7

Biased random walk and information diffusion

In this chapter we utilize the dependency of structure of the network and diffusion on the network to provide possible explanation for preferential growth of the real networks. Biased random walk is an observed phenomena in networks which is adopted to explain the biased growth of the networks such as Internet and WWW. Truncated random walk is another dimension of diffusion analysis in real-world systems which is considered in this chapter. The proposed biased random walk is utilized to calculate the *PageRank* efficiently.

Networks are ubiquitous in real world [Kelly, 2011]. For example, communication and information systems [Schwartz, 1977], transportation systems [Bell and Iida, 1997], biological systems [Girvan and Newman, 2002; Assenov *et al.*, 2008] and many socio-economic systems can often be interpreted as network systems. In general, these networks are dynamic in nature, where number of nodes and connectivity pattern between the nodes vary with time. However, it is observed that many real networks exhibit power-law in the tail of its degree distribution, for example, WWW and Internet. Often, specially in information and communication networks, and social networks, diffusion of information is considered on the networks, in which, an amount of a diffusive material gets diffused via links in the network, over the entire network. Thus in a dynamic network, connectivity pattern influences the diffusion dynamics. On the other hand, the diffusion pattern influences the nodes to add/remove links for acquiring or not acquiring the diffusive material. For example, temporal removal of links happen during an outbreak of disease/ infectious diffusive material in social networks. Theory of self-organization is employed to justify the growth of a network and diffusion dynamics on it in [Aoki *et al.*, 2015].

Thus a pertinent question about the structural dynamics of a network is the following: How does diffusion dynamics influence the nodes of a network such that power-law in degree distribution emerge or sustain in the network? Earlier, it has been speculated that preferential attachment is a procedure for link formation which can explain the arise of power-law in a growing network. Here we address whether the diffusion pattern can justify the existence or arise of power-law in the degree distribution of a real network.

We mention that several attempts have been made in literature to model and investigate diffusion phenomena on real networks. For example, information diffusion is often described as a random walk dynamics on a network. Indeed, a piece of information or data can be represented by a random walker, and the current and the immediate next position of the walker correspond to the sender and receiver nodes in an information network. Note that the activity level of the links plays a big role in information diffusion. Consider WWW network as an example which follows preferential attachment [Barabási *et al.*, 2000] and searching dynamics in it is modeled as a random walk phenomena [Adamic *et al.*, 2001]. In the preferential attachment, a new node prefers to get linked with a highly connected node in the network [Barabási *et al.*, 2000] whereas, according to random walk dynamics, average searching time also known as mean-first-passage-tie (MFPT) of a node in a network is inversely proportional to the degree of that node [Redner, 2001]. Thus a low degree node has high MFPT which does not justify the preferential growth of the network because if a node has low degree then it will have high searching time. Hereby one can claim that degree based preferential or biased growth and random walk based searching are not consistent to explain the dynamics of real-world networks which follow power-law degree distribution. Other existing studies on biased random walk dynamics are not able to justify the preferential or biased growth of the networks. From the above discussion, it is concluded that there is a hidden dynamics of diffusion which may not be captured by existing random walk dynamics (based or unbiased)[Redner, 2001; Fronczak and Fronczak, 2009].

In another direction of diffusion phenomena, we can observe that diffusion on real-world networks is not a continuous random walk process as considered in previous studies. A walker can stay at a node in for a while or jump to another node which may not be directly linked to the current location (node) of the walker. Transition probability of a link can vary with time or in a special case, it can be reduced. From the above discussion, we can deduce that a rigor analysis of diffusion as a random walk is required that can justify the preferential or biased growth of networks and such a walk can be adopted to study the diffusion and searching in networks. A biased random walk is usually defined by using property of links, interaction frequency of links and/or a function of degrees of incident nodes while defining the transition probability of a link.

In the existing literature on biased random walks, the bias is formulated based on the degree of a node, specifically, the receiver node in the context of information diffusion [Wang *et al.*, 2006]. However, we mention that there are examples of social networks in which activity level of a link (w_{ij}) depends on the structural property of both the participating nodes as

$$w_{ij} \propto (k_i k_j)^{\theta}, \tag{7.1}$$

where θ is a constant [Barrat *et al.*, 2004]. In this chapter, we define the transition probability of a link by using the activity level of that link given by

$$w_{ij} = \alpha k_i k_j, \tag{7.2}$$

where α is a proportionality constant such that $0 < \sum_j w_{ij} < 1$. Hence $\sum_j w_{ij}$ is the probability that a walker leaves the node i if it is at the node i and w_{ij} is the probability of leaving the node i via edge (i, j), hence it represents the transition probability. In this chapter, we incorporate the role of heterogeneous properties of links to define biased random walks. This is natural due to its importance in many real world applications, for example, if two places (nodes) (X and Y) are connected via two paths (P_1 and P_2) of multiple hops, people (walkers) prefer a path over another on the basis of its attributes such as traffic, the number of red signals and condition of the roads which are inherent properties of the links involved in the paths. Further, in literature, it is assumed that a walker (random and/or biased) never ends its walking. We explore and investigate random and biased random walks in which a walker can stop permanently at some node in the network with a non-zero probability during its walking. We call this phenomenon as "*walker may kill itself*". For example, observe that information diffusion in a social network is not a long lasting process.

Thus in this chapter, we introduce discontinued truncated biased random walk as a representation of information diffusion in social networks and this can have applications in truncated searching strategies. Truncated searching is defined as a searching process in which a walker has an upper time limit of searching.

The contribution in this chapter can be summarized as follows. We introduced a biased random walk on a network and interpret it as a diffusion dynamics on the network. We show that it can justify the preferential growth of the network. This results into existence of power-law in the degree distribution of the network. As mentioned above that it is realistic to assume that diffusion is not a continuous process forever, we introduce discontinued truncated random walk (DTRW) on networks and explain it as information diffusion dynamics on the network that ultimately shown to be useful to calculate *PageRank* ordering of nodes in the network. DTRW also supports the preferential growth of the network.

