\theta p}} \\
&= 1-\frac{1+\frac{1}{\beta+\theta p}}{k}+O\frac{1}{k^2}.
\end{aligned} \tag{6.12}$$

Thus comparing Eq. (6.11) and Eq. (6.12) we obtain

$$\gamma = 1 + \frac{1}{\beta+\theta p}.$$

Now it remains to show that $0 < M_k < \infty$ for $k \geq 2$. Let us rewrite the Eq. (6.6) as

$$p_i^j(t+1) = \left(\frac{1}{\gamma-1}\right)\frac{k_i(t)}{t} + \left(1-\frac{1}{\gamma-1}\right)r\frac{2}{t}, \tag{6.13}$$

where $r = \dfrac{1}{2(1-p)} \leq 1$ since $0 \leq p \leq 1/2$. Then the link formation in NRM can be described as follows. From Eq. (6.13), note that after $t > 2$ a new coming node either goes for preferential attachment (PA) with probability $\left(\frac{1}{\gamma-1}\right)$ or for random attachment (RA) with probability $\left(1-\frac{1}{\gamma-1}\right)$. Random attachment to two nodes in the network is restricted by probability r. Simply, the probability of getting two links under random attachment by newly appeared node is $\left(1-\frac{1}{\gamma-1}\right)r$. The graphical representation of the process is given in Fig. 6.2. Newly added node can also generate two links under preferential attachment scheme with some probability ≥ 0. So the probability that a newly appeared node has degree two is $\geq \left(1-\frac{1}{\gamma-1}\right)r$.

Thus for $k = 2$, we have $b_t = b = (1+\beta+2\theta p)$ and $c_t = c \geq \left(1-\frac{1}{\gamma-1}\right)r$. Then by Lemma 6.1.1 $\lim_{t\to\infty} \mathrm{E}(m_{2,t})/t = M_2$ exists and

$$M_2 = \lim_{t\to\infty}\frac{\mathrm{E}(m_{2,t})}{t} \geq \frac{\left(1-\frac{1}{\gamma-1}\right)r}{2+\beta+2\theta p}.$$

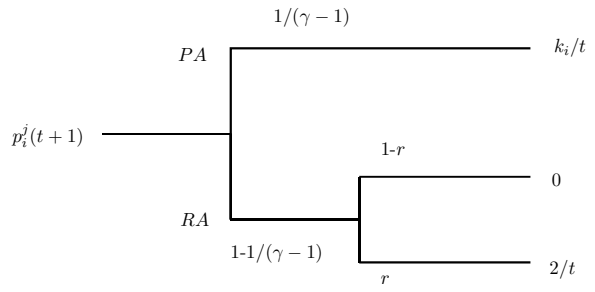


Figure 6.2 : Graphical representation of Eq. (6.13).

Thus $0 \neq M_k < \infty$ for all $k \geq 2$ by Eq. (6.10). This completes the proof.

Various structural and statistical properties of networks generated by NRM are investigated in Section 6.3. The SNR problem for a given real network can be solved by determining the values of β and p by using the degree sequence or degree distribution of the given scale-free network as follows.

6.1.1 Determination of the optimal values of β and p for network reconstruction

Now we describe a method for the determination of the optimal values of the parameters β, p that can produce a model network using NRM for reconstruction of a given scale-free network. Since any choice of $\beta \in [0,1]$ and $p \in [0,1/2]$ generates a model network with power-law in its degree distribution, we discritize both $[0,1]$ and $[0,1/2]$ into m_1 and m_2 equi-distant points, respectively. Let $S_1 \subset [0,1]$ and $S_2 \subset [0,1/2]$ denote the set of points obtained after discritization such that $|S_1| = m_1, |S_2| = m_2$. Then for each $(\beta, p) \in S_1 \times S_2$ the NRM produces a network having power-law degree distribution with exponent $\gamma = 1 + 1/(\beta + (1-\beta)p)$ in the tail of its degree distribution and there can be $m_1 m_2$ model networks.

Let $G(t)$ be the model network which contains the same number of nodes as the number of nodes in a given real network G_r. Then the optimal values of β, p is obtained by

$$(\beta_{\text{opt}}, p_{\text{opt}}) = \min_{(\beta,p)\in S_1\times S_2} \sum_{k=1}^{k_{max}} |P(k_i(t) < k) - P_r(k_i < k)|. \tag{6.14}$$

where $P(k_i(t) < k) = 1 - P(k_i(t) \geq k)$, the probability that a node has degree k_i smaller than k is calculated for the network generated by NRM for a given $\beta \in S_1, p \in S_2$; $P_r(k_i < k) = 1 - P_r(k_i \geq k)$ is calculated from G_r, and k_{max} denotes the maximum degree of nodes in G. Note that the time complexity in finding the values of $\beta_{\text{opt}}, p_{\text{opt}}$ rises if $m_1 m_2$ is a very large number as we need to generate $m_1 m_2$ model networks in order to obtain the desired result.

However the time complexity can drastically be reduced by utilizing the power-law exponent γ of the model network. Let $\gamma_r > 1$ be the power-law exponent of the given network G_r that can be obtained by using gradient descent technique [Kutner *et al.*, 2004] from the degree sequence of the given network. Setting

$$\gamma_r = 1 + \frac{1}{\beta + (1-\beta)p} \tag{6.15}$$

a relation between the parameters β, p can be obtained, and hence the search space for the optimal values of parameters (β, p) is reduced from $O(m_1 m_2)$ to linear search space of $O(m_1)$ or $O(m_2)$.

Table 6.1 : Real world networks.

Network	Size (n)	Edges
e-mail network	1133	5451
PPI network of YEAST	2361	6609
Collaboration network (CbN)	4158	13422
Collaboration network (ca-HepTh)	8638	24806
PGP network	10680	24316

