A Locality Model of the Evolution of Blog Networks

Mark Goldberg[*], Malik Magdon-Ismail[†], Stephen Kelley[‡], Konstantin Mertsalov[§]

[*]goldberg@cs.rpi.edu, [†]magdon@cs.rpi.edu, [‡]kelles@cs.rpi.edu, [§]mertsk2@cs.rpi.edu

Computer Science Department, RPI Troy, New York 12210

Abstract—We present a model of evolution of large social networks. Our model is based on the local nature of communication: a node's communication energy is spend mostly within it's small social area. We test our model on the Blog network hosted by LiveJournal. Our testing with different definitions of local areas shows that the best approximation to the observed statistics occurs when the area is defined as the union of clusters containing the node, and when nodes communicate with other nodes within their area using a preferential attachment strategy. We define clusters as locally maximally dense sets of vertices; our definition allows for clusters to overlap.

I. INTRODUCTION

A probabilistic model of the internal processes in the Blogosphere is needed for the development of a network generator to be used for testing hypotheses regarding the future activity of the Blogosphere. Using multiple simulations, one learns the expected level of the "life" of the Blogosphere, and then one is able to recognize unexpected deviations in the real-life network.

The structure of large social networks, such as the WWW, the Internet, and the Blogosphere has been the focus of intense research during the last decade (see [1], [6], [8], [12], [15], [17], [18], [19], [20]). One of the main foci of this research has been the development of dynamic models of network creation ([2], [11], [20], [16]) which incorporates two fundamental elements: network growth, with nodes arriving one at a time; and some form of preferential attachment in which an arriving node is more likely to attach itself to a high degree existing node than a low degree one (*the rich get richer*).

Once a network has been created, how does it evolve? This question has not been adequately addressed in previous literature, partially because the typical networks studied–the WWW, the Internet, collaboration networks,–mainly exhibit growth dynamics, and not *evolution dynamics*. In networks such as the WWW or the Internet, the growth dynamic is dominant. Models of these networks have been able to capture many interesting properties, in particular, scaling behavior with power-law degree distributions (the probability of a node having degree k is given by $P(k) \propto k^{-\alpha}$, where $\alpha \in [2, 3]$ for real networks). Analysis of the LiveJournal Blogosphere shows that the rate of growth of the network is significantly smaller than the rate at which new links are formed. In fact, over a 9 week period in 2006, the number of active bloggers remained approximately constant. However, looking at the communication between bloggers reveals that out of the total 400,000 unique pairs of bloggers that communicated via the blog comments in an arbitrary week, 240,000 were of those pairs did not exist in the previous week. Thus, it is the evolution dynamics of the Blogosphere which are dominant.

Yet, even in such an active environment, certain statistics are remarkably stable. For example: the power-law coefficient for the in-degree distribution, the clustering coefficient, and the size of the giant component (see Figure 1).

In this paper, we present a novel model for the evolution dynamics of social networks which supports public communication (communication which is visible to all members of the network). Though our model is general, it is particularly applicable to Blog-networks, which is the domain on which we do the testing.

We use a directed graph, the blogograph, to represent the communication activity in the Blogosphere. The nodes of the blogograph are bloggers and the edges are ordered pairs of bloggers. The edge $A \rightarrow B$ reflects the fact that blogger A visited the blog of blogger B and left a message there. The evolution of of the Blogosphere is represented by a sequence of such graphs and each next graph is generated based on the previous one. Our model is based on three fundamental principles for describing evolution dynamics in social networks:

Locality. Every blogger is aware of her local neighborhood. This local neighborhood could for example describe her interests, her social groups, etc. It is within this local neighborhood that the blogger will spend the bulk of her communication energy.

Attachment. Having determined her locality, the blogger spends her communication energy predominantly within her locale. In the context of Blog-networks, communication energy is expended through commenting on other blogs. We define the process by which a blogger distributes her communication within her area as the *attachment mechanism*.

Iteration. The evolution of the network is attained through repeated application of steps 1 and 2. After all bloggers have expended their communication energy in time step t, at time step $t + 1$ all bloggers re-evaluate their areas (for example, a blogger may discover some new blogs within her interests) and then re-attach their communication links according to the attachment model.

The definition of a model reduces to specifying the following two conditions:

(a) How to identify a blogger's area;
(b) What the attachment mechanism is.