The organization of the chapter is as follows. A background and literature survey is provided in next Section. Proposed dynamics of the biased random walk is discussed in Section 7.2. Section 7.3 is dedicated for discontinued truncated random walk which has application in calculating PageRank efficiently. In Section 7.4, detailed discussion of biased random walk and truncated random walk along with concluding remarks is provided.

7.1 BACKGROUND AND RELATED WORK

We recall that random walk dynamics is often used to study different diffusion dynamics on networks. Dynamics of a random walker is investigated in many contexts [Nash-Williams, 1959; Tetali, 1991], for example, navigation and centrality of networks [Perra *et al.*, 2012; Starnini *et al.*, 2012], routing of packets in Internet, and diffusion in communication networks. Traffic in transportation networks can also be represented as phenomena of random walkers [Wang *et al.*, 2006]. Random walk dynamics is applied to explore the dynamics of wealth's distribution in economic networks, gene expression pathways in biological networks and search (navigation) strategies in Internet [Adamic *et al.*, 2001; Tadić and Rodgers, 2002; Tadić and Thurner, 2004; Kim *et al.*, 2002; Rosvall *et al.*, 2005; Germano and de Moura, 2006]. In random walks, a walker positioned at a node i can move to any node j which is linked to the node i in the network with equal probability. Thus the selection of a node for a move is uniform among the neighbors of the current node occupied by a random walker [Akyildiz *et al.*, 2000; Noh and Rieger, 2004].

Thus the standard random walk dynamics can be explained as follows. Let $p_i(t)$ be the occupation probability of a node i at time t then,

$$p_i(t+1) = \sum_j a_{ij} \frac{p_j(t)}{k_j},$$

and the collective dynamics of the random walking in a network is given by

$$\mathbf{p}(t+1) = AD^{-1}\mathbf{p}(t), \tag{7.3}$$

where a_{ij} is the ij^{th} entry of adjacency matrix A of the network and D is the degree matrix associated with the network [Redner, 2001]. Solution of the Eq. (7.3) is given by

$$\mathbf{p}_\infty = \lim_{t\to\infty} \mathbf{p}(t) = \frac{\mathbf{d}}{\mathbf{1}^T\mathbf{d}},$$

where $\mathbf{d} = D\mathbf{1}$ and $\mathbf{1}$ is a all-one column vector of appropriate length.

Observe that a node of higher degree has higher probability of finding the random walker which can represent information or any other diffusive resources. This can motivate a new node to go for preferential attachment. But as soon as the new node joins an existing network, it becomes the part of a network and follows the same diffusion protocol and hence, it gets fewer resources indicated by occupation probability. This reduces the probability of existence of the newly formed link.

Next motivation for preferential attachment could be the searching time or mean-first-passage-time (MFPT) of a node. However, MFPT $\propto 1/k$, where k is the degree of the node which also does the negative effect in preferential growth of a network. Hence the standard random walk dynamics fails to interpret the phenomena of preferential attachment.

Later, preferential navigation (biased random walk) is explored in [Fronczak and Fronczak, 2009] as a model of information diffusion. In biased random walks, a walker shows bias while choosing the next move to one of the neighbors of the node occupied by the walker by setting priorities. The probability of passing (transition probability) of a walker from a node i to a neighbor node j via the link (i, j) is usually defined by

$$w_{ij} = \frac{k_j^{\alpha}}{\sum_{l \in \mathcal{N}_i} k_l^{\alpha}},$$

where α is a constant,k_j is the degree of a node j and $\mathcal{N}_i$ denotes the set of neighbors of i [Wang *et al.*, 2006]. A different variant of transition probability can also be defined to realize different phenomena in networks, for example, a localized biased random walk phenomenon is presented in [Sood and Grassberger, 2007] by using indirect neighbors (neighbors at 2 or more shortest distance) for the transition probability between two connected nodes (direct or indirect). In [Fronczak and Fronczak, 2009; Yuan and Chern, 1992; Sood and Grassberger, 2007; Motter *et al.*, 2005], attributes (properties) of nodes have been considered to define a biased random walk. We can observe the similar results as discussed for random walks, in the case of biased random walk in which transition probability depends on the degree of the receiver node only.

7.2 BIASED RANDOM WALK USING PROPERTY OF LINKS

In this section we introduce a biased random walk on networks by utilizig property of its links. We consider mean-first-passage-time (MFPT) [Redner, 2001] as a measure to investigate the dynamics of the proposed random walk. Recall that MFPT is defined as the average number of hops or time needed to reach a node i from any arbitrary node in the network [Perra *et al.*, 2012; Gefen and Goldhirsch, 1985]. Thus it is the average time taken in the process of network navigation to access some specific node. Let $q(t,i)$ be the probability that the walker arrives at the node i first time at time t. If μ_i is the probability of the walker to reach to a node i from its neighbors during the walking then

$$q(t,i) = \mu_i (1-\mu_i)^{t-1}. \tag{7.4}$$

Assume that the probability of travelling of the walker from a node j (neighbour of target node i) to i is $0 < \alpha k_i k_j < 1$. Then if the probability of finding the walker at node j at steady state is $\widehat{p}_j$ then we have

$$\mu_i = \sum_j \widehat{p}_j \alpha k_i k_j a_{ij}, \tag{7.5}$$

where a_{ij} is the ij^{th} entry of the adjacency matrix A associated with the network. Recall that $a_{ij} = 1$ if the nodes i and j are linked, otherwise $a_{ij} = 0$. The MFPT of node i (T_i) is then given by

$$T_i = \sum_{t=0}^{\infty} t q(t,i) = \frac{1}{\mu_i} = \frac{1}{\alpha k_i \sum_j \widehat{p}_j k_j a_{ij}}. \tag{7.6}$$

Then it follows from equation (7.6) that we need to calculate the occupation probability of node i, that is, $\widehat{p}_j$ to obtain an explicit expression of MFPT which can possibly explain the dynamics of the biased random walk proposed in this chapter.

