Noname manuscript No.
(will be inserted by the editor)

Locality and Attachedness-based Temporal Social Network Growth Dynamics Analysis

A case study on evolving nanotechnology scientific networks

Haizheng Zhang[1], **Baojun Qiu**[2], **Kristinka Ivanova**[3], **C. Lee Giles**[3], **Henry C. Foley**[3], **John Yen**[3]

[1] Microsoft

[2] Department of Computer Science and Engineering, The Pennsylvania State University

[3] College of Information Sciences and Technology, The Pennsylvania State University

Received: date / Revised version: date

Abstract The rapid advancement of nanotechnology research and development during the past decade presents an excellent opportunity for a scientometric study because it can provide insights on the dynamic growth of the fast evolving social networks associated with this exciting field. In this paper we conduct a case study on nanotechnology in order to discover the dynamics that govern the growth process of rapidly advancing scientific collaboration networks. This paper starts with the definition of temporal social networks and demonstrates that the nanotechnology collaboration network, in resemblance of other real-world social networks, exhibits a set of intriguing static and dynamic topological properties. Inspired by the observations that in collaboration networks new connections tend to be augmented between nodes in proximity, we explore the locality factor and the attachedness factor in growing networks. In particular,

we develop two distance-based computational network growth schemes, namely *DG* and *DDG*. The *DG* model considers only locality element while the *DDG* is a hybrid model that factors into both locality and attachedness elements. The simulation results of these models indicate that both clustering coefficient rates and the average shortest distance are closely related to the edge densification rates. In addition, the hybrid *DDG* model exhibits higher clustering coefficient values and decreasing average shortest distance when the edge densification rate is fixed, which implies that combining locality and attachedness can better characterize the growing process of the nanotechnology community. Based on the simulation results we conclude that social network evolution is related to both attachedness and locality factors.

Key words Social network analysis, growth dynamics, knowledge engineering

1 Introduction

Nanotechnology is a highly interdisciplinary field that is generally concerned with the control of matter on the molecular level in scales smaller than 1 micrometer, normally 1 to 100 nanometers, and the fabrication of devices within that size range. This field has been growing at an astounding pace in the last decades. This is reflected in the worldwide growth of funding from both government and industry, the increasing penetration into other disciplines, as well as the accelerating growth in the number of scientific publications and involved researchers. The explosive development of this field makes it ripe for in-depth scientometric analysis for this field. This paper conducts a study on the evolving nanotechnology collaboration network in order to develop insights into its social network growth dynamics.

This study is in line with the surging interests in social network and complex network studies in the recent decades. In particular, researchers have discovered that many social networks and other real-world complex networks exhibit a set of properties that distinguish them from random networks such

as *Erdos-Renyi* model [5]. These properties fall into two categories. The first category consists of a set of static topological properties that characterizes social network graphs, including (1) power-law degree distribution, (2) large clustering coefficient values, and (3) small average shortest path between two random nodes. In particular, the power-law degree distribution is a distinct feature of scale-free networks; large clustering coefficient values usually imply manifest community structures; and small average shortest path length values indicate short average separation between nodes. The latter two features constitute the so-called small-world network properties.

The second category is the kinetic properties exhibited in the growing process of social networks. For instance, researchers have reported the "shrinking diameter" phenomenon that the diameters of many real networks decrease over time [4,16,18]. These static and dynamic properties represent a significant departure from random networks. While these phenomena have been identified, there is no consensus on the cause of these features. Recently there has been a flurry of efforts by researchers from different disciplines exploring a variety of factors, including the attachedness factor (the degree of nodes) and the locality factor, to discover the growth schemes of social networks [3,10,17]. Section 2 provides an overview of the existing studies.

The nanotechnology collaboration network studied in in this paper, *NanoSCI*, is appealing for investigating social network growth dynamics for the following three reasons. Firstly, collaboration networks have been widely used in scientometrics and social networks study. It has been discovered that collaboration networks possess many static and dynamic properties that are similar to other social networks. In his early work on this domain, Newman studied several large collaboration networks and concluded that these networks exhibit all the general ingredients of small world networks, including short node-to-node distance and large clustering coefficient [21,22]. Moreover, researchers have recently shown that evolving collaboration networks exhibit similar dynamic patterns as do other social networks in the growth pro-

cess, such as shrinking diameters and high clustering coefficient values [4]. Secondly, *NanoSCI* offers one of the most extensive databases to date on social networks, including 292, 323 researchers and 368, 511 papers that are indexed by SCI (Science Citation Index) database [1] spanning from 1980-2006. Finally, the history of nanotechnology research is very short and the literature has developed so recently that the majority of its literature is on-line. Compared to other fields even new ones such as biotechnology or super conductivity, the short history of the field combined with its fully on-line character facilitates this kind of meta-scientific study. Thus *NanoSCI* provides unique research opportunities for us to investigate the characteristics of the formation stage of collaboration networks.