There are a number of intuitively plausible models that one could identify, and the test of any such model has to be done against the observed Blog-network. In particular, does the convergent dynamics of the model reproduce, to a desirable accuracy, the observed stable properties of the real-world blogograph? The aim of this paper is twofold. The first

w	$|V|$	$|E|$	GC	C	d	α
35	111,248	376,878	96.0%	0.0788	5.336	2.87
36	118,664	411,294	96.0%	0.0769	5.327	2.74
37	120,549	410,735	96.0%	0.0752	5.375	2.79
38	119,451	386,962	95.8%	0.0728	5.455	2.82
39	113,296	323,062	95.2%	0.0666	5.641	2.80
40	124,329	430,523	96.3%	0.0764	5.332	2.77
41	121,609	380,773	95.9%	0.0705	5.471	2.81
42	124,633	415,622	96.2%	0.0739	5.349	2.74
43	123,898	403,309	96.5%	0.0713	5.425	2.81

Fig. 1. Statistics for observed blogograph: order of the graph ($|V|$), graph size ($|E|$), fraction of vertices that are part of giant component (GC size), clustering coefficient (C), average separation (d), power law exponent (α)

is to present a set of models which are instances of our framework, and the second is to test the validity of these models against the observed Blog-network. Our approach is to experimentally simulate the evolution according to the model until it converges to an equilibrium steady state, and compare the statistics of the resultant simulated blogograph with the observed blogograph. As we will demonstrate, the model based on defining areas using clusters (defined below) combined with the preferential attachment mechanism yields the locality-based model which best describes the evolution of Blog-networks. The other models we tested, in order of performance, were: several variations of the two-neighborhood areas with uniform and preferential attachment; and the global area with uniform and preferential attachment, which can be viewed as an extension of the preferential attachment growth model to a model for evolution dynamics.

We break down the paper as follows. First, we give an overview of the LiveJournal Blogosphere, which is the source of our data. We then describe the locality based models which we will consider and describe how we test their validity. Finally we present the experimental results using all the models, and end with some concluding remarks and future directions.

II. LiveJournal Blog Data

We define the blogograph to represent the communication within a fixed time-period; for our experiments, one week. The blogograph is a directed unweighted graph with a node for every blogger and a directed edge from the author of any comment to the owner of the blog where the comment was made during the observed time period. Parallel edges are not allowed and a comment is ignored if the corresponding edge is already present in the graph. To study the evolution dynamics, we considered consecutive weekly snapshots of the network; the communication graph contains the bloggers that either posted or commented during this week and the edges represent the comments that appeared during the week.

Our data was collected from the popular blogging service LiveJournal. LiveJournal imposes few restrictions on communication. Much of the communication in LiveJournal is public, which allows for easy access, especially given the real time RSS update feature provided by LiveJournal that publishes all open posts that appear on any of the hosted blogs. In our experience, the overwhelming majority of comments appear on a post within two weeks of the posting date. Thus, our screen-scraping program visits the page of a post after it has been published for two weeks and collects the comment threads. We then generate the communication graph.

We have focused on the Russian section of LiveJournal. as it is reasonable but not excessively large (currently close to 250,000 bloggers) and almost self contained. We identify Russian blogs by the presence of Cyrillic characters in the posts. Technically this also captures the posts in other languages with Cyrillic alphabet, but we found that the vast majority of the posts are actually Russian. The community of Russian bloggers is very active. On average, 32% of all LiveJournal posts contain Cyrillic characters. Our work is based on data collected during September and October of 2006.

The observed communication graph has interesting properties. The graph is very dynamic (on the level of nodes and edges) but quite stable if we look at some aggregated statistics. About 70% of active bloggers will also be active in the next week. Further, about 40% of edges that existed in a week, will also be found in the next week. A large part of the network changes weekly, but a significant part is preserved. Some of the important parameters of the blogograph illustrating their stability are presented in Figure 1. The giant component (GC) is the largest connected subgraph of the undirected blogograph. A giant component of similar size was observed in other large social networks [16], [14]. The clustering coefficient (C) refers to the probability that the neighbors of a node are connected. The clustering coefficient of a node with degree k is the ratio of the number of edges between it's neighbors and $k(k-1)$. The clustering coefficient of the graph is defined to be the average of the node clustering coefficients. The observed clustering coefficient is stable over multiple weeks and significantly different from the clustering coefficient in a random graph with the same out-degree distribution, which is 0.00029. The average separation (d) is the average shortest path between two randomly selected vertices of the graph. We computed it by sampling 10,000 random pairs of nodes and finding the undirected shortest path between them. The blog communication graph is not significantly different with respect to this parameter from some other social networks [16], [21].