Let $p_i(t)$ be the probability that random walker is at node i at time t. Then we define the dynamics of the proposed biased random walker as

$$p_i(t+1) = \sum_j a_{ij} w_{ji} p_j(t) + p_i(t) \left(1 - \sum_j a_{ij} w_{ij}\right) \tag{7.7}$$

where $w_{ji} p_j(t)$ represents the probability that the walker will reach to the node i from its neighbour j, if the walker is present at node j. The second term corresponds to the probability that the walker will not leave the node i if it is there. Replacing $w_{ij} = \alpha k_i k_j$ in equation (7.7) we obtain

$$p_i(t+1) = \sum_j \alpha k_i k_j a_{ij} p_j(t) + p_i(t) \left(1 - \sum_j a_{ij} \alpha k_i k_j\right). \tag{7.8}$$

Hence the collective dynamics of the proposed biased random walk in a network is given by

$$\begin{aligned} \mathbf{p}(t+1) &= \alpha DAD\mathbf{p}(t) + (\mathrm{I} - \alpha \mathrm{Dia}\,(A\mathbf{d})\, D)\, \mathbf{p}(t) \\ \Rightarrow \mathbf{p}(t+1) - \mathbf{p}(t) &= -\alpha\, (\mathrm{Dia}\,(A\mathbf{d})\, D - DAD)\, \mathbf{p}(t) \end{aligned}$$

where $\mathbf{p}(t) = [p_1(t),\ p_2(t),\ p_3(t),\ ...,\ p_n(t)]^T$, I is the identity matrix of appropriate dimension, and Dia($\mathbf{x}$) is the diagonal matrix with i^{th} diagonal entry as the i^{th} entry of the vector $\mathbf{x}$. After approximating the difference equation in to a differential equation, we obtain

$$\dot{\mathbf{p}}(t) = -\alpha M \mathbf{p}(t), \tag{7.9}$$

where $M = (\mathrm{Dia}\,(A\mathbf{d})\, D - DAD)$. Observe that M is a symmetric diagonally dominant matrix with positive diagonal entries, hence M is a positive semi-definite matrix. Further, $M\mathbf{1} = \mathbf{0}$ since $M_{ii} = \sum_{j \neq i} |M_{ij}|$ and $\mathbf{1}$ is the all-one vector of length n. Solution of the Eq. (7.9) is then given by (for a detailed proof see chapter 3, [Mesbahi and Egerstedt, 2010])

$$\mathbf{p}(t) = e^{-\alpha M t} \mathbf{p}_0 = \sum_i e^{-\alpha \lambda_i(M) t} \left(\mathbf{u}_i^T \mathbf{p}_0\right) \mathbf{u}_i, \tag{7.10}$$

where $\mathbf{u}_i$ is an eigenvector corresponding to the ith eigenvalue $\lambda_i(M)$ of M. For a connected network, the least eigenvalue $\lambda_1(M) = 0$, $\mathbf{u}_1 = \mathbf{1}/\sqrt{n}$ and $\lambda_i(M) > 0$ for $i \geq 2$. Thus as $t \to \infty$,

$$\mathbf{p}(t) \to \widehat{\mathbf{p}} = \frac{\mathbf{1}^T \mathbf{p}_0}{n} \mathbf{1}. \tag{7.11}$$

The biased random walk process defined by the Eq. (7.9) is verified by performing a simulation on a connected test network on 17 nodes having average degree of the network 2.37. A random vector of unit length under 2 norm, that is, $||\mathbf{p}_0||_2 = 1$ is considered as initial condition. Convergence of biased random walk is shown in Fig. 7.1. The occupation probability of each node for a biased random walker, defined by (7.9), is same at steady state. Now one can say that how this could be motivation for preferential attachment while the occupation probability is same for all nodes?

The average searching time of a node i (defined as MFPT) can be calculated by using Eqs. (7.6) and (7.11) as

$$T_i = \frac{n}{\alpha k_i \left(\mathbf{1}^T \mathbf{p}_0\right) \sum_j k_j a_{ij}}. \tag{7.12}$$

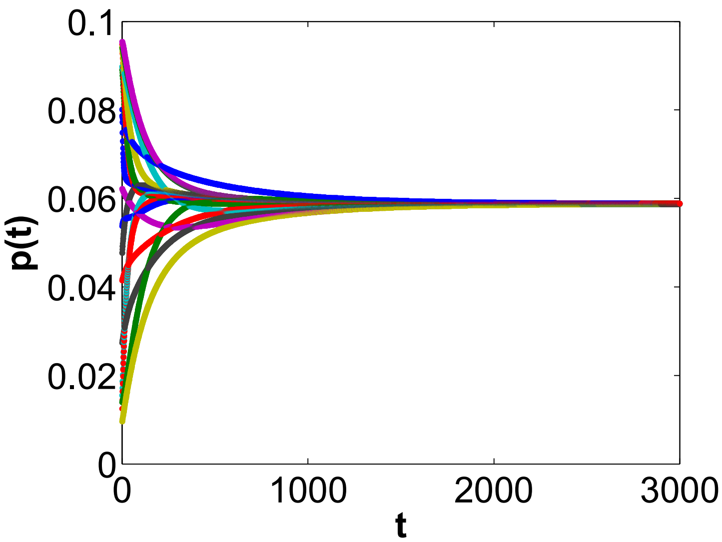

Figure 7.1 : Single particle diffusion defined by Eq. (7.9) is simulated over a test network of 17 nodes.

Observe that the average searching time T_i not only depends on the connectivity (degree) of i but also on the distribution of its neighbors' connectivity. We mention that in other searching strategies under random or biased random walk that are available in literature, for example in [Fronczak and Fronczak, 2009] the MFPT of a node i is of the form $1/k_i^\gamma$ or some other similar function that clearly indicates that higher degree nodes have less MFPT value, that is, they receive more visits. Hence there are no benefits of being linked with a node of high degree unless it does not have a higher degree. Thus there is no motivation to stay linked with a higher degree node in a network. In contrast, the result obtained in Eq. (7.12) has a different story to offer. The proposed biased random walk explains that even if a node is not well connected with the entire network it can receive more visits if it is linked with highly connected nodes. This phenomenon motivates a node to get linked and stay connected with high degree nodes and hence the biased random walk provides a possible explanation of preferential growth in real world networks, for example, on the internet or the web graph WWW. Thus it would not be exaggerated to say that the proposed random walk process proves the saying that '*being a friend of a rich person is beneficial*' which promotes the preferential growth of a network. This is an example of self-organization in a network which leads to power-law or preferential growth of Internet and similar networks.