In particular, this paper reports a set of static properties and dynamic patterns observed in the evolving nanotechnology collaboration network. Based on these observations, we explore the joint effect of attachedness (degree) factor and locality factor for network growth dynamics. This paper proposes two distance-based computational growth schemes, namely *DG* (distance-based growth model) and *DDG* (hybrid degree and distance-based growth model), and compares them with other growth models. In the *DG* model, the probability of building a new connection between two nodes is in inverse proportion to their distance. The *DDG* model, similar to *Law of Gravity*, specifies that the attractiveness between two nodes is determined by their degree and the distance. Based on the simulation results, we discover that both clustering coefficient rates and the average shortest distance are closely related to the edge densification rate, a metric that measures the relative growth speed of edges and nodes. In addition, the hybrid *DDG* model exhibits higher clustering coefficient values and decreasing average shortest distance when the edge densification rate is fixed, which implies that combining locality and attachedness can better characterize the growth of the nanotechnology community. To summarize, the contributions of this paper include: (1) exploring the locality and the attachedness-based network growth paradigm and the corresponding

dynamics and patterns, and (2) discussing factors that can cause high clustering coefficient values and the "shrinking diameters" phenomenon in temporal social networks.

The rest of this paper is organized as follows: Section 2 introduces the background of this study and gives a brief review of the related works; Section 3 defines terminology and concepts used in this paper; Section 4 presents several observations in social networks that motivate our work; Section 5 describes two computational models that incorporate distance and attachedness factors and presents quantitative analysis on the impact of these two factors on the topological properties of graph. Finally, Section 6 states some potential future work and concludes this paper.

2 Related work

A proliferation of work studying the evolution dynamics of complex networks has happened in the last decade. These related studies include static analysis on social network evolution [9], and a variety of models to reproduce the static topological properties and dynamic patterns observed in social networks. The majority of the related studies focus on either the degree factor or the locality factor. This section provides a list of representative studies along this line:

Attachedness-based Growth Schemes: Attachedness is a concept that measures how well nodes are connected and therefore it is usually reflected by the degree of nodes in complex networks. In their well-received article, Barabasi and Albert develop the notable *Preferential Attachment* theory that specifies high degree nodes are always favored when building new connections [3]. In this paper, the authors develop a model in which new nodes are added to the network one by one. The authors claim that the probability that a node v_n will be linked to a vertex v_i depends on v_i's degree, $\frac{d_i}{\sum_j d_j}$, where d_i is the degree of node v_i. Each new node attaches itself (creates a link) to one of the existing nodes with a certain probability that is proportional to the number of links that the existing nodes possess. The authors demonstrate

that this simple scheme results in power-law degree distribution and "rich get richer" phenomenon. In a later paper [4], the authors developed a "continuum theory" based on the preferential attachment theory and use an Mante Carlo approach to simulate the network growth process. The authors show that the clustering coefficient value can decrease or increase by adjusting the parameter that specifies the number of newly created internal links per node in unit time. However, in contrast with the decreasing average shortest distance observed in reality, this approach results in increasing average separation, which was attributed to the incomplete data by the authors. In this study, we demonstrate that the trend of average shortest distance is closely related to edge densification rate. In addition, combining both locality and attachness rate tends to result in decreasing average nodes separation.

Locality-based Growth Scheme: Degree-based models assume that the attractiveness between nodes only depends on their degrees and is independent of the distance between them. In contrast, many of the existing models explicitly or implicitly exploit the locality factor and assume that the generation of a new connection between two arbitrary nodes is related to how far apart they are in the existing topology. The following is a list of these models:

1. *Copying mechanism:* This model specifies that at each time step a new node is added to the network by connecting to a constant number of existing nodes in the network [17]. The new node copies a number of links from a "prototype" node that is selected randomly from the existing nodes whereas choosing the remaining neighbors is random. The author show that this process can result in scale-free distribution.
2. *Walking on a network:* Inspired by citation networks, Vazquez designed the *Walking on a network* scheme to simulate the graph growth process [24]. The model specifies that a network always starts with an isolated node. At every time step, a new node v_i is added and linked to a randomly selected node v_j through a directed edge. The node v_i then mimics a "random walk" on the network by

following the edges starting from node v_j and linking to their end points with probability p. This step is repeated for those nodes to which new connections were established, until no new target node is found.