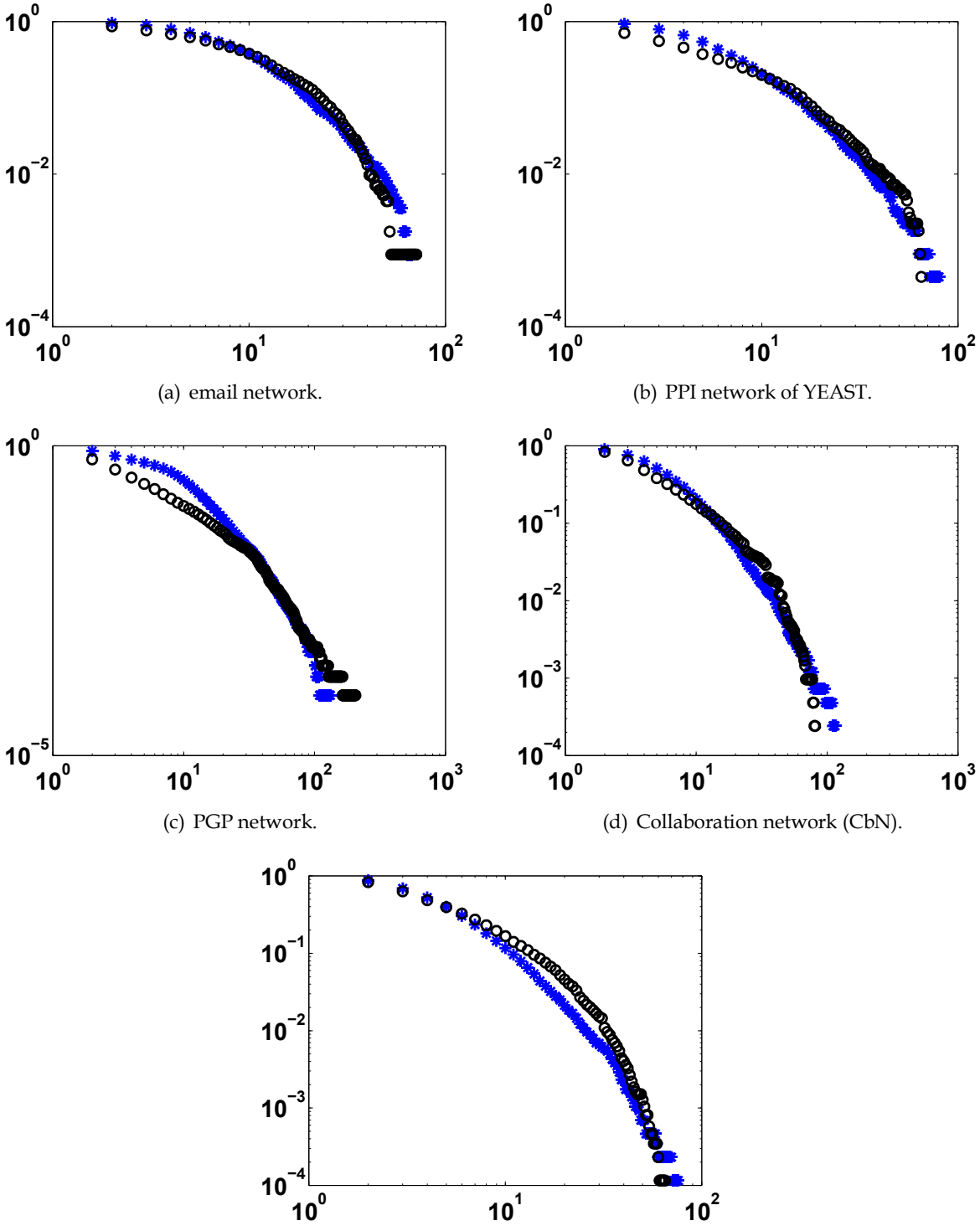
Let n be the number of nodes in $G(t)$. Then the degree of each node in $G(t)$ can be calculated at time t in constant run time using Eq. (6.2). Thus it needs $O(n)$ space to store the degree distribution. Determination of p_{opt} and β_{opt} requires approximately $O(2n)$ space.

We mention that due to the lack of a functional relationship between the power-law exponent and model parameters in existing parametric models, the structural reconstruction of a given network by using existing models is computationally challenging. For example, how the power-law exponent in the degree distribution of networks in CGA model is related to the model parameters i.e. difficulty constant (> 0) and branching factor (> 0) is not known. Same is true for FFM in which the model parameters are known as forward burning probability ($\in (0,1)$) and the backward burning ratio (> 0). Besides, determination of values of the parameters involved in CGA and FFM is difficult as the corresponding domain spaces are unbounded. Hence one has to be careful while restricting a domain to be bounded for reconstruction of a given network using these models.

We determine the pair $(\beta_{\text{opt}}, p_{\text{opt}})$ after discretizing the domains of β, p into $m_1 = 100 = m_2$ points to generate model networks using NRM for certain real world networks which include email network [Guimerà *et al.*, 2003], PPI network of YEAST [Bu *et al.*, 2003], PGP network [Boguñá *et al.*, 2004], Collaboration Network (CbN) [Leskovec *et al.*, 2007] and the Collaboration network (ca-Hep Th) [Leskovec *et al.*, 2007]. Note that the degree distribution of each of these networks follows power law in the tail as investigated in [Boguñá *et al.*, 2004; Clauset *et al.*, 2009b; Barabási and Albert, 1999b; Albert and Barabási, 2002], see Table 6.1. The datasets of these networks can be found in [Leskovec and Krevl, 2014]. The degree distribution (in $\log - \log$ scale) of each of these real world networks and the same for respective model networks obtained by NRM using $(\beta_{\text{opt}}, p_{\text{opt}})$ are plotted in Fig. 6.3. The detailed analysis of the reconstruction capability of NRM and comparison with other relevant models is provided in next section.

6.1.2 Limitation of NRM

The power-law exponent γ in the networks generated by NRM ranges from 2 to ∞. Hence NRM is not applicable for reconstruction of scale-free networks having power-law exponent less than 2. The optimal values of β and p are obtained by discretizing the domains of these parameters and thus β_{opt} and p_{opt} need not be the global optimal values. For a given pair (β, p), the NRM produces networks having a constant γ throughout the evolution of the network. Therefore it does not capture the class of networks which have time varying power-law exponent.



(a) email network.

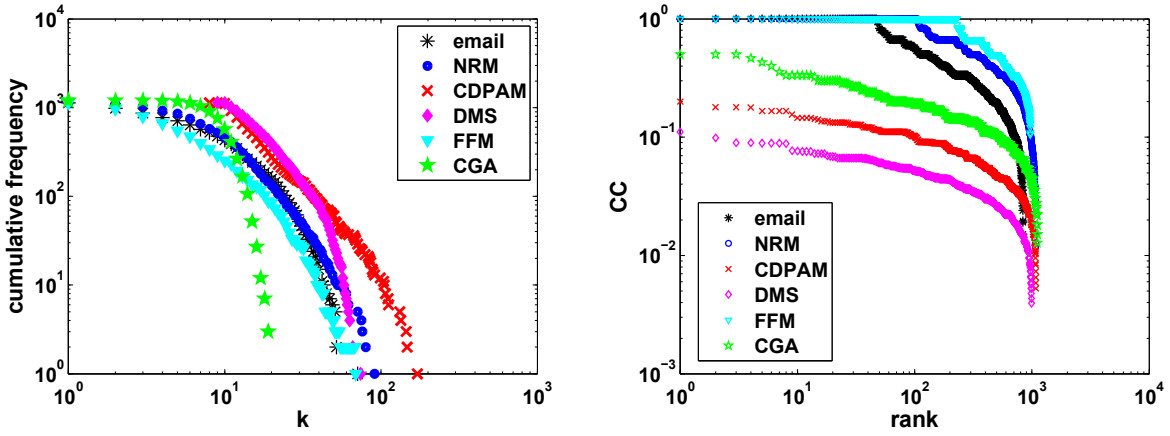
(b) PPI network of YEAST.