We compare the dynamics of the biased random walk defined by Eq. (7.9) with the random walk proposed in [Noh and Rieger, 2004] on a scale-free network of size 5000 with average degree 4 generated by using the Barabasi-Albert model. In Figs. 7.2 and 7.3 the MFPT of all the nodes in that network are plotted under biased and unbiased navigation schemes respectively. If the probability of successful transfer of a walker depends only on the degree of the node occupied by the walker then the nodes of the lower degree show higher MFPT, and nodes of the higher degree have lower MFPT, see Fig. 7.3. Similar results are stated in the case of degree dependent biased navigation in which the probability of successful transfer of a walker depends on the degree of receiver node [Fronczak and Fronczak, 2009]. Biased random walk defined in this chapter has an advantage in which a node of lower degree connected with higher degree nodes can have lower MFPT, see Fig. 7.2. The comparison of improved MFPT of lower and higher degree nodes defined by the proposed biased random walk and considered random walk is provided in Fig. 7.4.

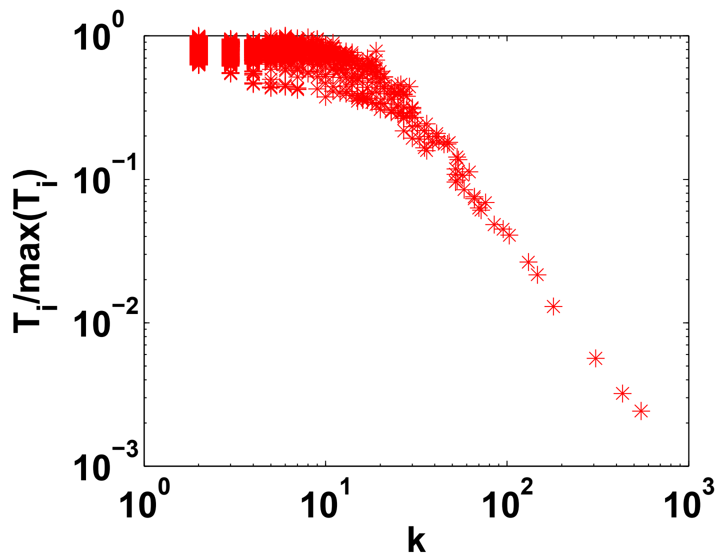

Figure 7.2 : Single particle diffusion defined by (7.9) is simulated over the network of 5000 nodes which is generated under BA model. Relative MFPT of nodes is plotted.

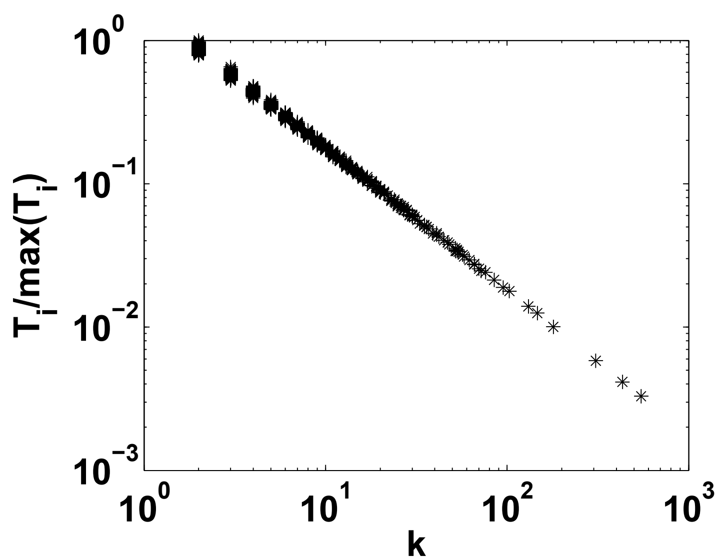

Figure 7.3 : Single particle diffusion under random walker is simulated over the network of 5000 nodes which is generated under BA model. Relative MFPT of nodes is plotted.

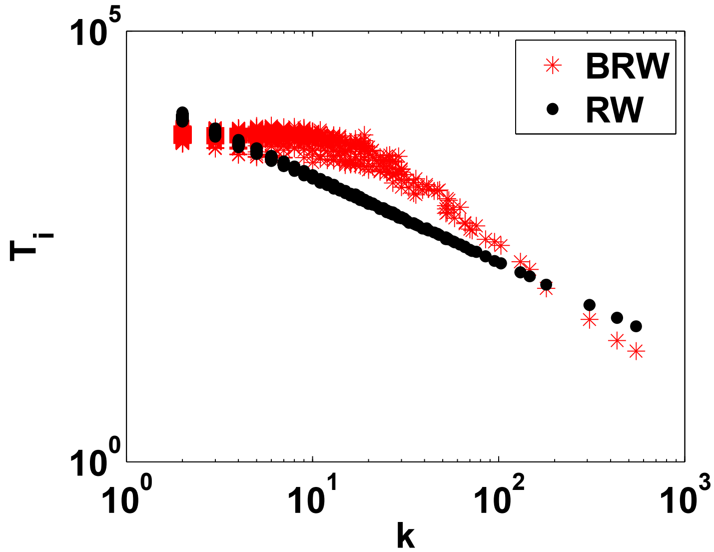

Figure 7.4 : Comparison of MFPT under simple random walk (RW) and biased random walk (BRW). Simulation is done on a network of 5000 nodes which is generated under BA model.

