A Random Walk Approach to Modeling the Dynamics of the Blogosphere

Muhammad Shafiq, Alex Liu

A Random Walk Approach to Modeling the Dynamics of the Blogosphere

M. Zubair Shafiq and Alex X. Liu

Department of Computer Science and Engineering
Michigan State University
East Lansing, MI 48824, U.S.A.
Email: {shafiqmu,alexliu}@cse.msu.edu

Abstract. It is important to develop intuitive and tractable generative models to simulate the topological and temporal dynamics of the blogosphere because these models provide insights about its structural evolution. In such generative models, independent instances of individual bloggers are initiated and these instances interact with each other to simulate the evolution of the blogosphere. Existing generative models of the blogosphere have certain limitations: (1) they do not simultaneously consider the topological and temporal properties, or (2) they utilize the global information about the blogosphere that is typically not available. In this paper, we propose a novel generative model for the blogosphere based on the random walk process that simultaneously considers both the topological and temporal properties and does not utilize the global information about the blogosphere. The results of our experiments show that the proposed random walk based model successfully captures the scale-free nature of both topological and temporal dynamics of the blogosphere.

Keywords: Blogosphere, Random walks, Network Science

1 Introduction

1.1 Background

The blogosphere constitutes an important niche of online social networks. It consists of blogs and each blog usually contains several posts. As shown in Figure 1, the blogosphere can be envisioned as consisting of two separate networks: (1) blog network, and (2) post network. In these networks, nodes represent blogs and posts respectively; whereas edges are directed hyperlinks from a source node to a destination node.

In recent years, researchers have carried out independent measurement-driven studies to understand various global properties of the blogosphere [6, 7, 9, 12]. These studies have reported several interesting global patterns in the blogosphere, such as distributions of post in-degree, blog in-degree, size of cascades, inter-posting times, and post popularity over time follow power-laws. These discoveries have important implications in terms of evolution of network graphs and information flow, search, and retrieval.

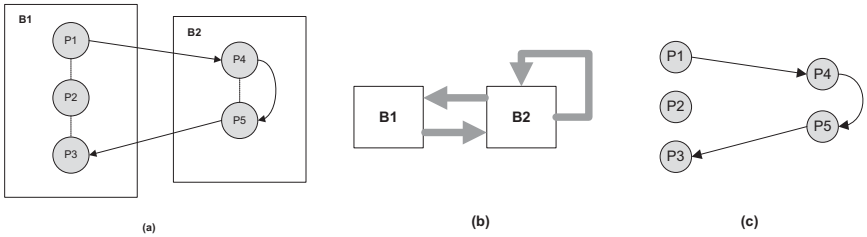


Fig. 1. (a) The blogosphere can be visualized as two networks: (b) blog network, (c) post network

1.2 Problem Statement

A significant amount of research effort has been put in to understand the global properties of the blogosphere. However, little is yet known about the underlying local mechanisms that have led to these topological and temporal patterns at the global scale. One way to understand these local mechanisms is to develop a formal model of individual bloggers in the blogosphere. This model is then replicated for all bloggers and allowed to execute over a given period of time. Finally, global properties of the evolved synthetic network graph are computed and matched to those of real network graphs. A high degree of correlation in compared metrics validate the developed model. In this paper, we aim to develop a realistic, intuitive, and tractable generative model for the blogosphere.

1.3 Proposed Approach

In this paper, we present a generative model based on the random walk process to simulate the evolution of the blogosphere. This model is intuitive and it produces realistic blogosphere networks whose topological and temporal properties match to those discovered in real-world networks. It has the following desirable properties. First, it controls the subtle structural properties of the evolved graph. Second, it does not utilize any global information that may not be locally visible to individual bloggers. Finally, the scale-free distributions evolve only *implicitly* as our model does not explicitly bias link creation probability to nodes' attributes.

1.4 Key Contributions

We summarize our key contributions as follows:

1. We propose a novel multiple random walk based model that can evolve realistic blogosphere networks by mimicking the micro-level interactions of individual bloggers under real-world constraints. To the best of our knowledge, this is the first generative model that *utilizes only locally visible information* to accomplish this task.
2. We propose a novel methodology to quantify the burstiness of blogging activity. Towards this end, we provide a formal analysis of the burstiness of a finitely-bounded, one-dimensional random walk.