(c) PGP network.

(d) Collaboration network (CbN).

(e) Collaboration network (ca-HepTh).

Figure 6.3 : Matching of degree distribution of real world networks and the same produced by the NRM. The degree distributions of different networks generated by NRM are plotted (in $\log - \log$ scale) in blue and the same of the corresponding real-world networks are plotted in black. The horizontal axis represents degree (k) and vertical axis represents cumulative probability $P(k_i \geq k)$.



(a) Degree distribution.

(b) Distribution of clustering coefficient of nodes.

(c) Distribution Triangles.

(d) Distribution of hops.

Figure 6.4 : Numerical results of different properties of network generated by models corresponding to email network are plotted.

Table 6.2 : Dis-similarity between degree distributions.

Models / Networks	email	CbN	ca-HepTh	PGP	YEAST
NRM	**0.04**	**0.04**	**0.0045**	**0.067**	**0.073**
CDPAM	0.20	0.27	0.24	0.28	0.26
DMS	0.20	0.27	0.01	0.28	0.25
FFM	0.07	0.10	0.0065	0.12	0.13
CGA	0.15	0.13	0.105	0.12	0.15

6.2 COMPARISON OF NRM WITH CDPAM, DMS, FFM AND CGA MODEL

Now we compare the performance of the proposed NRM with the existing parametric models FFM, CGA model, DMS, CDPAM as follows. We consider the domains of the forward burning probability and backward burning ratio in FFM as $(0, 1)$ and $(0, 10)$ respectively for computing a model network using FFM. In CGA model we consider the branching factor to lie in $(2, 5)$ and the difficulty constant belongs to $(1, 10)$. For DMS, we consider the domain of the model parameter 'initial attractiveness' to lie in $(1, 10^5)$. Finally, the model control parameter β in CDPAM is assumed to lie in $(0.6, 10^5)$. We discretize each of these domains for all the models into 100 points. Model networks are generated using each of these values and the optimal values of the parameters are determined that minimize $|P(k_i(t) < k) - P_r(k_i < k)|$ (see Eq. (6.14)).

For NRM, as mentioned above, we discretize the domains of $\beta \in [0, 1]$ and $p \in [0, 1/2]$ into 100 points. Recall that due to the functional relation between the model parameters in NRM, the running complexity for the generation of a model network is equivalent to the single parameter models, when the power-law exponent of the given scale-free network is incorporated. The optimal values of β, p are obtained as defined in Eq. (6.14).

We call the model network generated by the optimal values of the parameters for a given real network as an optimal model network.

6.2.1 Comparison of NRM, DMS, CDPAM, FFM and CGA using Jensen-Shannon Index

The Jensen-Shannon index [Lin, 1991] measures the similarity between probability distributions of two different systems defined by [Boas *et al.*, 2010]

$$J(p,q) = \frac{1}{2}\left(D(p,m) + D(q,m)\right),$$

where $m = (p+q)/2$, p and q are the probability mass functions of the two different systems. $D(p,m)$ is defined by

$$D(p,m) = \sum_x p(x) \log \frac{p(x)}{m}$$

which gives the distance in bits between two distributions p and m.

We employ the Jensen-Shannon index to calculate the similarity of degree distributions of the optimal model network generated by a given model and the corresponding real-world network. The corresponding dissimilarity indices are provided in Table 6.2 for certain real networks discussed above. The table establishes that NRM is better in the context of degree distribution for reconstruction of real world networks compared to the other models which we consider in this chapter.

Table 6.3 : Average error in different structural properties of the networks generated by different models corresponding to the given networks. It is averaged over error between 10 ensembles of generated model networks and real world network.

Network	Model	E_{DD}	$E_{\overline{CC}}$	E_{ATS}	E_{AH_s}	E_{AC}	E_{SV}	E_{NV}
email	NRM	**1.76**	**0.14**	3.65	**0.10**	**0.19**	**0.28**	**0.002**
	CDPAM	9.74	0.17	**1.71**	1.01	6.82	1.26	0.008
	DMS	9.80	0.19	12.64	0.92	6.71	1.46	0.007
	FFM	2.59	0.26	3.64	0.91	0.22	0.32	0.031
	CGA	4.86	0.15	20.11	0.28	0.85	0.81	0.028
YEAST	NRM	**1.89**	0.22	**4.84**	**0.13**	**0.02**	0.55	**0.003**
	CDPAM	7.66	0.17	8.57	1.43	4.50	1.33	0.008
	DMS	7.70	0.18	11.70	1.28	4.43	1.47	0.009
	FFM	2.02	0.32	5.32	1.05	0.05	**0.37**	0.012
	CGA	3.66	**0.14**	14.23	0.60	0.06	0.80	0.018
ca-HepTh	NRM	**0.71**	**0.017**	**1.92**	1.27	**0.008**	**0.15**	**0.002**
	CDPAM	6.21	0.47	17.10	2.52	3.65	0.93	0.006
	DMS	6.22	0.48	18.52	2.23	3.58	1.12	0.007
	FFM	1.62	0.021	5.06	0.93	0.03	0.24	0.004
	CGA	2.73	0.43	17.87	**0.03**	0.024	0.51	0.009
CbN	NRM	**1.56**	0.19	**54.01**	**1.32**	**0.026**	**0.44**	**0.006**
	CDPAM	7.43	0.54	63.74	2.95	4.45	1.24	0.010
	DMS	7.45	0.55	66.16	2.74	4.42	1.41	0.010
	FFM	2.06	**0.16**	63.44	1.76	0.029	0.57	0.008
	CGA	3.69	0.49	66.58	1.83	0.035	0.67	0.013
PGP	NRM	**3.30**	**0.10**	**15.22**	2.48	**0.044**	**0.75**	**0.003**
	CDPAM	5.42	0.26	29.56	3.86	2.83	1.05	0.005
	DMS	5.49	0.26	30.31	3.47	2.73	1.26	0.005
	FFM	8.10	0.15	17.24	3.73	0.30	1.20	0.006
	CGA	3.89	0.23	30.01	**0.50**	0.063	0.61	0.004

6.2.2 Comparison of NRM, DMS, CDPAM, FFM and CGA on the basis of reconstruction of structural and spectral properties

In this section, we introduce error functions which measure the amount of deviation of certain structural and spectral properties of the optimal model network and a given real network. Thus we compare the potential of the network generative models described above to resolve the SNR problem for a given real network.