7.3 DISCONTINUED TRUNCATED BIASED RANDOM WALK

Till now we discussed biased random walks in which a walker continues its walking as $t \to \infty$. However, there are evidences of real-world processes in which a walking does not necessarily continue forever, for example, information diffusion (spreading of a message) and searching in WWW. In the following, we investigate single walker navigation on networks in which a walker may stop permanently during its travel. This happens, for example, in the case of single message forwarding in a network. Truncated searching is also an application of such a biased random walk. We propose two such processes for real-world applications.

Process-I Consider that a walker presents at a node i can walk to one of its neighbors j with probability $w_{ij} = \alpha k_i k_j$ and the probability that it will not leave the node i is $1 - \sum_j w_{ij}$ where α is a positive constant such that $0 < \sum_j w_{ij} < 1$. If the walker is alive at the node j at time t with probability $p_j(t)$ then the probability that the walker is present at node i alive at time $t+1$ is given by

$$p_i(t+1) = \sum_j \alpha a_{ij} k_i k_j p_j(t).$$

This leads to a collective dynamics of the random walker under discontinued walking as

$$\mathbf{p}(t+1) = \mathscr{T}\mathbf{p}(t). \tag{7.13}$$

where $\mathscr{T} = \alpha DAD$ is the transition matrix in which $0 < \sum_j \mathscr{T}_{ij} < 1$ and $\mathscr{T}$ has maximum eigenvalue positive and it is less than 1.

Let $\mathbf{p}_0$ be the initial condition of the random walker. Then the recursive relation in Eq. (7.13) can be converted to

$$\mathbf{p}(t) = \mathscr{T}^t \mathbf{p}_0.$$

For a large value of t

$$\mathbf{p}(t) = \lambda_{\mathscr{T}}^t \mathbf{v}_{\mathscr{T}},$$

where $\mathbf{v}_{\mathscr{T}}$ is the leading eigenvector of matrix $\mathscr{T}$ and $\lambda_{\mathscr{T}}$ is corresponding maximum eigenvalue of τ, see [Cull *et al.*, 2005] for detailed proof of such recurrence relations.

Process-II This process is motivated by the following example. Consider a simple real world example of migration of people from cities and villages. Cities are considered as well-connected nodes and villages are not so well-connected. The four cases of migration can be discussed as follows. (i) City to city, (ii) village to village, (iii) village to city and (iv) City to village. We may assume that the probabilities of migration between cities and between villages are similar. The probability of migration of population from villages to cities is much higher as compared to the other cases. For (iv) the probability may be a very small number.

Thus we define the transition probability $w_{ij} = \alpha \frac{k_j}{k_i}$ where α is a positive constant such that $0 < \sum_j w_{ij} < 1$ and then the proposed biased random walk is given by

$$p_i(t+1) = \alpha \sum_j a_{ij} \frac{k_i}{k_j} p_j(t).$$

The collective dynamics of the walker can be written in the matrix form as

$$\mathbf{p}(t+1) = \alpha D A D^{-1} \mathbf{p}(t). \tag{7.14}$$

Let $\mathbf{p}_0$ be the initial condition of the random walker. Then we obtain

$$\mathbf{p}(t) = \alpha D A^t D^{-1} \mathbf{p}_0.$$

For a large value of t

$$\mathbf{p}(t) = (\alpha\lambda_1)^t D\mathbf{v}_1$$

where v_1 is the leading eigenvector and λ_1 is corresponding eigenvalue of the adjacency matrix of the network (see [Cull *et al.*, 2005] for a detailed proof). Note that $p_i(t) = (\alpha\lambda_1)^t (D\mathbf{v}_1)_i$ where $\alpha\lambda_1 < 1$ since $0 < \left(\sum_j w_{ij}\right) < 1$. Thus $p_i(t)$ depends on both the eigenvector centrality and degree centrality of the node i.

Now the question is that can a random walker be able to visit a node i before stopping? Note that at steady state $\widehat{p} \to \mathbf{0}$ which indicates that $\mu_i \to 0$ as defined by Eq. (7.5). However for a large value of t $\mu_i = \sum_j p_j(t)\alpha k_i k_j a_{ij}$ and $q(t,i) = \mu_i (1-\mu_i)^{t-1}$. Then the probability of visiting a node i by the random walker before its stopping is given by

$$\begin{aligned}
\mathbf{P}_i(\tau) &= \sum_{t=0}^{\tau} q(t,i) \\
&= \sum_{t=0}^{\tau} \mu_i (1-\mu_i)^{t-1} \\
&= \sum_{t=0}^{\tau} \sum_j p_j(t)\alpha k_i k_j a_{ij} \left(1 - \sum_j p_j(t)\alpha k_i k_j a_{ij}\right)^{t-1} \\
&\le \sum_{t=0}^{\tau} \left(\sum_j p_j(t)\alpha k_i k_j a_{ij}\right).
\end{aligned}$$

Then for **Process-I** we have

$$
\begin{aligned}
\mathbf{P}_i(\tau) &\leq \sum_{t=0}^{\tau}\sum_{j}\lambda_{\mathscr{T}}^{t} k_j v_{\mathscr{T}}^{j} \alpha k_i k_j a_{ij} \\
&\leq \alpha k_i \sum_{t=0}^{\tau}\lambda_{\mathscr{T}}^{t} \sum_{j} v_{\mathscr{T}}^{j} k_j^2 a_{ij} \\
&\leq \alpha k_i \left(\frac{1-\lambda_{\mathscr{T}}^{\tau+1}}{1-\lambda_{\mathscr{T}}}\right) \sum_{j} v_{\mathscr{T}}^{j} k_j^2 a_{ij}
\end{aligned}
$$

and for **Process-II**

$$
\mathbf{P}_i(\tau) \leq \alpha k_i \left(\frac{1-(\alpha\lambda_1)^{\tau+1}}{1-\alpha\lambda_1}\right) \sum_{j} v_1^{j} a_{ij}.
$$

Thus observe that, in the context of truncated searching Eqs. (7.14) and (7.13) have different dynamics as compared to the continued ($t \to \infty$) searching process defined by Eq. (7.9). In continued searching, the occupation probability of the walker at each node is same or proportional to the degree of the node, whereas, in truncated searching, the probability of finding the walker at a node depends on the degree of the node and degree distribution of its neighbors.