Many large social networks [2], [14] have a power law degree distribution, $P(k) \propto k^{-\alpha}$, where $P(k)$ is the probability a node has degree k. Figure 2 shows the in-degree distribution averaged over the collected blogographs. We observed power
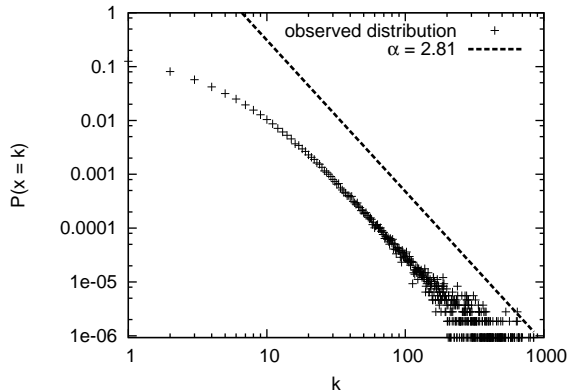
Fig. 2. Average in-degree distribution in the blogograph observed over 9 weeks from Aug. 29, 2006 and Oct. 29, 2006
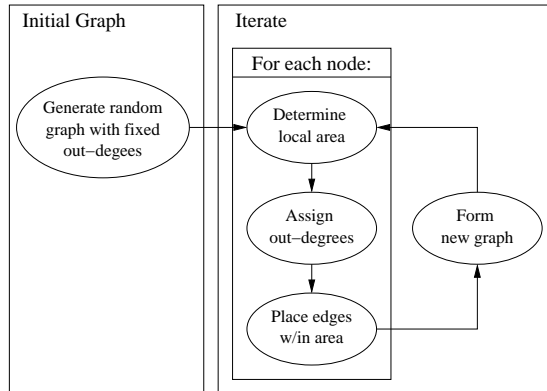

Fig. 3. Model execution flow

law tail with parameter $\alpha \approx 2.81$, which is stable from week to week. This value was computed using maximum likelihood method described in [10] and Matlab code provided by Aaron J. Clauset.

III. Locality Models of Evolution

Our goal is to model how the communication structure evolves, and equilibrates to its observed steady state. Our model specifies how nodes update their edges in response to the observed communication activity. In specifying this model of evolution, we take as given the out-degree distribution of the Blogograph. The justification for this is that, while the out-degree distribution would be an interesting object to model, it mainly reflects the individual properties of the users in the network (the level of energy and involvement of the user). Such quantities tend to be innate to a user (different people have different social habits: some manage to communicate with hundreds of people while others interact with only a small group), and hence out-degrees should be specified either *ab initio* (e.g. from social science theory) or extracted directly from the observed data. We will take the latter approach to specifying the out-degree distribution when it comes to testing

our model.

Given the out-degrees for all nodes, the task is now to specify how to attach the out-edges of the nodes, and in particular, to obtain the in-degree distribution. It is the in-degree distribution that characterizes the global communication structure of the network (for example, who is considered by others to be important). Clearly, the out-degree distribution of a graph alone does not determine its in-degree distribution. Algorithms for generating undirected random graphs with a prescribed degree distribution are well known (see [9], [13], [22]). However, even if those algorithms can be expanded to the domain of directed graphs, they would still be insufficient for our purpose of modeling evolution which requires repeated generation of the next graph given the previous one.

To summarize, we are interested in models which reproduce the observed evolution, *given* the out-degrees of the nodes. Thus, all our locality models assume that a node when deciding where to attach its communication links, has some fixed budget of emanating edges which it can attach. The main task of our model is to develop an evolution mechanism that re-creates an in-degree distribution close to the observed one.

We will use standard graph theory terminology in describing our model (see for example [23]). The sequence of Blogographs are represented by directed graphs $G_0, G_1, G_2, \ldots$, where at every time step t, $G_t = (V, E_t)$. V is the common vertex set of all known bloggers, $V = \{v_1, \ldots, v_n\}$. An edge (v_i, v_j) is in the edge set E_t if blogger v_i commented on a post by v_j during the time period t. We take one time period to equal one week, which appears to be the natural time scale in the blogosphere.