We define the error in degree distribution as

$$E_{DD} = \sum_{k=1}^{k_{max}} |P_M(k_i < k) - P_r(k_i < k)|.$$

where $P_r(k_i < k)$ is the degree distribution of the given real-world network and $P_M(k_i < k)$ is the same of the optimal model network. Similarly, error in clustering coefficient of the networks is defined as

$$E_{\overline{CC}} = |\overline{CC}_r - \overline{CC}_M|,$$

where $\overline{CC}_r$ is the average clustering coefficient of the real world network and $\overline{CC}_M$ is the average clustering coefficient of the optimal model network. Error in the average number of triangles shared by a node is defined as

$$E_{ATS} = \frac{1}{n}|\text{Trace}(A_r^3) - \text{Trace}(A_M^3)|$$

where A_r is the adjacency matrix of the real world network and A_M is the same for the optimal model network of the same size. Error in average hops between a pair of nodes E_{AH_s} is defined as the error in average path length of the real network and the optimal model network. Error in algebraic connectivity is defined as

$$E_{AC} = |\lambda_2(L_r) - \lambda_2(L_{M)})|.$$

where $\lambda_2(L_r)$ and $\lambda_2(L_M)$ are the second smallest eigenvalues of the combinatorial Laplacian matrices associated with the real world network and the optimal model network respectively.

Finally we define errors in the singular values (SV) and the values of principal eigenvector (also called network values (NV)) of the adjacency matrices as

$$E_{SV} = \frac{1}{n}\sum_{k=1}^{n}|SV_k(A_r) - SV_k(A_M)|.$$

and

$$E_{NV} = \frac{1}{n}\sum_{k=1}^{n}|NV_k(A_r) - NV_k(A_M)|$$

respectively.

The errors, thus calculated, are given in Table 6.3 after determining the optimal model network for each of the network generation models. Observe that the error is minimum for NRM in almost all the cases. Thus we deduce that NRM has the capability to capture multiple structural and spectral properties simultaneously for a given scale-free network better than the other models. Besides, FFM and CGA models perform marginally better than DMS and CDPAM.

Here we mention that algebraic connectivity is small for most of the real networks and it is an indication of modular structure in the network [Jamakovic and Uhlig, 2008; Newman, 2004, 2006a]. From Table 6.3 observe that the optimal model networks generated by NRM have lower algebraic connectivity which is close to the algebraic connectivity of the corresponding real networks. Hence NRM has more potential compared to other models to inherit modular structure in the reconstructed copies of real networks. Moreover, the spectral properties such as singular values, network values of the optimal model networks generated by NRM are close to that of the real networks. This means that the rate of convergence of different diffusion protocols on these model networks should be same as that of the real networks [Newman, 2010].

In Fig. 6.4 we plot the distribution of clustering coefficient, degree distribution, distribution of sharing of triangles and hop's distribution of the reconstructed copies of email network and that of the email network. Observe that the optimal model network generated by NRM outperforms all the other optimal model networks. Hence these numerical results confirm that NRM can solve the SNR problem efficiently than the other models for the choice of domains of its parameters considered in this chapter.

Table 6.4 : approximated execution time (in sec) of different methods adopted for network reconstruction.

Models/ Network	email	Yeast	CbN	PGP	ca-HepTh
NRM	**13**	**43**	**138**	**610**	**538**
CDPAM	139	347	1256	5789	5392
DMS	156	423	1163	7143	5643
FFM	32132	78264	123462	892335	297436
CGA	213	443	1326	8976	7364

6.2.3 Comparison of NRM, DMS, CDPAM, FFM and CGA on the basis of execution time

We evaluate the execution time for solving the SNR problems for the real networks considered in this chapter, by using CDPAM, DMS model, FFM, CGA model and NRM. The optimal model network is produced by discretizing all the domains of parameters discussed before into 100 equidistant points. For NRM, the relationship between the model parameters is used while incorporating the power-law exponent of a real network (see Eq. 6.15). The execution time for each of these models for the corresponding real network is calculated in seconds and presented in Table 6.4. All the computation is in a machine of configuration 4 GB RAM, Intel(R) Core(TM) $i5-3230$M CPU 2.6 GHz, 64 bit OS. The execution times in Table 6.4 confirm that the one parameter models and the CGA model have execution time almost 10 times more than NRM while FFM has almost 10^4 times more execution time compared to NRM. Thus we conclude that NRM is an efficient model to resolve the SNR problem.

6.3 PROPERTIES OF NETWORK RECONSTRUCTION MODEL

In this section, we discuss different statistical properties of a network generated by NRM. Note that NRM is similar to the preferential attachment scheme as described in [Chapter 3, [Chung and Lu, 2006]]. Following a similar derivation as provided in [Chapter 3, [Chung and Lu, 2006]], it is easy to prove that the degree distribution of a network generated by NRM exhibits power law in their tail given by $P(k) \propto k^{-\gamma}$ where $\gamma = 1 + \frac{1}{\beta+(1-\beta)p}$. Now we derive computable expressions of the expected diameter, expected number of triangles and expected clustering coefficient of a node in a model network generated by NRM. We also provide an algebraic relation for the model parameters β and p such that any chosen values for these parameters that satisfy the algebraic relation would produce networks which follow edge-densification and densification power law.