3. *Referral Model:* Davidsen et al. present a simple scheme that connections are always formed between two nodes that share a common neighbor[7]. This model emulates the real-world scenario that one person introduces two of his acquaintances get to know each other. Such a simple evolution scheme is viewed as the basis of the evolution of social networks. The authors demonstrate that this simple scheme is able to reproduce major nontrivial features of social networks including short path length, high clustering and scale-free or exponential degree distribution.
4. *Distance Preference Model:* Jost and Joy describe a purely distance-based scheme where each new node is connected to a randomly selected node and the subsequent connections are related to the distance of the destination node[11]. This computational model resembles the *DG* model in our work. However, the authors focus on the discussion of degree distribution and assume that new nodes are always connected to the rest of the networks upon joining. Thus at any given time, there is only one GCC in the network. This assumption can lead to very different dynamics of network growth.

Similar graph growth mechanisms also include models that implicitly or explicitly rely on the locality heuristics [8,13,15,18,19,25] or specified feature similarity (correlation) between nodes [26]. In particular, Guimera et al. [8] propose a team assembly mechanism by investigating the interplay between "incumbents" and "newcomers" in the context of collaboration networks. This mechanism focuses on the establishment of "collaborations" rather than "links". The authors attempt to reproduce a variety of networks by adjusting the likelihood of different types of agents participating in the collaborations and evaluate the generated networks based on their degree distribution. While this novel team assembly model can be generalized to more generic networks, we study explicitly how proximity and degree play

a role in network growth processes. Another more recent work on this front includes Morris and Goldstein's team-based growth model, known as the Yule model, for the bipartite networks that consist of both papers and authors [20]. While both the Yule model and our paper are concerned with the growth mechanism of collaboration networks, there exist three main differences between the two papers. First, as opposed to the Yule model that focuses on the bipartite networks, we only focus on collaboration networks, thus our model does not model the productivity (i.e., the number of papers generated) by a team of collaborators. However, our model can be potentially applied to other complex networks, especially those in which the cost of connecting two nodes is related to their distance. Second, the team-based Yule model uses preferential attachment for within-team author selection for a new paper, and random selection of new authors outside the team. Hence, it adopts a binary locality measure (i.e., whether an author is within a team or outside of a team). In contrast, the proposed hybrid model (DDG) in this paper uses a continuous locality measure based on the distance between two authors in the collaboration network. The third difference between the two growth models is that the team-based Yule model does not use preferential attachment for selecting new authors outside of the team, whereas our proposed hybrid model applies preferential attachments to all nodes, regardless of whether they are close or far away on the network.

These papers provide valuable insights into the dynamics of complex networks. However, there are inadequate studies in investigating how proximity and degree quantatively contribute to the network growth scheme and the implications on the static and dynamic properties of social network. In this paper, we provide in-depth analysis on the growth scheme of social networks and develop network growth models that incorporate both global attachedness and locality factors. Hence, both degree and distance factors are taken into consideration in the hybrid model proposed in this paper. The simulation results

indicate that the evolution process of social networks can be better characterized by combining these two factors.

3 Temporal social networks

In order to investigate the growth and evolution process of social networks, we define evolving social networks as *temporal social networks*. In a temporal social network $G(t) = \{V(t), E(t)\}$, the vertex set $V(t)$ and the edge set $E(t)$ evolve over time. The snapshot of a temporal social network at time t_k, $G(t_k) = \{V(t_k), E(t_k)\}$, is a static graph. The two vertices of an edge $e_i = (v_j, v_k)$ is denoted as $V(e_i) = \{v_j, v_k\}$. The set of new connections that are built at time t_k is $\Delta E(t_k) = \{e_1^{t_k}, e_2^{t_k}, ..., e_m^{t_k}\}$. The corresponding vertex set is $\Delta V(t_k)$. And we have