The input to the model is the set of out-degrees at time t for each vertex, $\{k_t^1, \ldots, k_t^n\}$, and G_{t-1}, the Blogograph at time $t - 1$. The output of the model is G_t, the Blogograph at time t. Our model is locality based: at time t, every node v_i identifies its area, and assigns it out-edges with destinations in its area.

More formally, denote the area of v_i at time t by $A_t^i \subseteq V$. A_t^i represents the locality of node v_i at time t. Typically, a node's locality at time t will depend on G_{t-1}, the Blpgograph at time $t - 1$. The attachment mechanism is probabilistic (for each node). Node v_i attaches its k_t^i out-edges according to its own probability distribution p_t^i, where $p_t^i(v)$ specifies the probability for node v_i to attach to node v for $v \in V$. The probability distribution p_t^i may depend on A_t^i and G_{t-1} (e.g. higher degree nodes may get higher probabilities). In particular, we assume that $\sum_{v \in A_t^i} p_t^i(v) = 1$, which corresponds to the assumption that every nodes expends all its communication energy within its local area. Since we do not allow parallel edges, if $k_t^i > |A_t^i|$, it is not possible for node v_i to expend all its communication energy within its local area A_t^i. In this case we assume that $k_t^i - |A_t^i|$ edges are attached uniformly at random to nodes outside its area, and the remaining edges are attached within its area.

The evolution model is illustrated in Figure 3. In more detail, the evolution model first obtains the out-degree distribution (which is exogenously specified); from G_{t-1} it computes

A_t^i and p_t^i for all nodes $v_i \in V$. For all nodes, it then attaches edges according to p_t^i resulting in new graph G_t. This entire process is iterated for a user specified number of time steps. The model is configured with a procedure that determines the out-degree of a given node in a given iteration, a procedure to determine the area based on the graph of the previous iteration and a procedure to determine the attachment probabilities according to a given attachment model. We will now discuss some approaches to defining the areas and the attachment probabilities. The procedure for obtaining the out-degrees is discussed in Section IV.

A. Locality Models

The node expends the majority of it's communication energy within it's local area. This captures the intuition that people mostly communicate with in a small group that contains friends, family, colleagues, etc. We propose the following definitions of a node's area:

Global. Every node v_i is aware of the whole network, the local area of v_i is $A_t^i = V$ at every time period t.

k-**neighborhood.** The local area A_t^i of node v_i at time t consists of all v_j such that the undirected shortest distance $\delta_t(v_i, v_j) \leq k$.

Union of clusters. The area is defined as the union of *clusters* containing a node. A cluster containing a node is a proxy for a social group that a node belongs to. Our definition of clusters ([4], [3], [5]) allows for clusters to overlap. First, we define the notion of a density of a set. For our application, we use the ratio of number of edges within the subset to the number of edges with at least one endpoint in the subset. Having defined the density, a cluster is then defined to be any subset of nodes that is locally maximal w.r.t. to the density. Our experiments show that clusters overlap quite frequently. This is expected in social networks, where the same member may belong to more than one community.

B. Attachment Models

Given the local area A_t^i of the node v_i at time t, the attachment model describes the probability $p_t^i(v)$ of occurrence of an edge (v_i, v) at time t for $v \in V$. We propose the following attachment modes:

Uniform. Node v_i attaches to any $v \in A_t^i$ with uniform probability, so that $p_t^i(v) = (|A_t^i|)^{-1}$ and $p_t^i(v) = 0$ for $v \notin A_t^i$.

Preferential Attachment. Node v_i attaches to any $v \in A_t^i$ with probability

$$p_t^i(v) \propto indeg_{t-1}(v) + \gamma \tag{1}$$

where $indeg_{t-1}(v)$ is the in-degree of vertex v in G_{t-1} and γ is a constant. Setting $\gamma = 0$ corresponds to a pure preferential attachment.

Markov Chain. To obtain the attachment probabilities for vertex v_i we simulate the particle traveling over undirected edges of graph G_t starting from the node v_i and randomly selecting edges to travel until it arrives at the first node $v_e \notin A_t^i$. Every time the particle arrives at some node $v_j \in A_t^i$,

the counter c_j^i is incremented. This simulation is repeated with out resetting the counters c_j^i a number of times and we determine the attachment probability as $p_t^i(v) \propto c_j^i$.