6.3.1 Expected Diameter

The expected diameter of a random network is defined as the maximum of average path lengths between any pair of nodes [Fronczak *et al.*, 2004]. Recall that, if $A_1, A_2, ...A_n$ are mutually independent events and their probabilities fulfil the conditions $P(A_i) \leq \varepsilon$ for all i then

$$P(\cup_{i=1}^{n} A_i) = 1 - \exp\left(-\sum_{i=1}^{n} P(A_i)\right) - Q, \tag{6.16}$$

where $0 \leq Q < \sum_{j=0}^{n+1} (n\varepsilon)^j / j! - (1+\varepsilon)^n$ [Fronczak *et al.*, 2004]. In order to compute the expected diameter of a network generated by NRM, we consider $A_i \in \mathscr{A}$ to be an event which assumes the existence of a walk $(i, v_1, \ldots, v_{l-1}, j)$ of length l between two nodes i and j in $G(t)$. The set $\mathscr{A}$ consists of all such events and hence $|\mathscr{A}| = |V(t)|^{l-1}$ where $|V(t)|$ denotes the number of nodes in

$G(t)$ [Fronczak *et al.*, 2004]. Note that a walk is called a path if $v_i \neq v_j, 1 \leq i, j \leq l-1$. Then the probability of the existence of a path of length not more than l between any two nodes i, j in $G(t)$ is given by

$$\begin{aligned} P_{ij}(l) &\approx P\left(\cup_{A_i \in \mathscr{A}} A_i\right) \\ &\approx 1 - \exp\left(-\sum_{v_1=1}^{|V(t)|} \cdots \sum_{v_{l-1}=1}^{|V(t)|} p_i^{v_1} \cdots p_{v_{l-1}}^{j}\right) \end{aligned} \tag{6.17}$$

by using Eq. (6.16) and p_x^y denotes the probability of having a link between the nodes x and y. Consequently,

$$P_{ij}(l) = 1 - \exp\left(-\left(\frac{k_i^0}{\gamma - 1} + \theta\right) \frac{h_n^{l-1}}{t_i^{\frac{1}{\gamma-1}} t_j^{1-\frac{1}{\gamma-1}}}\right)$$

by using Eqs. (6.3) and (6.17), where $h_n = \sum_{i=1}^{|V(t)|} \frac{\left(\frac{k_i^0}{\gamma-1}+\theta\right)}{i}$. Thus, the expected length of a shortest path between two nodes $i, j \in V(t)$ is given by

$$l_{ij} = \frac{\frac{1}{\gamma-1}\ln t_i + \left(1 - \frac{1}{\gamma-1}\right)\ln t_j - \ln\left(\frac{k_i^0}{\gamma-1} + \theta\right) - \mathscr{E}}{\ln h_n} + \frac{3}{2} \tag{6.18}$$

where $l_{ij} = \sum_{l=0}^{\infty} F(l)$, $F(l) = 1 - P_{ij}(l)$ and $\mathscr{E} = 0.5772$ is the Euler's constant [Fronczak *et al.*, 2004]. Moreover, it follows from Eq. (6.18) that l_{ij} is an increasing function of t_i and t_j when other parameters are fixed. This observation indicates that the expected diameter of the network generated by NRM is the expected length of a shortest path between the first node and the last node added in the network. Hence, setting $t_j = |V(t)|$ and $t_i = 1$, the expected diameter of a complex network generated by NRM is given by

$$D_G = \frac{\left(1 - \frac{1}{\gamma-1}\right)\ln|V(t)| - \mathscr{E}}{\ln h_n} + \frac{3}{2}.$$

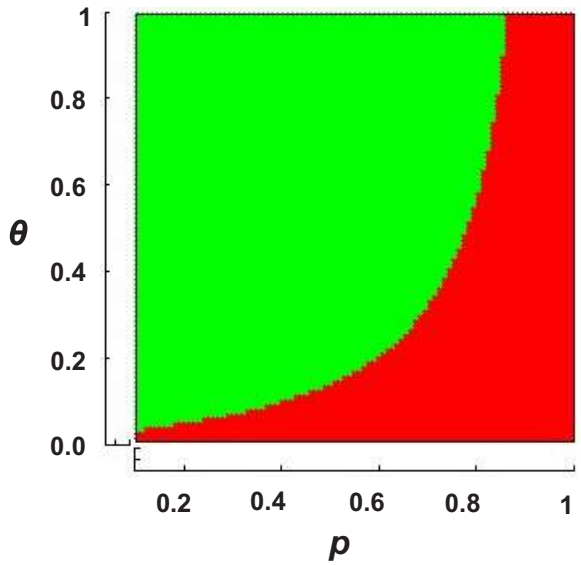


Figure 6.5 : (p, θ) in green region shows shrinking in expected diameter and in red region first expected diameter grows then stabilizes. $k_i^0 = 2 \ \forall \ i < 100$.

In Fig. 6.5 a parameter grid is plotted that shows the behavior of expected diameter of NRM for given range of values. The combination of parameter values in the green region produces stabilization of diameter after shrinking and red region shows the stabilization of diameter with very slow growth. Note that here we allowed p to lie in $[0, 1]$ and recall that the NRM produces a network whose degree distribution follows power-law in the tail only when $0 \leq p \leq 1/2$.

6.3.2 Expected Number of Triangles and Clustering Coefficient

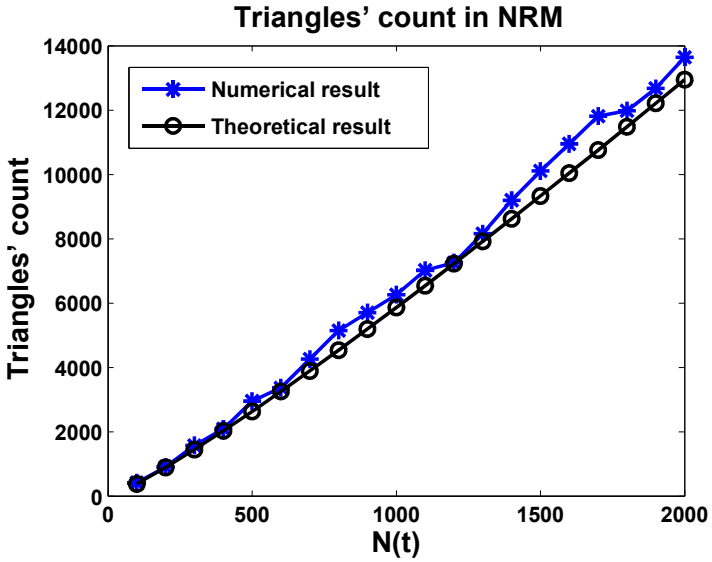


Figure 6.6 : Comparison between numerical and theoretical results for triangle counts in a network generated by NRM at $p = 0.4, \beta = 0.2$ is plotted.