3. Our proposed model has the ability to control different structural properties of the generated blogosphere network by varying the length of the random walk.

The rest of the paper is organized as follows. We review the related work in Section 2. In Section 3, we provide background and overview of the proposed random walk based model. In Section 4, we present formal analysis and details of our proposed generative model. In Section 5, we provide experimental results. Finally, we give concluding remarks in Section 6.

2 Related Work

Several research studies have recently focused on the formal modeling of the blogosphere. Some studies have focused on developing formal models of individual bloggers that can evolve graphs with properties similar to those of the real-world blogosphere graphs. Researchers have also tried to model the topological and temporal patterns of the blogosphere. The relevant research studies can be categorized into three types based on the patterns that they analyze or model. We now present a brief overview of each category.

2.1 Topological Properties.

In [8], the authors proposed a generative model for the blogosphere called second space. Second space uses several parameters to generate graphs with properties similar to those of real-world graphs. The authors computed multiple topological properties of the simulated blogosphere such as degree distributions, diameter, reciprocity, *etc.* The model developed by second space does not simultaneously consider topological and temporal patterns, which is a major limitation. Moreover, it has a lot of input parameters, which is undesirable. In another study [10, 11], the authors used Kronecker graphs to model the structure of networks. They further proposed KronFit algorithm to fit a Kronecker graph generation model to real networks. However, this approach is not specifically designed for modeling the blogosphere.

2.2 Temporal Properties.

In [9], the authors performed an analysis of time evolution of the blogosphere using time graphs. They observed that the blogosphere expanded in a bursty manner. Using this observation, the authors created a graph model called the randomized blogspace. This model focuses on temporal patterns observed in the evolution of the blogosphere and does not generalize to other patterns.

2.3 Topological and Temporal Properties.

In [5], the authors have proposed the zero-crossing ($\mathcal{ZC}$) model to generate synthetic blogospheres. It is the first work to jointly model temporal and topological

properties of the blogosphere. This model also has an intuitive appeal as all procedures closely relate to the mentality of real-world bloggers. The blogosphere generated by $\mathcal{ZC}$ model conforms with several global patterns including but not limited to post in-degree, blog in-degree, and inter-posting times. However, $\mathcal{ZC}$ model has several limitations. First, It cannot control other important structural properties of the evolved graphs such as clustering, centrality, connected components, *etc.* Second, it uses global information (such as total number of in-links to a blog) that is not locally visible to individual bloggers. Finally, the power-laws are explicitly programmed into the model that reduces its intuitive appeal. For example, the probability of creating links to other posts is proportional to their number of in-links (defined as *preferential attachment* process which automatically leads to power laws). In a more intuitive model, a blogger will parse the existing network graph while probabilistically creating new links without any explicit bias.

3 Overview of the Proposed Model

In this section, we first provide the goals of our proposed model and then present its overview. It has been shown earlier that several topological patterns of real-world blogosphere, including blog degree distributions, post degree distributions, inter-posting times, *etc.* follow the power-law distribution [4, 15–17]. The first goal of our random walk based model is to implicitly produce these power-laws using only local interactions between individual bloggers. We also aim to reproduce the burstiness and self-similarity in the temporal dynamics of the blogosphere using the random walk process [18]. Finally, we want to control various properties of the evolved blogosphere without using multiple parameters.

We now provide an overview of our proposed random walk based model. The *random walk* process constitutes the core of our proposed model [1, 14]. To explain the random walk process, an analogy often cited in the literature is of a drunkard walking in city streets with junctions. At each junction, the drunkard chooses one of the possible paths randomly. The rationale behind using the random walk procedure to model the dynamics of the blogosphere is to mimic the local behavior of bloggers in blog reading and writing. When multiple instances of the same random walk based model are invoked, they interact with each other to generate a blogosphere network.

Figure 2 shows an overview of the proposed random walk based model called $\mathtt{m}\mathcal{RW}$. Each instance of $\mathtt{m}\mathcal{RW}$ model corresponds to an individual blogger. $\mathtt{m}\mathcal{RW}$ model consists of five modules: $\mathcal{RW}$-1, $\mathcal{RW}$-2, $\mathcal{RW}$-3, $\mathcal{RW}$-4, and $\mathcal{RW}$-5. $\mathcal{RW}$-1 is a random walk on the discrete one-dimensional number line and aims to mimic the posting behavior of a blogger. A blogger creates a post P every time the zero mark on the number line is crossed.