We mention that finding an explicit expression for MFPT in truncated biased random walks is difficult as the probability $p_i(t)$ depends on t. Indeed one can obtain a pattern of the MFPT of the random walker at time τ by dividing the entire time span into small intervals of size $\triangle$. Then the initial times of these intervals are $b_1 = 1$, $b_2 = \triangle + 1$, $b_3 = 2\triangle + 1, \ldots,$ $b_{\tau/\triangle} = \tau - \triangle + 1$. In each interval, MFPT is calculated by considering the initial time of that interval and constant μ_i which is calculated by using the minimum or maximum value of $p_i(t)$ in the same interval. If we consider the minimum (maximum) value of $p_i(t)$ in each interval then we can obtain the upper (lower) bound of MFPT by calculating the expression in Eq. (7.15) as follows.

$$
\mathscr{F}(\triangle, \mu_i^l) = \sum_{t=1}^{\triangle}\left(1-\mu_i^l\right)^l - \triangle\left(1-\mu_i^l\right)^{\triangle+1}, \tag{7.15}
$$

where μ_i^l is in l^{th} interval which is calculated using maximum or minimum of $p_i(t)$ in the same interval. In the context of information diffusion (message passing), $\mathbf{P}$ is the probability distribution of information diffusion which is viewed as a biased random walk in the network under the given condition that the walker can stop during the diffusion. Within the time limit τ, a node i can receive the first message (information) on average time $T_i(\tau)$ with probability $\mathbf{P}_i(\tau)$.

Information diffusion phenomena are simulated under discontinued truncated biased random walking in which the number of walkers increases as each node which has information can start sending it to its neighbors. The diffusion is simulated over the same scale-free network considered previously. Spreading of information started from a random node. Time of reaching the information at a node, first time, is also calculated. This simulation repeated 1000 times and average reaching time, MFPT, is plotted in Figs. 7.5 and 7.7 for truncated biased random walk defined by Eqs. (7.13) and (7.14), respectively. The probability that a node i receives information within time τ having average reaching time $T_i(\tau)$ is also calculated and plotted in Figs. 7.6 and 7.8 for the corresponding biased diffusion given by Eqs. (7.13) and (7.14) respectively.

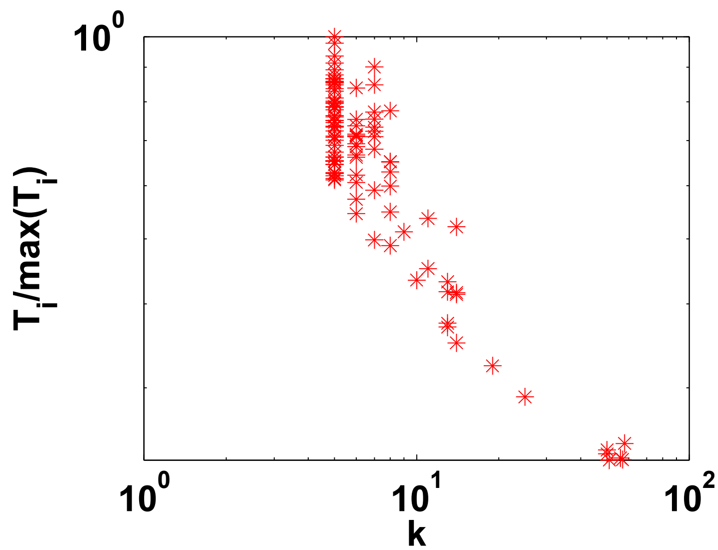

Figure 7.5 : Single particle diffusion defined by Eq. (7.13) is simulated over a network of 100 nodes which is generated unde BA model.

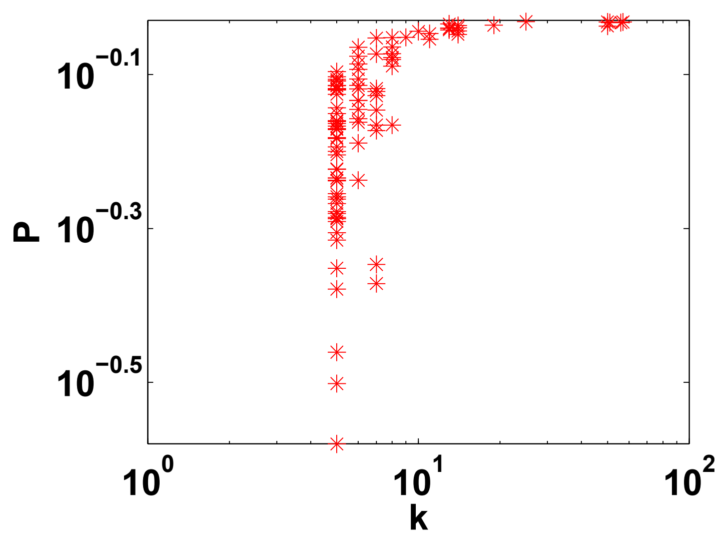

Figure 7.6 : Information spreading is simulated as a diffusion defined by Eq. (7.13) over a network of 100 nodes which is generated unde BA model. Probability distribution is plotted for the reachability of information to a node within considered time τ. Simulation performed 1000 times with same parameter α and different source node which is selected randomly to calculate the probability distribution.