Inverse distance. Node v_i attaches to some node $v \in A_t^i$ with probability

$$p_t^i(v) \propto \frac{1}{\delta_{t-1}^\rho(v_i, v)} \tag{2}$$

where $\delta_{t-1}(v_i, v)$ is the shortest undirected distance between vertices v_i and v in G_{t-1} and ρ is a constant. We experimented with $\rho \in [1, 5]$.

The combination of the locality model and attachment model specifies the evolution model that, given the out-degree distribution, will produce a series of graphs that represent the Blogograph at different time periods.

IV. TESTING THE MODEL

In parallel to the design of the evolutionary procedures, we need to independently design the testing routines to be used to decide if those procedures are "good" or "bad." Our experiments with LiveJournal reveal the remarkable stability (over time) of the out-degree and in-degree distributions in the graph.

Any testing of a model should involve the comparison of selected parameters of the networks generated by the model with those of the observed networks. Frequently considered parameters include the degree distribution, clustering coefficients, the size of the giant component, and its diameter ([2]). Our decision to consider directed graphs led us to adopt the out-degree distribution as an input parameter to the model, and thus, our task was to develop a mechanism for graph generation that re-creates the in-degree distribution of the observed graph. The use of the in-degree distribution as a test of the model is justified since the in-degrees can not be inferred from the out-degree distribution.

To obtain the numerical value of the model's quality, we compare the in-degree distributions of the graphs generated by the model to the graphs observed in LiveJournal. To compare the distributions, for each iteration t of the experiment we compute the normalized in-degree distribution

$$D'_i(k) = \frac{D_i(k)}{T_i}$$

where $D_i(k)$ is the number of nodes with in-degree k and T_i is the total number of nodes with in-degree greater then 0. Similarly, we compute the normalized in-degree distribution $D'_{ob}(d)$ for the observed graph. The error measure between the generated and observed degree distribution is then defined as:

$$E_t = \sum_{k>0} |D'_t(k) - D'_{ob}(k)|$$

Notice, $E_t \in [0, 2]$ and lower values of E_t correspond to a closer match. To determine the steady state error of the model, we executed it for enough iterations until the error stabilized. For all combinations of local area definitions and attachment methods it took under 30 iterations for the model to stabilize.
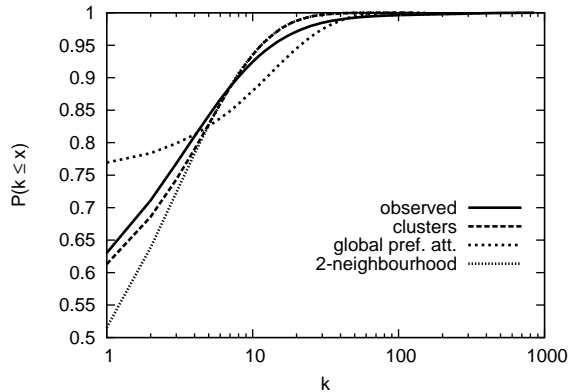
Fig. 4. Cumulative in-degree distributions for different model instances

Fig. 5. Generated in-degree distribution. $\alpha = 2.81$ corresponds to observed distribution.

To test the model we need to provide a mechanism for generating the out-degree for every node at each time period. We used blogographs observed at LiveJournal to develop such a mechanism. First, we introduced the type of a node to indicate the energy level of the node. We defined a type to be the average degree of the observed node over the 9 weeks and computed the distribution of types. Then, within each type we computed the distribution of out-degrees. The input to the model consisted of the distribution of types and the distribution of out-degrees conditioned on type. During the model initialization, each node was assigned some type that was selected randomly from the type distribution. After the initial assignment, the type of a node did not change for the duration of the model execution. Each time period the out-degree of a node was selected randomly from the distribution of out-degrees conditioned on the type of the node. This simulates the observation that the communication energy of a blogger expended during particular time period (in out case one week) depends on their personal activity level, which in our setup is the type.

We initialized the model to a random graph with assigned out-degrees, based on the distribution of the out-degrees observed in LiveJournal blogosphere.