A blogger has two options after creating a post in $\mathcal{RW}$-1. Either with probability p_l, a blogger decides to include out-links in the created post. Or with probability $1 - p_l$, a blogger publishes a post without any out-links. If a blogger decides to include out-links in the created post, it further has to choose between $\mathcal{RW}$-2 and $\mathcal{RW}$-3, with probabilities $1 - p_e$ and p_e respectively. In $\mathcal{RW}$-2, also

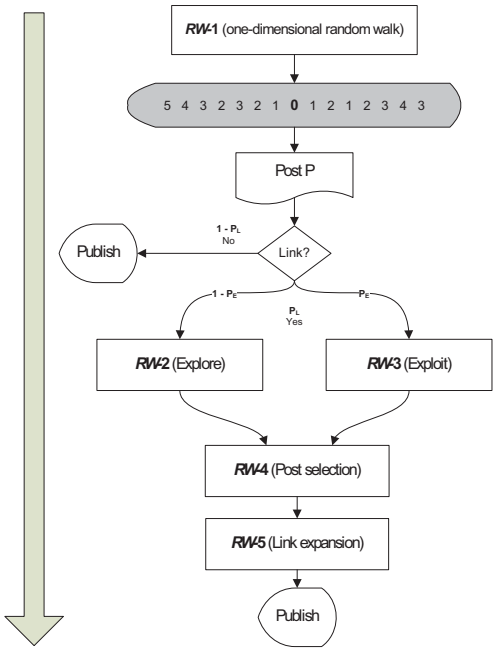


Fig. 2. Overview of the proposed random walk based model

called *explore mode*, a blogger aims to create a link to the blog to which it has not linked earlier. In $\mathcal{RW}$-3, also called *exploit mode*, a blogger aims to randomly revisit a previously linked blog. In contrast to $\mathcal{RW}$-1, $\mathcal{RW}$-2 and $\mathcal{RW}$-3 are random walks on the blog graph. At each step of these random walks, a blogger moves to one of the neighbors of the current blog graph node and the blog reached at the end of a finite length random walk is selected for link creation. Furthermore, note that p_l and p_e are randomly sampled from a uniform distribution for every blogger and thus are not input parameters of our proposed model.

After a blog is selected for link creation, a blogger has to choose one of its posts by $\mathcal{RW}$-4. $\mathcal{RW}$-4 is a random walk carried out on the post list of the selected blog starting at the most recent post. The post reached at the end of a finite length random walk is eventually linked in the post P. Finally in $\mathcal{RW}$-5, also called *link expansion*, the blogger conducts a random walk to further link to an arbitrary number of posts referred by the selected post. In contrast to $\mathcal{RW}$-2 and $\mathcal{RW}$-3, $\mathcal{RW}$-4 and $\mathcal{RW}$-5 are carried out on the post graph rather than the blog graph.

In the next section, we provide the details and formal analysis of all five random walks. Here, we iterate that the only input parameter of our $\mathtt{m}\mathcal{RW}$ model is the length of the random walk, which controls the properties of the generated blogosphere networks. Moreover, our proposed model simulates local

interactions between individual bloggers without utilizing any global information that is not available in the real-world.

4 Formal Analysis of the Proposed Model

4.1 Formal Definitions and Notations

We first formally define the basic concepts and notations that will be utilized in our $\mathtt{m}\mathcal{RW}$ model. The blogosphere can be mapped to two separate directed and simple graphs, blog graph $\mathbb{B} = (V_B, E_B)$ and post graph $\mathbb{P} = (V_P, E_P)$. V_B and V_P are sets of *vertices* and E_B and E_P are sets of *edges* for blog and post graphs, respectively. The elements of sets E_B and E_P (edges) are ordered pairs of the form $e = (v_1, v_2)$, where v_1 is source vertex and v_2 is destination vertex, because both $\mathbb{B}$ and $\mathbb{P}$ are directed graphs. A *path* $\mathcal{P}$ of finite length l from vertex u to v on graph is defined as the sequence of edges $e_1, e_2, ..., e_n$. A path does not have any repeated edges. A *walk* $\mathcal{W}$ is also defined as a sequence $e_1, e_2, e_3, ..., e_{n-1}, e_n$, however, it can have repeated edges. Let $E_i = \{e_1, e_2, ..., e_j\}$ denote the set of all possible edges for i^{th} step $(0 < i \leqslant l)$ of a walk. A walk is called *random walk* $(\mathcal{RW})$ if, at every step i of the random walk, the next edge is uniformly randomly selected from E_i. We now separately provide formal analysis of all five modules of our proposed $\mathtt{m}\mathcal{RW}$ model.