Note that, in NRM, a new coming node can get attached with previously existing nodes in two ways: either it goes for preferential attachment with probability β or it chooses random attachment having local growth with probability $1-\beta$. Thus, the expected number of triangles $\triangle_G$ in a network G generated by NRM is given by

$$\begin{aligned}
\triangle_G &= \sum_{i \in G(t)} \triangle_i^0, \text{ where} \\
\triangle_{i+1}^0 &= (1-\beta)\left(\sum_{i_1=1}^{i} \frac{p(k_{i_1} + p\triangle_{i_1}(t_i))}{t_i}\right) \\
&\quad + \beta \sum_{i_1, i_2 = 1, i_1 < i_2}^{i} \frac{k_{i_1}(t_i)}{t_i} \frac{k_{i_2}(t_i)}{t_i} p_{i_1}^{i_2}
\end{aligned}$$

and hence

$$\begin{aligned}
\triangle_{i+1}^0 &= (1-\beta)p\left(\bar{k}_{t_i} + \sum_{i_1=1}^{i} \frac{p\triangle_{i_1}(t_i)}{t_i}\right) \\
&\quad + \sum_{i_1, i_2 = 1, i_1 < i_2}^{i} \frac{\beta\left(k_{i_1}^0 + \theta(\gamma-1)\right)^2 \left(k_{i_2}^0 + \theta(\gamma-1)\right)}{t_i^{2-\frac{2}{\gamma-1}} (\gamma-1) t_{i_1}^{\frac{2}{\gamma-1}} t_{i_2}}
\end{aligned} \tag{6.19}$$

where $\triangle_i^0$ is the expected growth of triangles in the network when a node i is added to the network given that $\triangle_1^0, \triangle_2^0 = 0$.

In the right had side of the Eq. (6.19), the first term represents the triangles generated by random attachment with local growth and second term counts the same for preferential attachment. In random attachment phenomena, let a new node $i+1$ get attached with node i_1 having degree k_{i_1} at time t_i then the node $i+1$ creates approximately $\frac{pk_{i_1}}{t_i}$ triangles by connecting the neighbours of node i_1 and i_1 is also a part of these triangles. Next let at time t_i, node i_1 be attached with $\triangle_{i_1}(t_i)$ triangles. Then the node $i+1$ will get attached to these triangles with probability p^2. Thus $i+1$ gets $\frac{p^2\triangle_{i_1}(t_i)}{t_i}$ expected triangles and hence $i+1$ gets $\frac{p(k_{i_1}+p\triangle_{i_1}(t_i))}{t_i}$ total expected triangles from the node i_1.

Observe that when a node i gets attached in the network at time step t_i, it generates $\triangle_i^0$ triangles which also increases the count of triangles formed by the existing nodes. Thus the growth of triangles at a particular node is given by

$$\begin{aligned}\triangle_i(t+1) = \triangle_i(t)\left(1+\frac{2\theta p^2}{t}\right) + 2\theta p\frac{k_i(t)}{t} \\ + \beta\sum_{i_1>i}\frac{k_i(t)}{t}\frac{k_{i_1}(t)}{t}p_i^{i_1}.\end{aligned} \tag{6.20}$$

The number of triangles attached to a node increases with the time as the network grows. When a new node appears at time t and selects random attachment with local growth step to form links with the existing nodes, the node i gets $\theta p\frac{k_i(t)}{t}$ expected triangles. Secondly, the new node can also select a neighbour of node i with probability $\frac{1}{t}$ so the probability that the new node will connect with the neighbours of node i first, is $\frac{k_i(t)}{t}$ then by probability p, i will be selected. This way the node i gets $\theta p\frac{k_i(t)}{t}$ expected triangles. Thirdly, let (i,j,k) be a triangle. The new node can connect with j (k) first with probability $\frac{1}{t}$ then with nodes i and k (j) with probability p. Let there be $\triangle_i(t)$ triangles attached to node i at time t. This increases $\frac{2\theta\triangle_i(t)p^2}{t}$ expected triangles of node i. Next, the growth of triangles at a node also depends on preferential attachment that is given by last term of Eq. (6.20). Finally, the overall growth of triangles attached to node i is the summation of all cases described above and given by Eq. (6.20).

The expected number of triangles calculated theoretically and numerically calculated number of triangles in a network generated by NRM are plotted in Fig. 6.6. The numerical count of triangles is averaged over 100 ensembles of growing networks generated by NRM.

Consequently, the expected clustering coefficient of a node in the network generated by NRM is given by

$$c_i(t) = \frac{\triangle_i(t)}{k_i(t)\left(k_i(t)-1\right)}.$$

where $c_i^0 = \frac{\triangle_i^0}{k_i^0\left(k_i^0-1\right)}$ and

$$k_i^0 = \sum_{f=1}^{i-1} p_f^i = \sum_{f=1}^{i-1}\left(\frac{k_f^0}{\gamma-1}+\theta\right)\frac{1}{t_f^{\frac{1}{\gamma-1}}t_i^{1-\frac{1}{\gamma-1}}}.$$

6.3.3 Edge densification and densification power law

Edge densification or increasing average degree of a growing network is an interesting phenomena which is observed in many real networks [Leskovec *et al.*, 2007]. We show that the same is true for networks generated by NRM when the model parameters β, p satisfy an algebraic equation. Let $\bar{k}_t$ be the expected average degree of a network N_t generated by NRM. Assume that

a new node i is added to the network at time-step $t+1$ such that the degree of node i becomes $k_i^0 = \bar{k}_t(\beta+\theta p)+\theta$. Then the new expected average degree of the network is given by

$$\bar{k}_{t+1} = \frac{t\bar{k}_t + 2\bar{k}_t(\beta+\theta p)+2\theta}{t+1}. \tag{6.21}$$

followed by Eq. (6.1). Thus,

$$\bar{k}_{t+1} - \bar{k}_t = \frac{\frac{2\bar{k}_t}{\gamma-1}+2\theta-\bar{k}_t}{t+1}. \tag{6.22}$$

since $(\beta+\theta p) = \frac{1}{\gamma-1}$ (see Eq. (6.15)). Hence, a network generated by NRM exhibits edge densification if

$$\frac{2\bar{k}_t}{\gamma-1}+2\theta-\bar{k}_t > 0 \Rightarrow \bar{k}_t\left(\frac{3-\gamma}{\gamma-1}\right)+2\theta > 0.$$

Thus, when $\gamma \leq 3$, that is, $\beta+(1-\beta)p \geq \frac{1}{2}$, the network exhibits edge densification. Nevertheless, when $\gamma > 3$, that is, $\beta+(1-\beta)p < \frac{1}{2}$, the edge densification happens up to time-step t if

$$\bar{k}_t < \frac{2\theta(\gamma-1)}{\gamma-3} = \frac{2\theta}{1-2(\beta+(1-\beta)p)}.$$

Further, converting Eq. (6.22) into a differential equation by mean field approximation, we have

$$\frac{\partial\bar{k}_t}{\partial t} = \frac{\bar{k}_t\left(\frac{2}{\gamma-1}-1\right)+2\theta}{t+1}. \tag{6.23}$$

Considering the initial condition as $\bar{k}(0)=1$ and solving Eq. (6.23), we obtain

$$\bar{k}_t = \left(1+\frac{2\theta}{\delta}\right)(t+1)^{\delta} - \frac{2\theta}{\delta}.$$

where $\delta = 2(\beta+(1-\beta)p)-1$.