V. EXPERIMENTAL RESULTS

We tested a variety of local area definitions and attachment methods. The global area model did not produce good results in combination with any of the attachment mechanisms. The combination of global area and the uniform attachment mechanism is a variation of Erdös-Rényi random graph. Such random graphs are known to produce an in-degree sequence that is very different from a power-law. Global area models combined with preferential attachment yielded a model similar to scale-free graphs model by Bollobás et al [7] and the Barabási-Albert scale-free model [1]. Our model is different from these due to the introduction of evolution. For $\gamma = 0$ (see equation (1)), the attachment probability is strictly proportional to the in-degree of the node in the previous iteration of the model. If the node had zero in-degree, then the probability of
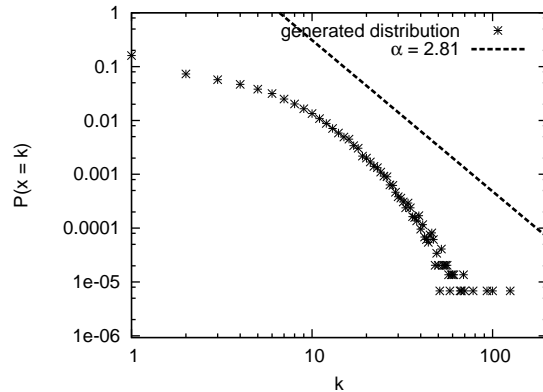
linking to this node is nearly zero (some nodes with area size smaller then out-degree still may select this node in random selection from the whole graph). For this reason the model with $\gamma = 0$ predictably tends to develop power houses - a small set of very high in-degree nodes that attract almost all the out-degree and very few nodes with non-zero out-degree. A distribution that contains "power houses" is considerably different from the observed power law distribution. Because of this, we did not pursue the combination of global area and preferential attachment further. For $\gamma = 1$ the model converged to error value of 0.89 after about 30 iterations. The in-degree distribution obtained from this combination is shown along with the observed and other distributions on Figure 4.

We experimented with global area combined with inverse distance attachment and attachment based on a combination of inverse distance and preferential attachment (in this combination the probability of attachment is computed as the product of the attachment probabilities of each component). The experiment was executed for different values of ρ (see equation 2). None of these instances of the model arrived at a reasonable steady state. Global area with the combination of inverse distance and preferential attachment tended to develop "power houses" for all exponents. The combination of global area and inverse distance produced an average error of 0.62 for $\rho = 2$. Other values of ρ that were tried produced greater error.

The k-neighborhood locality model with $k = 2$ improved the error considerably. Combining the random attachment method with 2-neighborhood area resulted in graphs which converged to the error value of 0.485 after about 10 iterations. The combination of 2-neighborhood with preferential attachment tended to a power house.

Next, we experimented with a cluster locality model. We used the following density function:

$$F(S) = \frac{|E_{in}(S)|}{|E_{in}(S)| + |E_{out}(S)|}$$

where $E_{in}(S)$ is a set of all edges (i, j) such that $i \in S, j \in S$ and $E_{out} = E(S, V \setminus S) + E(V \setminus S, S)$ where $E(A, B)$

	2-neighb.	clusters	global
pref. att.	0.814	**0.152**	0.822
random	0.485	0.759	1.148

Fig. 6. In-degree error for different experiment setups

Fig. 7. Edge history for different model instances. Cumulative distribution of the shortest distance in graph G_{t-1} of nodes that are connected by an edge in graph G_t.

denotes a set of all edges (a, b) with $a \in A, b \in B$. In other words, the density is the ratio of the number of edges inside the subset to the number of edges where either or both ends of inside the subset. The Iterative Scan algorithm [4] that was used to find clusters takes seeds as an input. These seeds are optimized by the algorithm into clusters which are locally optimal subsets with respect to a density function. We experimented with different seeding approaches including node seeds, edge seeds and the seeds found by Link Aggregate algorithm [4]. The best results were obtained using edges as seeds. We tested this area model with random attachment and preferential attachment methods. The model instance with clusters and random attachment converged to error 0.759 after about 25 iterations and the instance with preferential attachment converged to 0.152 after about 30 iterations. The latter combination produces the best results. The in-degree distribution obtained by this model instance is presented in Figure 5. The results of the most interesting combinations are presented in Figure 6.