4.2 $\mathcal{RW}$-1: One-dimensional random walk modeling the temporal dynamics of posting behavior

$\mathcal{RW}$-1 module models the temporal dynamics of a blogger's posting behavior. $\mathcal{RW}$-1 is a random walk on a discrete number line starting at origin. Let $s(t)$ denote the position of a random walker at time tick t. A fair coin is flipped at each time tick $(t \geqslant 0, s(0) = 0)$. For all head outcomes the position is incremented by a unit $(s(t+1) = s(t)+1)$, and for tail outcomes the position is decremented by a unit $(s(t+1) = s(t)-1)$. We also finitely bound the length of random walk. In our initial analysis we assume that the bound is fairly large *i.e.* $n \to \infty$. For the time tick when current position crosses origin (*i.e.* $s(t) = 0$) a post is published by a blogger. This is also known as *zero-crossing*. Let t_{origin} denote this time tick. Consequently, inter-posting time is denoted by $t_{origin} - 0 = t_{origin}$. In the following text, we provide a mathematical analysis of two important properties of $\mathcal{RW}$-1.

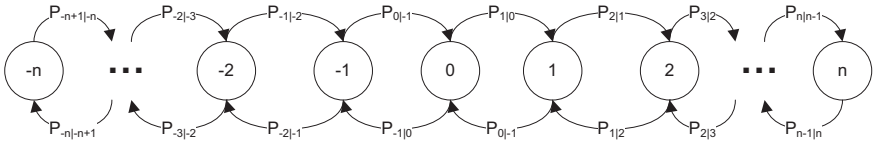
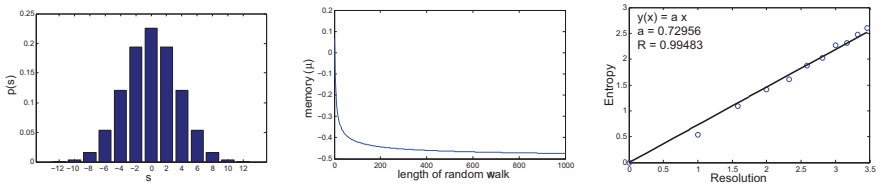


Fig. 3. $\mathcal{RW}$-1: One-dimensional random walk

Probability Distribution of Inter-posting Times: We want to find the probability distribution function of zero-crossing, *i.e.* returning to origin. Note that the probability of returning to origin is zero for odd number of time ticks. Also the probability of zero-crossing is $2^{-t}\binom{t}{t/2}$ for even number of time ticks.



(a) Probability distribution function of final position after $l(= 12)$ steps

(b) Variation in memory with respect to walk length

(c) Entropy plot to characterize self-similarity in posting times

Fig. 4. Properties of $\mathcal{RW}$-1 module

It can be shown after applying Stirling's formula to inter-posting time: $p \propto t^{-a}$, where a is a positive constant. This shows that the distribution of inter-posting time should follow a power law in the graphs generated by our model.

Burstiness of Inter-posting Times: To mathematically model burstiness, we model the probability distribution function of zero-crossing after l steps. Towards this end, we first model the random walk process using a Markov chain shown in Figure 3. Note that $n \gg l$ and conditional probabilities $P_{0|1}, P_{1|0}, ... = \frac{1}{2}$. It can be empirically shown using the Pascal's triangle that the probability distribution of landing at state s after l steps decays quickly for the increasing values of s (see Figure 4(a)) [14]. Equivalently, we can show using the Stirling's formula that: $P(s,l) = \frac{2}{\sqrt{2\pi l}} e^{\frac{-s^2}{2l}}$. Since we are interested in finding the probability distribution of zero-crossing, *i.e.* $s = 0$. Hence, we can simplify this relationship to: $P(0,l) = \frac{2}{\sqrt{2\pi l}}$.