$$E(t_k) = E(t_{k-1}) \cup \Delta E(t_k)$$

$$V(t_k) = V(t_{k-1}) \cup \Delta V(t_k)$$

Note that in the context of collaboration networks, connections can be constructed repeatedly between the same two nodes at different times, which implies that the set $E(t_{k-1})$ and $\Delta E(t_k)$, and $V(t_{k-1})$ and $\Delta V(t_k)$ may not be disjoint. More formally, $E(t_{k-1}) \cap \Delta E(t_k) \supseteq \Phi$ and $V(t_{k-1}) \cap \Delta V(t_k) \supseteq \Phi$, where Φ represents the empty set. We also define the edge density rate of a temporal social network as $\chi(t) = \frac{|E(t)|}{|V(t)|(|V(t)|-1)}$ and the *edge densification rate* as $\Delta\chi(t) = \frac{|\Delta E(t)|}{|\Delta V(t)|}$. The edge density rate is a static concept and describes the density of edges versus nodes at a particular timestamp t. In contrast, the edge densification rate is a dynamic concept and it characterizes the speed of edges growth versus nodes growth. As will be shown in Section 5, the *edge densification rate* is a crucial factor in determining topological properties of temporal social networks.

The proximity of two individual nodes in a social network is often defined in the context of the investigated application domains. The most widely used measure is the shortest distance betweeen the

two nodes. In addition to shortest distance, researchers have also discovered that two additional factors can help: for instance Koren et al. proposed a cycle-free effective conductance (CFEC) to measure distance between network nodes by accouting for the multiple and disparate paths between nodes [12]. However, the pairwise CFEC computation is prohibitive for large-scale social networks; Liben-Nowell et al. have shown that the number of common neighbors is a helpful proximity indicator [19]. In the proximity-based model we develop and describe in the following sections, we adopt shortest distance for individual proximity measure. However, the other proximity measures can fit in the model alternatively.

Accordingly the aggregate proximity properties of a social network can be evaluated by a variety of measures, including average shortest distance, diameter, and effective diameter. The diameter dt of a social network is defined as the largest shortest path between any two nodes, i.e $dt = Max_{v_i,v_j \in V} r(v_i, v_j)$. Some researchers use *effective diameter*, a measure that is obtained by taking $90th$ percentile of the largest shortest distance combined with interpolation, to reduce variance. However, the diameter, effective diameter, and average shortest distance tend to exhibit similar dynamics in our experiments. Thus, for the sake of simplicity, we use average shortest distance for measuring the aggregate proximity of social networks in the rest of this paper.

In temporal social network, the distance between two nodes changes over time. The shortest distance from v_i to v_j at time t_k is denoted as $r_{t_k}(v_i, v_j)$. The average shortest distance for a graph at time t_k is denoted as

$$\overline{r}(t_k) = \frac{\sum_{i,j} r_{t_k}(v_i, v_j)}{|V(t_k)|(|V(t_k)| - 1)}.$$

Similarly, the diameter $dt(t_k)$ of a temporal social network at time t_k is defined as the largest shortest path between any two nodes, i.e $dt_(t_k) = Max_{v_i,v_j \in V(t_k)} r_{t_k}(v_i, v_j)$. Some researchers use *effective diameter*, a measure that is obtained by taking $90th$ percentile of the largest shortest distance combined with interpolation, to reduce variance. However, the diameter, effective diameter, and average shortest distance

tend to exhibit similar dynamics in our experiments. Thus, for the sake of simplicity, we use average shortest distance measure in the rest of this paper.

The clustering coefficient $\aleph_{v_i}(t_k)$ for a node v_i at time t_k is defined as the proportion of links between the vertices within v_i's neighborhood divided by the number of links that could possibly exist between them. More formally,

$$\aleph_{v_i}(t_k) = \frac{|\{e_{jk}\}|}{d_{v_i}(t_k)(d_{v_i}(t_k) - 1)} \tag{1}$$

where $v_j, v_k \in V(t), e_{jk}, e_{ji}, e_{ki} \in E(t)$ and $d_{v_i}(t_k)$ is the degree of node v_i at time t_k. The clustering coefficient of a node v_i measures how well v_i's neighbors are connected to each other. The average clustering coefficient $\overline{\aleph(t)}$ characterizes the modularity of the social network at time t [23].