To evaluate how well we simulate the evolution that takes place in the LiveJournal Blogosphere we devised the following metric called "edge history". For every edge (i, j) that appears in time step t we compute the shortest distance between i and j in graph corresponding to time step $(t - 1)$ and compute distribution of such distance for every edge that appears at time t. Intuitively this metric measures how far an individual blogger has to travel to leave a comment, which describes one of the core parameters that influences the evolution of the blogograph. The distribution of edge histories for different locality models with preferential attachment is given in Figure 7. It is evident from the plot that while the cluster-based area curve is somewhat different from the observed one, it follows its general shape. Alternatively, the global area model shows strong deviation which corresponds to the fact that bloggers have to travel far until they find the place to comment. In 2-Neighborhood model most bloggers only travel 2 steps, but a few whose area was smaller then their out-degree travel further.

REFERENCES

[1] R. Albert and A.-L. Barabási. Statictical mechanics of complex networks. *Reviews of Modern Physics*, 74(47-97), 2002.
[2] A. Barabási, J. Jeong, Z. Nëda, E. Ravasz, A. Shubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica*, A 311(590-614), 2002.
[3] J. Baumes, M. Goldberg, M. Krishnamoorthy, M. Magdon-Ismail, and N. Preston. Finding comminities by clustering a graph into overlapping subgraphs. *Proceedings of IADIS International Conference, Applied Computing 2005*, pages 97–104, 2005.
[4] J. Baumes, M. Goldberg, and M. Magdon-Ismail. Efficient identification of overlapping communities. *IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 27–36, May, 19-20 2005.
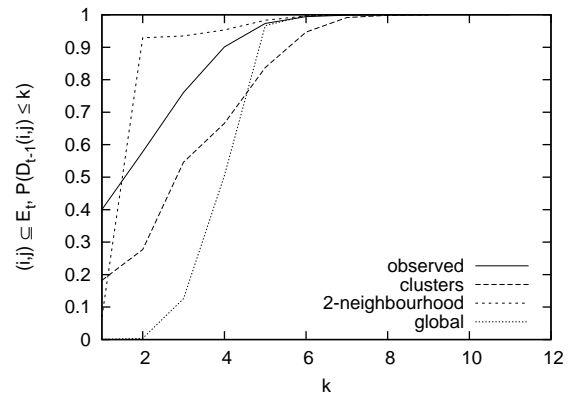[5] J. Baumes, M. Goldberg, M. Magdon-Ismail, and W. Wallace. Identification of hidden groups in communications. *Handbooks in Information Systems, Volume 2: National Security, 2007*, 2007.
[6] T. Berger-Wolf and J. Saia. A framework for analysis of dynamic social networks. *DIMACS Technical Report*, 28, 2005.
[7] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Directed scale-free graphs. In *SODA '03: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 132–139, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics.
[8] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stat, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33(1-6):309–320, 2000.
[9] F. Chung and L. Lu. Connected components in random graphs with given degree sequence. *Annals of Combinatoreics*, 6:125–1456, 2002.
[10] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data, 2007.
[11] P. Doreian and E. F.N. Stokman. Evolution of social networks. *Gordon and Breach*, 1997.
[12] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM*, pages 251–252, 1999.
[13] C. Gkantsidi, M. Mihail, and E. Zegura. The markov chain simulatiopn methods for generating connected power law random graphs. *Proc. of ALENEX'03*, pages 16–50, 2003.
[14] K.-J. Goh, Y.-H. Eom, H. Jeong, B. Kahng, and D. Kim. Structure and evolution of online social relationships: Heterogeneity in unrestricted discussions. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 73(6):066123, 2006.
[15] J. M. Kleinberg and S. Lawrence. The structure of the web. In *Science*, pages 1849–1850, 2001.
[16] G. Kossinets and D. J. Watts. Empirical analysis of an evolving social network. *Science*, 311:88–90, 2006.
[17] R. Kumar, J. Novak, P. Raghavan, and A. Tomkins. Structure and evolution of blogospace. *Communications of the ACM*, 33(1-6):309–320, 2004.
[18] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In *KDD'06*, 2006.
[19] M. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.
[20] M. Newman, A.-L. Barabási, and D. Watts. The structure and dynamics of networks. *Princeton University Press*, 2006.
[21] M. E. J. Newman. The structure of scientific collaboration networks. *PROC.NATL.ACAD.SCI.USA*, 98:404, 2001.
[22] A. O. Stauffer and V. C. Barbosa. A study of the edge-switching markov chain method for the generation of random graphs. `arxiv: cs. DM/0512105`, 2006.
[23] D. B. West. Introduction to graph theory. *Prentice Hall, Upper Saddle River, NJ*, 2003.