In order to model the burstiness of zero-crossing process, we collapse all non-zero states to a single state for simplification. Moreover, we exploit the symmetry of the problem to combine positive and negative states. As a result, we get a 1^{st} order, 2 state, discrete time Markov chain. It is also known as the Gilbert-Elliot model [13]. The states here represent the position after l steps. Using the above-mentioned model, state 0 refers to origin on the number line and state 1 jointly models the rest of non-origin positions. A suitable measure to model the burstiness of the Gilbert-Elliot model was proposed in [13]. It is called memory μ and is defined as: $\mu = 1 - P_{0|1} - P_{1|0}$, where $-1 \leqslant \mu \leqslant 1$. $\mu = 0$ corresponds to zero memory, $\mu > 0$ corresponds to *persistent* memory, and $\mu < 0$ corresponds to *oscillatory* memory. Note that $P_{1|0} = P_{0|0} = 0.5$ and $P_{0|1} = 1 - P_{1|1}$; therefore, the relationship for μ can be reduced to: $\mu = P_{1|1} - \frac{1}{2}$. After plugging the probability of staying in state 1 after l steps from our previous analysis (while assuming homogeneity), we get: $\mu = \frac{2}{\sqrt{2\pi l}} - \frac{1}{2}$. Figure 4(b) shows the variation in memory of zero-crossing with respect to walk length. Note that $\mu \to -0.5$ as $l \to \infty$.

Given the posting times in a real-world blog network, a well-known method to empirically quantify the burstiness of inter-posting times is using the entropy plot. The entropy plot of a self-similar process is linear and the slope of the entropy plot quantifies the burstiness. The slope of the entropy plot varies in the

range $[0, 1]$, where 0 refers to a highly bursty process and 1 indicates a periodic (non-bursty) process. In [5], the authors showed that the slope of the entropy plot was approximately 0.88 for a sample of the blogosphere, which is in the middle of the two extremes. This observation could not be explained by their proposed model.

On the other hand, the burstiness of inter-posting times predicted by our proposed model varies as the function of the length of random walk. Therefore, we can tune this parameter to generate realistic blogosphere networks. To perform a one-to-one comparison between the burstiness predicted using our proposed model and that empirically measured using the entropy plot, we generate a sample blogosphere with a total of $100,000$ blogs, $500,000$ posts, and the length of random walk was set to 10. Figure 4(c) shows the entropy plot of inter-posting timings in a blogosphere network generated using $\mathtt{m}\mathcal{RW}$ model. It is evident from the regression fit that the plot is almost linear while the slope of the fit is about $0.7 < 1$, which is close to what was observed for the real-world blogosphere in [5]. Furthermore, our model can also establish the type of burstiness. The negative sign shows that the burstiness of zero-crossing process is in fact oscillatory. This shows that the probability of remaining in a given state is lower than the steady-state probability of being in that state [13].

4.3 $\mathcal{RW}$-2/3: Random walk on blog graph (explore/exploit)

$\mathcal{RW}$-2 and $\mathcal{RW}$-3 modules model the linking behavior of a blogger. Some previous studies have proposed to choose blogs proportional to their in-degree [5]. However, in real-world scenarios, a blogger does not know all blogs linking to a given blog. Rather, only a small fraction of in-linking blogs may be known. In fact, many well-known blogging platforms, such as `Blogger` and `Wordpress`, explicitly provide some of the in-linking blogs. We conclude that the total number of in-links to a blog is a global information and is not locally visible to an individual blogger. Therefore, a realistic and intuitive generative model of the blogosphere should only use information locally visible to individual bloggers to mimic their linking behavior. In this paper, we propose to use random walk on the directed blog graph to model the linking behavior of bloggers while using only locally visible information. The properties of random walks on graphs have been well-studied in previous research [16, 17].