Let $e(t)$ be the number of edges in $G(t)$. Then we have $2e(t) = t\bar{k}_t + 2$ which yields

$$e(t) \approx t\left(t^{\delta}\left(\frac{1}{2}+\frac{\theta}{\delta}\right)\left(1+\frac{1}{t}\right)^{\delta} - \frac{\theta}{\delta}\right)+1 \tag{6.24}$$

when $\delta = 2(\beta+(1-\beta)p)-1 \neq 0$. Therefore, if $2(\beta+(1-\beta)p) \neq 1$, a network generated by NRM exhibits densification of power law given by

$$e(t) \approx \left(\frac{1}{2}+\frac{\theta}{\delta}\right)t^{\delta+1}. \tag{6.25}$$

when the time-step t is sufficiently large.

The expected number of edges calculated theoretically and the numerically calculated number of edges in a network generated by NRM are plotted in log-log scale in Fig. 6.7. The numerical count of edges is averaged over 100 ensembles of growing networks generated by NRM. Hence we establish that a growing network $G(t)$ generated by NRM inherits the phenomena of edge densification when t is sufficiently large.

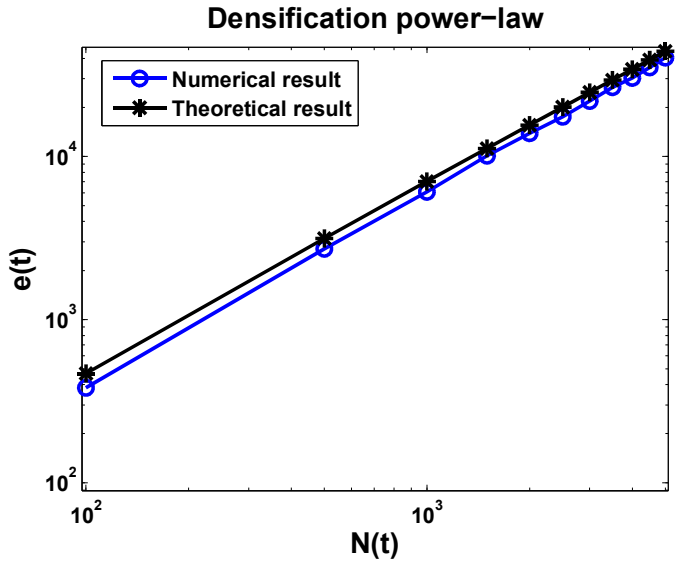


Figure 6.7 : Comparison between numerical and theoretical results for densification power-law (DPL) in a network generated by NRM at $p = 0.4, \beta = 0.2$ is plotted. The count of triangles is averaged over 100 ensembles of growing networks generated by NRM.

6.3.4 Properties of NRM by Numerical Simulation

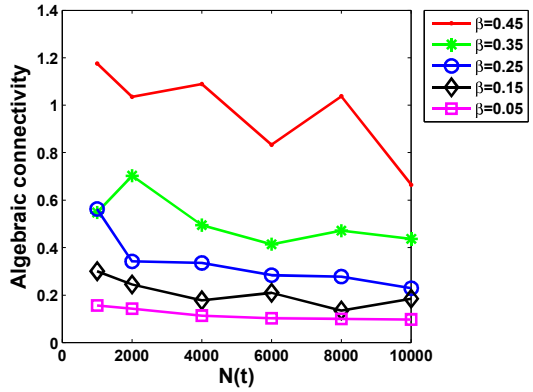
Even though the growth process of NRM is simple, closed-form expressions of several properties of networks generated by NRM are in general difficult to obtain due to cumbersome calculations. Thus in this section we investigate certain structural properties of such networks by setting fixed values of the parameters β and p. We focus on the dynamics of clustering coefficient, assortativity, algebraic connectivity, spectral radius and modularity of model networks during its formation by setting $p = 0.5$ and $\beta \in \{0.05, 0.15, 0.25, 0.35, 0.45\}$.

In Figs. 6.8(b), 6.8(e) and 6.8(d), we observe that the clustering coefficient (CC), assortativity, and modularity (Q) [Blondel *et al.*, 2008] are negatively correlated with the parameter β. As the value of β increases the value of clustering coefficient, assortativity, and modularity, decrease. Whereas, algebraic connectivity (AC) and spectral radius (SR) are positively correlated with β, see Figs. 6.8(a), 6.8(c).

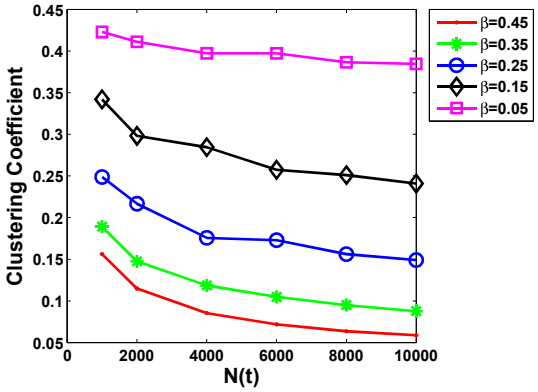
6.4 DISCUSSION AND FUTURE DIRECTION

Understanding the mechanism of network evolution lies at the core of many real-world problems and applications.

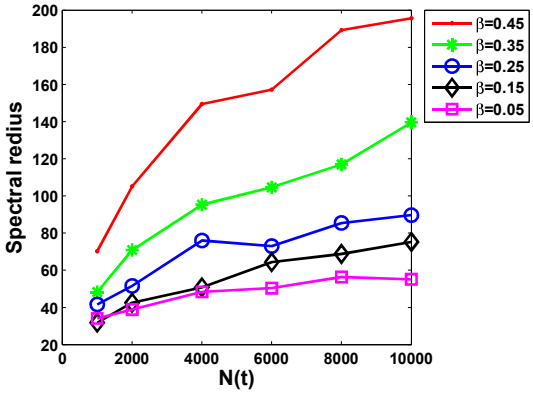
One of the fundamental problems in dealing with large real networks is the space complexity. In general, the space complexity of storing a network is $O(|V|^2)$ or $O(\bar{k}_t|V(t)|)$, where $\bar{k}_t$ and $|V(t)|$ denote the average degree and the number of nodes in a growing network at time-step t respectively. In addition, $\bar{k}_t$ increases with time t in many networks during the growth of the networks due to edge densification. We introduce a parametric growing scale-free random network model in this chapter that can inherit multiple structural properties of a given real network. Thus, in turn, we attempt to solve the problem of structural reconstruction for scale-free real networks. We develop a method to determine the optimal values of the parameters involved in the model by using only the power-law exponent or degree sequence of the given scale-free