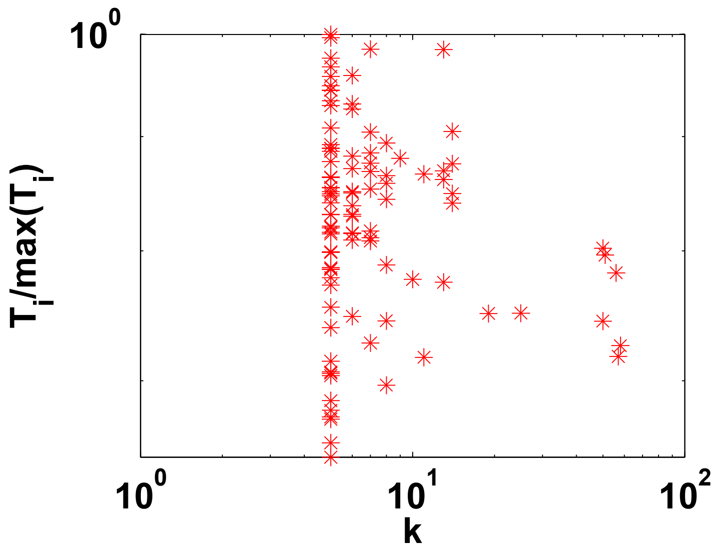

Figure 7.7 : Single particle diffusion defined by Eq. (7.14) is simulated over a network of 100 nodes which is generated unde BA model.

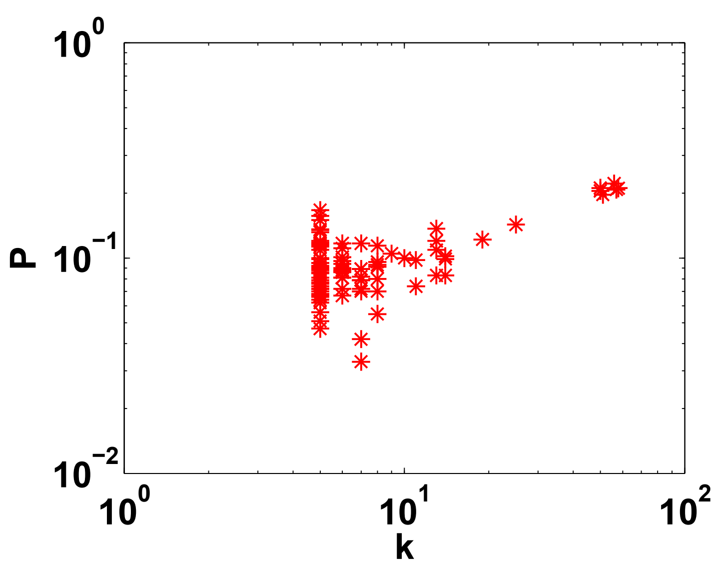

Figure 7.8 : Information spreading is simulated as a diffusion defined by Eq. (7.14) over a network of 100 nodes which is generated unde BA model. Probability distribution is plotted for the reachability of information to a node within considered time τ. Simulation performed 1000 times with same parameter α and different source node which is selected randomly to calculate the probability distribution.

The results obtained for average reaching time of information to a node under simulation is same as obtained theoretically under the defined truncated diffusion protocols. Nodes of higher degree are more probable to get information in less time while the distribution of MFPT and the corresponding probabilities have wide distribution for lower degree nodes. The lower degree nodes which are linked with higher degree nodes have lower MFPT value with high probability, see Figs. 7.5 and 7.6. But in the case of diffusion under Eq. (7.14), higher degree nodes act like sinks which do not release the information easily. Assortativity in the network improves the diffusion in this case. A higher degree node has lower MFPT with high probability. As before, in this case, the lower degree nodes have a wide distribution of MFPT and probability **P**, however, the reasoning is different. The lower degree nodes which are linked with lower degree nodes have a tendency to receive information from lower degree nodes if they have it. In this case (second case) assortativity for lower degree nodes is preferable. There is high positive correlation 0.94 and 0.788 between quantity $f_i = \sum_j \alpha k_i k_j$ and $\mathbf{P}_i$ for the diffusion under Eqs. (7.13) and (7.14), respectively.

The distribution of MFPT in the biased random walk, and probability distribution in discontinued truncated biased random walk depend on the structural properties of neighboring nodes. This dependency provides an initial feed for preferential growth. A newly connected node can be benefited from its neighbors similar to social networks. Strategic network model also suggest the similar dynamics in the form of utility [**?**].

Thus we conclude that the discontinued truncated diffusion can have practical applications in web-searching, the study of information diffusion, packet routing on Internet etc. The results obtained in this chapter also establish an analytical reasoning for the preferential growth of real world networks.

7.3.1 PageRank and discontinued truncated random walk

In this section, we discuss a method to calculate *PageRank* search which is used to rank web pages in WWW network [Newman, 2010] by utilizing the concept of discontinued truncated random walk. *PageRank* is an ordering metric which signifies the importance of nodes [Langville and Meyer, 2011; Pasquinelli, 2009].

We consider discontinued random walk in which a random walker can continue its random walking with probability β and otherwise it stops. The collective dynamics is given by

$$\mathbf{p}(t+1) = \beta A \mathrm{D}^{-1} \mathbf{p}(t). \tag{7.16}$$

Assume that a random walker starts its walking from some other arbitrary random node. Thus the random walking with jumping phenomena (also called as repeated truncated random walk) can be modelled as follows [Sarma *et al.*, 2015].

Let the walker be at a node i at time t with probability $p_i(t)$. Next, it can go either for a walk to its neighbor with probability β or do a random jump with probability $1-\beta$. In a network on n nodes, $1/n$ is the probability of being occupied by the walker in a single jump. Thus the collective dynamics of the random walk can be written as

$$\mathbf{p}(t+1) = \beta A \mathrm{D}^{-1} \mathbf{p}(t) + \frac{1-\beta}{n}\mathbf{1}. \tag{7.17}$$

At converging state, solution of the Eq. (7.17) is $\mathbf{p} = \frac{1-\beta}{n}\left(I - \beta A \mathrm{D}^{-1}\right)^{-1}\mathbf{1}$. The probability of visiting a node by random walker under this diffusion scheme is proportional to PageRank of