$\mathcal{RW}$-2 module mimics the explore mode of a blogger and initiates at a randomly chosen vertex v in the blog graph. The blog reached at the end of a random walk is selected. The length of random walk l for every blogger is selected at the start of blogging activity. For the i^{th} step of random walk ($i \leqslant l$), let $\mathbb{O}_v(i)$ and $\mathbb{I}_v(i)$ denote the set of all outgoing and incoming links respectively for vertex $v(i)$. Also, let $\mathcal{I}(i) \in \mathbb{I}_v(i)$ denote the singleton set containing the incoming traversal link for i^{th} step of the random walk. The outgoing traversal link at step i is uniform-randomly chosen from the set of links $\mathbb{S}_i$, which is defined as: $\mathbb{S} = \mathbb{O}_v(i) \cup \mathcal{I}(i)$. $\mathcal{I}(i)$ provides 'immediate backtrack' functionality via incoming traversal link.

$\mathcal{RW}$-3 module captures the "exploit" mode of a blogger. It is the same as $\mathcal{RW}$-2 module except for the fact that the walk always starts at the correspond-

ing node of blogger conducting the walk. We show later in Section 4.6 that the probability of link creation in both $\mathcal{RW}$-2 and $\mathcal{RW}$-3 is proportional to in-degree while using only locally visible information.

4.4 $\mathcal{RW}$-4: Random walk on reverse chronological post list

In $\mathcal{RW}$-4 module, a blogger chooses a post from the blog selected in modules $\mathcal{RW}$-2 or $\mathcal{RW}$-3. The post list is typically ordered in the reverse-chronological order and the links between posts are undirected. The length of random walk l in $\mathcal{RW}$-4 is the same as used in $\mathcal{RW}$-1/2/3. The random walk is always initiated at the post $P1$ which is denoted by $v(1)$. Note that $deg(v(i)) = 2, \forall i \in \{2, 3, ..., l-1\}$.

The formal analysis of $\mathcal{RW}$-4 is similar to that of $\mathcal{RW}$-1. $\mathcal{RW}$-4 is essentially one-sided subset of $\mathcal{RW}$-1 if $l \gg k$, where k is the size of post list. The distribution of final position s is similar to the one plotted in Figure 4(a). It is evident that the more recent posts have higher probability of being selected for link creation and this procedure is reflective of real-world behavior of bloggers.

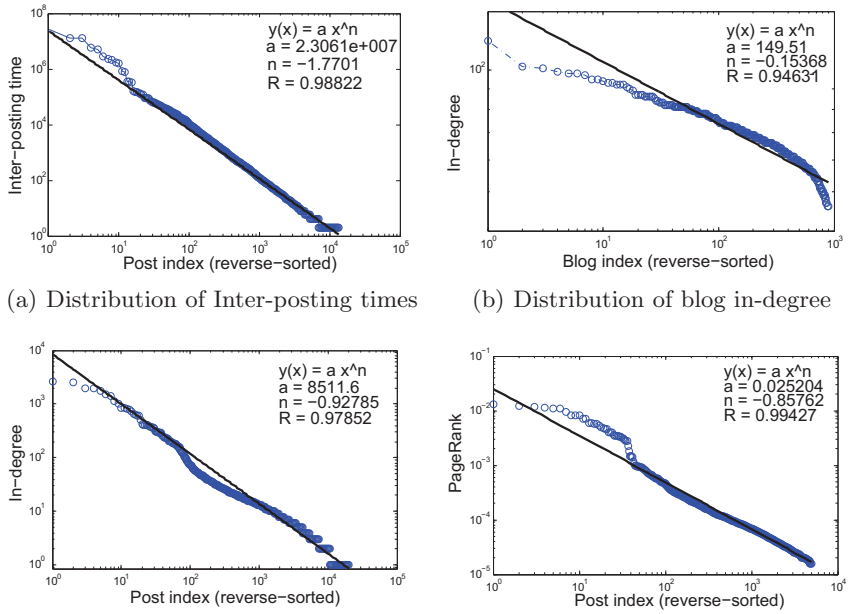
4.5 $\mathcal{RW}$-5: Random walk on post graph (link expansion)