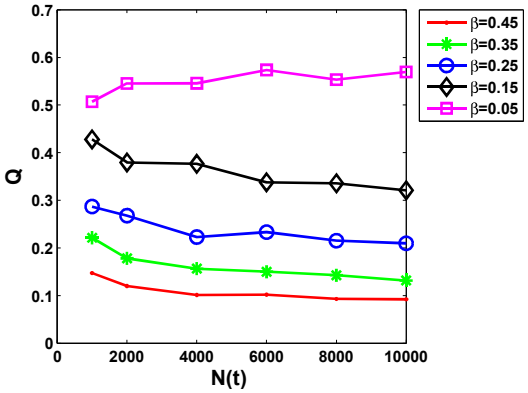
(a) Algebraic connectivity.



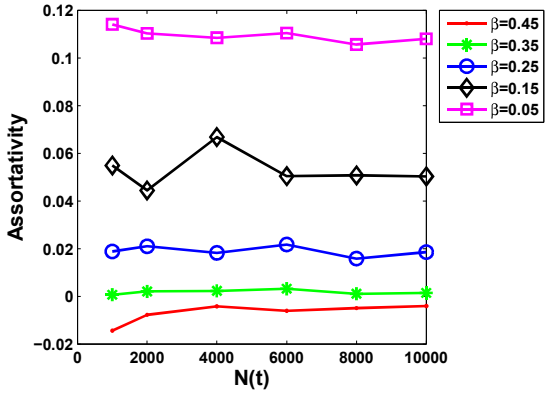
(b) Clustering coefficient.



(c) Spectral radius.



(d) Modularity index Q.



(e) Assortativity index.

Figure 6.8 : Numerical results of different properties of networks generated by NRM is plotted for different values of parameter β after setting $p = 0.5$. Horizontal axis represents size of the network $N(t)$ and vertical axis represents numerical values of considered property.

real network. Thus we show that the proposed approach can efficiently resolve the structural reconstruction problem for scale-free real networks efficiently. As a byproduct of this method, we deliver a novel technique to reduce the space complexity of storing a scale-free real network by storing only its degree sequence which is of $O(|V|)$.

We mention that it would be impossible to develop a mathematical model that can reproduce any given scale-free network. This follows the philosophy that the growth process of networks in different contexts could significantly differ. The crucial observation here is that if the scale-free real network is grown from a 'social' perspective, that is, based on preferential attachment (PA) and random attachment(RA) with local growth (LG) then NRM can better reproduce the network. We mention that PA and RA+LG are considered as building blocks for several growing real networks and network models [Chakrabarti and Faloutsos, 2006]. However, as we discussed before, the PA which we define in this chapter is not the standard concept of PA [Albert and Barabási, 2000]. Here PA is defined by a sequence of independent Bernoulli trials with varying success probabilities of formation of links and the success probabilities depend on the degrees of the existing nodes during the growth of the network. The novelty of the growth process adopted in NRM proves to be efficient as its performance is better than CDPAM, DMS, FFM, and CGA model. Besides, the method for structural reconstruction proposed in this chapter is computationally efficient while estimating the optimal values of the model parameters compared to the other models.

A natural question regarding NRM is to find out what role do the model parameters β, p play into determining specific structural and spectral properties of networks generated by it. We numerically investigate the influence of β in the model networks while p is fixed. In Fig. 6.8, we plot clustering coefficient, spectral radius, algebraic connectivity, modularity and assortativity index of the model networks by setting $p = 0.5$ and $\beta \in \{0.45, 0.35, 0.25, 0.15, 0.05\}$. We observe that as β (that is the contribution of PA) increases, the algebraic connectivity and spectral radius of the network increases. Thus it positively affects the spectral properties of the network. Whereas, the increment of β decreases modularity, assortativity, and clustering of the network. Thus β contributes negatively to the structural organization of the model networks. Finally, we can say that the local growth is responsible for the existence of community structure and clustering in the model networks. Recall that, community formation in social context relates to the formation of groups of similar social elements who have more connections inside the group compared to its connections outside the group. Moreover, connectivity of similar nodes increases the assortativity of the network. We mention that NRM has similar dynamics during its formation. Indeed, PA increases the connectivity of the network in NRM. As the value of β increases, the value of γ decreases while p is fixed. It is interesting to note that as the fraction of connections based on preferential attachment increases, the growth rate of diameter becomes slower, and at $\beta = 1$ NRM has the constant diameter.

It is well known that many real-world networks follow edge densification with shrinking diameter during its evolution. In several cases they follow densification power law (DPL) $|E(t)| \propto |V(t)|^{\alpha}$ where α lies in the interval $(1, 2)$. We prove that NRM demonstrates a similar pattern. In fact, when the power-law exponent γ lies between 2 to 3, NRM shows DPL without any constraint and most of the real-world networks have power-law exponent between 2 to 3. Thus the model networks generated by NRM can help to track the dynamics of growing real-world networks. An optimal model network which can inherit certain structural properties of a given real network can be used to extrapolate the future properties of the corresponding real network. In addition, a significant difference between extrapolated results and real results can be considered as an indicator of the abnormality in the real network. For example, link failure cascade, spamming or artificial generation of unnecessary links in a network can lead to irregularities in the edge densification law.

Note that, it is often difficult to access the real-time diffusion dynamics on a real network. The optimal model network generated by NRM for a real network can be used to study different diffusion protocols on the real network.

6.5 CONCLUSION

In this chapter, a solution is provided to reduce the space complexity used to store the social or complex networks. We can store degree sequence of a network and later using this sequence and NRM, we can reproduce the other properties with a good approximation (low error) compared to other models considered here. A parametric network generation model is proposed for reconstruction of scale-free networks with given degree sequence or degree distribution. A sufficient condition for the model parameters is provided to generate networks which follow edge-densification and densification power-law. Computable expressions for expected number of triangles and expected diameter for the networks generated by the proposed model are derived. Reconstruction of several real-world networks is done in support of the novelty of the proposed model. By numerical computations, it is established that the proposed model can generate networks with shrinking diameter and modular structure when specific model parameters are chosen. In this chapter, we focus on reduction of space complexity of complex networks. The method of network reconstruction discussed in this chapter has linear space complexity of order $O(|V|)$ and linear runtime complexity of order $O(m_1)$ or $O(m_2)$.

…