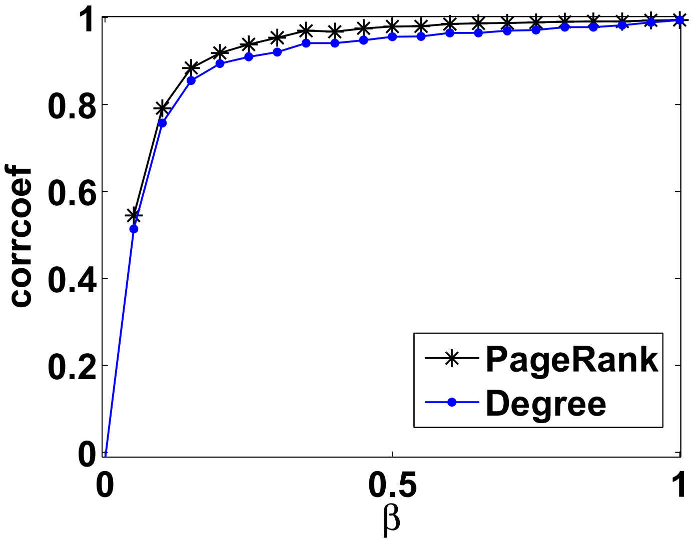

Figure 7.9 : Vertical axis represents correlation coefficient (correcoef) between considered metric (Degree and PageRank) and distribution of density of walkers and and horizontal axis corresponds to parameter β.

the network that is well-defined centrality used by google to rank the relative importance of web pages [Sarma *et al.*, 2015]. Solution of the Eq. (7.17) follows the *PageRank* ordering which requirs inverse of a matrix to be calculated. Can we reduce the complexity of calculating the *PageRank* ordering without computing the inverse of $I - \beta AD^{-1}$?

A search engine based on the random walker with random jumps converges to the desired outcome. The random walk dynamics defined by Eq. (7.17) is based on local dynamics that does not require complex computation and detailed global structure of the network. The proposed discontinued truncated random walk based method for calculating *PageRank* is a distributed diffusion process which reduces the omputational complexity. Similar results are provided in [Sarma *et al.*, 2015]. Converging state of the probability distribution, in Eq. (7.17), follows same distribution as *PageRank* which can be easily calculated by the density distribution of walkers or users in almost constant time.

Let a network on n nodes have m walkers per node which follows the searching dynamics defined by Eq. (7.17). After some time, the density of walkers in a node remains unchanged. The distribution of density of walkers at steady state is directly proportional to PageRank which can be used to rank the web pages. We consider an LFR network on 500 nodes with average degree 5.176 for simulation. Initially, we consider $m = 50$ per node, after that each walker performs random walk 50 times and then the density of random walkers is calculated. For validation, we calculated the correlation between density distribution of walkers and PageRank.

In Fig.7.9, the correlation between density distribution of walkers and considered metrics (PageRank and Degree) are plotted for different values of parameter β. It is observed that at the higher value of β, correlation is high for both degree and PageRank but PageRank always performs better than the degree. In Fig. (7.10), PageRank of the LFR network is plotted for different values of β. It is noticed that PageRank is widely distributed and shows more clear ranking of nodes. It promotes the consideration of the higher value of β (between 0.8 to 0.9) for better resolution of

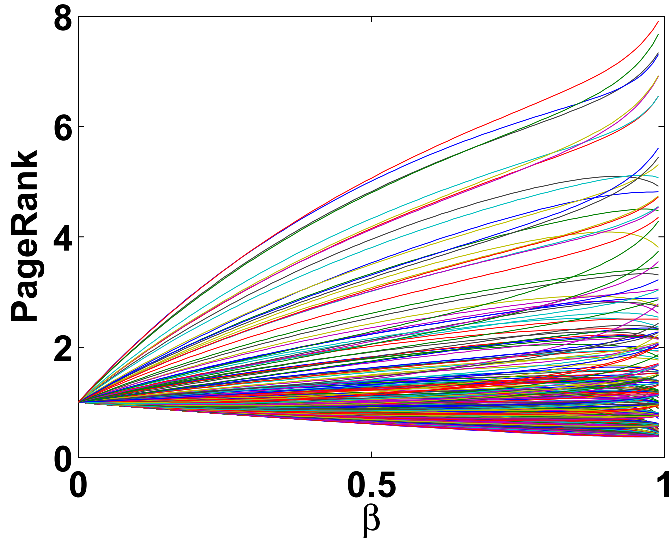

Figure 7.10 : Vertical axis represents PageRank and horizontal axis corresponds to parameter β.

ranking of nodes. From Figs. 7.9 and 7.10, we can conclude that the values of β should be higher but not very close to one otherwise Pagerank approaches to degree distribution.

7.4 DISCUSSION AND CONCLUSION

In a network with power-law degree distribution, there are a large number of nodes which have fewer connections and a few nodes in the network are well connected in the network and these are known as *hubs*. Preferential growth of the nodes in a network is considered to explain such behavior of degree distribution in real networks. In this chapter, we establish an explanation for such behavior of the growth process in a network considering the diffusion dynamics on the networks in the form of biased random walks. We studied simple biased random walks in which a random walker can walk forever and transition probability of a link depends on the degrees of the participating nodes of that link. The assumption of transition probability is adapted from [Barrat *et al.*, 2004]. Secondly, we also consider a discontinued truncated biased random walk on networks, in which a walker can stop while travelling in the network. Information diffusion is a real-life example of such phenomena. The results obtained in the form of MFPT in continuous biased random walk and the probability distribution of visiting a node in discontinued truncated biased random walk supports the preferential growth of real networks. We can conclude that diffusion dynamics of a node which considers the effect of its neighboring nodes provide a feedback force for biased growth similar to preferential attachment. The defined biased random walks can also be adopted in packet switching, searching, marketing in online social networks.

Discontinued truncated random walk is also adopted to calculate PageRank in a network which is simply projected as the density distribution of users in a network under repeated truncated random walk dynamics defined by Eq. (7.17). In this method of PageRank calculation, distributed local dynamics leads to desired collective behavior in the form of the density distribution of users in the network. In the simulation, we find very high correlation (≈ 0.99) between density distribution of users and PageRank which justify the consideration of density distribution of users for the purpose of ordering or ranking.

…