$\mathcal{RW}$-5 module models the link expansion behavior, which is found in real-world blogs. In link expansion, a blogger recursively refers to some out-links of the selected post. In $\mathcal{RW}$-5 module, a blogger probabilistically creates links to posts linked by the post that is selected by $\mathcal{RW}$-4. This random walk is initiated at the selected post on the out-directed post graph. The length of random walk l is the same as used in all other random walks. For the i^{th} step of random walk ($i \leqslant l$), let $\mathbb{O}_v(i)$ denote the set of all outgoing links for vertex $v(i)$. Also, let $\mathcal{I}(i) \in \mathbb{I}_v(i)$ denote the singleton set containing the incoming traversal link for i^{th} step of random walk. The outgoing traversal link is uniform-randomly chosen from the set of links $\mathbb{S}_i$ which is defined as: $\mathbb{S}_i = \mathbb{O}_v(i) \cup \mathcal{I}(i)$. We provide the formal analysis for $\mathcal{RW}$-5 in Section 4.6.

4.6 Formal analysis of $\mathcal{RW}$-2/3/5

We now provide a formal analysis of random walks on blog and post graphs in modules $\mathcal{RW}$-2, $\mathcal{RW}$-3, and $\mathcal{RW}$-5. Note that random walk procedure uses only locally visible information. The goal of our formal analysis is to show that the random linking via random walks is equivalent to the well-known preferential attachment principle found in real-world blogs [15].

Towards this end, let $P(v)$ denote the probability that vertex v is chosen for linking at the end of random walk which starts from vertex w out of total N vertices. Also let $P(w)$ denote the probability that vertex w is chosen for start of random walk, so $P(w) = 1/N$. Using Bayes rule, we can show that: $P(v) = \frac{P(v|w)P(w)}{P(w|v)}$. We note that the destination vertex at every step of random walk is chosen randomly from the set of existing links, *i.e.*, $P(v|w) = 1/deg(w)$. Similarly, it can be shown that $P(w|v) = 1/deg(v)$. We combine these observations to get the following result: $P(v) = \frac{deg(v)}{Ndeg(w)}$. Note that if w is chosen randomly then $deg(w) = \mu_{deg}$, where μ_{deg} is the average degree of the graph which is constant. We conclude that: $P(v) \propto deg(v)$. This result is essentially the mathematical formulation of the preferential attachment principle.



(a) Distribution of Inter-posting times (b) Distribution of blog in-degree

(c) Distribution of post in-degree (d) Distribution of PageRank

Fig. 5. Properties of Static Blog Snapshots

5 Experimental Results

In this section, we gauge the properties of the blogosphere networks generated using our proposed $\mathtt{m}\mathcal{RW}$ model. We analyze various properties of static snapshots of the generated blog networks.

5.1 Inter-posting time

In Section 4.2, our formal analysis showed that inter-posting times of simulated bloggers are self-similar and that the distribution of inter-posting times follows a power-law. Figure 5(a) shows the distribution of inter-posting times for a randomly selected blogger in our proposed model with length of random walk equal to 10. Note that we have sorted the post-index with respect to its inter-posting time. The data clearly follows a straight line on a log-log scale. The high value of goodness-of-fit parameter $R \approx 0.98$ for the regression line confirms that the distribution of inter-posting times follow a power law distribution. This indicates the fact that blogging activity can be characterized by long periods of inactivity separated by sudden and short periods of activity.

5.2 Blog in-degree

In Figure 5(b), we show the blog in-degree distribution of a randomly chosen generated blogosphere. Here we have sorted the blog-index with respect to its

degree. A power-law curve is fitted on the data at log-log scale. The high value of goodness-of-fit parameter $R \approx 0.95$ shows that we have reasonable confidence in this observation. This highlights the fact that only a few blogs receive large number of in-links and a majority of blogs remain unnoticed.

5.3 Post in-degree

Figure 5(c) shows the distribution of post in-degree for one of the blogospheres generated using $\mathtt{m}\mathcal{RW}$ model. A reasonable portion of the observed data follows the straight line of the log-log scale. The high value of goodness-of-fit parameter $R \approx 0.98$ highlights the scale-free nature of the graph evolved using our proposed model. Similar to our observation for blog in-degree, this also highlights that only a few posts get linked by a large number of other posts.

5.4 PageRank

PageRank is a well-known measure which forms the core of Google web search engine developed by Brin *et al.* [2,3]. PageRank value is essentially an importance weight assigned to every node in a network. The links from blogs with larger in-degrees are considered relatively more significant in PageRank computation. PageRank value is established from the principal eigenvector of the adjacency matrix of a graph. In our analysis, we have found that the distribution of PageRank values of blogs follows a power-law. Figure 5(d) shows the distribution of PageRank of blogs for one of the blogospheres generated using our proposed $\mathtt{m}\mathcal{RW}$ model. It is evident that the PageRank distribution follows a straight line on the log-log scale and has a high value of goodness-of-fit parameter $R \approx 0.99$.

6 Conclusion

In this paper, we have presented a novel and intuitive model using multiple random walks for generating the blogosphere. This model overcomes the limitations of previously proposed models. Furthermore, the results of our experiments show that the blogosphere generated using our proposed $\mathtt{m}\mathcal{RW}$ model possess well-studied properties of the real-world blog graphs. In future, we plan to explore the control provided by our proposed model over other structural properties of the generated blogosphere using a single input parameter *i.e.* the length of random walk.

The successful modeling of topological and temporal dynamics of the blogosphere using random walk has several several interesting applications: For example, graph extrapolation can be done by predicting the trajectory of individual random walkers (bloggers) constituting the blogosphere. The only parameter to be estimated in this regard is the length of random walk. (2) The developed model can also be utilized to study and predict the effect of *active probing* (*e.g.* applying certain constraints on link creation) on topological and temporal properties of networks.

Acknowledgements

The authors would like to thank Dr. Habib Salehi for valuable comments and suggestions on an initial draft of the paper.

This material is based in part upon work supported by the National Science Foundation under Grant Number IIS-0968495. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Aldous, D., Fill, J.: Reversible Markov Chains and Random Walks on Graphs. Book Draft (2001)
2. Brin, S., Pag, L.: The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems 33, 107–117 (1998)
3. Brin, S., Page, L., Motwami, R., Winograd, T.: The pagerank citation ranking: bringing order to the web. Tech. rep., Computer Science Department, Stanford University (1998)
4. Evans, T.S., Saramaki, J.P.: Scale-free networks from self-organization. Physical Review E 72(026138) (2005)
5. Gotz, M., Leskovec, J., McGlohon, M., Faloutsos, C.: Modeling blog dynamics. In: AAAI ICWSM (2009)
6. Gruhl, D., Guha, R., Liben-Nowell, D., Tomkins, A.: Information diffusion through blogspace. In: WWW. pp. 491–501 (2004)
7. Gurzick, D., Lutters, W.G.: From the personal to the profound: understanding the blog life cycle. In: CHI. pp. 827–832 (2006)
8. Karandikar, A., Java, A., Joshi, A., Finin, T., Yesha, Y., Yesha, Y.: Second space: Generative model for the blogosphere. In: AAAI ICWSM. pp. 198–199 (2008)
9. Kumar, R., Novak, J., Raghavan, P., Tomkins, A.: On the bursty evolution of blogspace. In: WWW. pp. 568–576 (2003)
10. Leskovec, J., Chakrabarti, D., Kleinberg, J., Faloutsos, C., Ghahramani, Z.: Kronecker graphs: An approach to modeling networks. arXiv:0812.4905v2 (2009)
11. Leskovec, J., Faloutsos, C.: Scalable modeling of real graphs using kronecker multiplication. In: ICML (2007)
12. Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., Hurst, M.: Cascading behavior in large blog graphs. In: SDM (2007)
13. Mushkin, M., I.B-David: Capacity and coding for the gilbert-elliot channels. IEEE Transaction on Information Theory 35(6), 1277–1290 (1989)
14. Rudnick, J., Gaspari, G.: Elements of the random walk: an introduction for advanced students and researchers. Cambridge University Press (1944)
15. Saramaki, J., Kaski, K.: Scale-free networks generated by random walkers. Physica A: Statistical Mechanics and its Applications 341, 80–86 (2004)
16. Vazquez, A.: Knowing a network by walking on it: emergence of scaling. Statistical Mechanics (2000)
17. Vazquez, A.: Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. Physical Review E 67(056104) (2003)
18. Vazquez, A., Oliveira, J.G., Dezso, Z., Goh, K.I., Kondor, I., Barabasi, A.L.: Modeling bursts and heavy tails in human dynamics. Physical Review E 73(036127) (2006)