Evolving graphs $G(t)$ usually consist of a number of isolated subgraphs. It is particularly interesting to investigate the patterns and behaviors of the largest connected cluster, which is referred as the **Giant Connected Cluster (GCC)** in this paper, denoted as $GCC(t) = \{V_{GCC}(t), E_{GCC}(t)\}$ where $V_{GCC}(t) \subseteq V(t)$, $E_{GCC}(t) \subseteq E(t)$ and

$$\forall v_i, v_j (v_i, v_j) \in E_{GCC}(t) \Rightarrow v_i \in V_{GCC}(t) and v_j \in V_{GCC}(t).$$

Table 1 lists a number of important notations for the concepts and terminology used in this paper. Also note we use "node" and "vertex"; "edge" and "connections" synonymously.

4 Observations and motivations for social network growth models

NanoSCI is a collection of nanotechnology-related articles published and indexed by *SCI* (Science Citation Index) in 1980-2006 period. The records are acquired by inquiring at the *Thomson Scientific* website [1] directly. Using keyword-based queries generated based on an iterative relevance feedback technique [14], we obtain 368, 511 SCI-indexed papers regarding nanotechnology. The essential idea of this approach

Table 1 Terminology and notations for temporal social networks

Notation	Meaning
v_i	a vertex
e_i	an edge
$V(t)$	the set of vertices at time t
$E(t)$	the set of edges at time t
$G(t) = (V(t), E(t))$	the graph G at time t
$\Delta V(t)$	the set of new vertices at time t
$\Delta E(t)$	the set of new edges at time t
$\chi(t)$	$\frac{\lvert E(t)\rvert}{\lvert V(t)\rvert(\lvert V(t)\rvert-1)}$ edge density ratio
$\Delta\chi(t)$	$\frac{\lvert\Delta E(t)\rvert}{\lvert\Delta V(t)\rvert}$ the densification rate of edges versus nodes at time t
$V(e_i)$	the two vertices of edge e_i
$\aleph(v_i)$	the clustering coefficient of node v_i
$\overline{\aleph(G(t))}$	the average clustering coefficient of graph $G(t)$
$r_t(v_i, v_j)$	the shortest distance between nodes v_i and v_j at time t
$\overline{r}(G(t))$	the average shortest distance for Graph $G(t)$
$d_t(v_i)$	the degree of node v_i at time t
$C_k(t)$	the expected number of vertices whose degree are k at time t

is to augment the keyword set until the returned results converge. In addition, we extract several sub-communities of nanotechnology from the *NanoSCI* dataset using keywords such as *NanoTube*, *NanoWire*, *NanoParticle*, *Fullerene*, etc. These sub-communities vary with each other in terms of start year, the number of papers and authors. We consider that the *NanoSCI* and each of the sub-communities represent a scientific collaboration network. In each network the nodes are the researchers and two researchers are connected if they have coauthored a paper, which is represented as a link. The number of papers and the

Table 2 Statistics for different nanotechnology communities as of year 2006

Dataset	*Researchers*	*Papers*	$\|E_{GCC}\|$	$\|V_{GCC}\|$
NanoSCI	292,393	368,511	1,836,499	268,594
NanoTube	31,688	25,285	149,138	26,849
NanoWire	86,234	80,645	435,451	77,304
NanoParticle	81,734	69,530	400,749	72,905
Fullerene	97,641	96,331	515,898	88,496

number of researchers for the *NanoSCI* and for each of the four nanotechnology communities as of year 2006 are listed in Table 2.

This paper compares the proposed social network growth scheme with existing models in Section 5.5 using the collaboration networks constructed for these communities.

Next we assess how the number of links (edges) between researchers and the number of researchers (nodes) increase with time. Figure 1 shows in log-log scale the edge growth versus node growth for the *NanoTube* and *NanoSCI* communities, respectively. It appears that the growth speed is almost linear in the log-log scale which implies that the edge growth increases as power law as a function of the nodes growth. This finding justifies using the densification laws suggested previously [18]. These Regression results show that their growth rates are $|E(t)| = 2.5173 * |V(t)|^{1.1049}$ and $|E(t)| = 3.0459 * |V(t)|^{1.1141}$ respectively. Thus, the corresponding edge densification rates for the two communities are $\Delta\chi(t) = 2.78 * |V(t)|^{0.10}$ and $3.39 * |V(t)|^{0.11}$ respectively. These edge densification rate of *NanoSCI* communities is used in comparing the simulation results of different network growth models, which will be discussed in Section 5.5.

Figure 2 demonstrates the temporal changes of average shortest distance, $\overline{r}(GCC(t))$, of the giant connected component, $GCC(t)$ for *NanoSCI*, *NanoTube*, *Fullerene*, *NanoParticle*, and *NanoWire*, re-

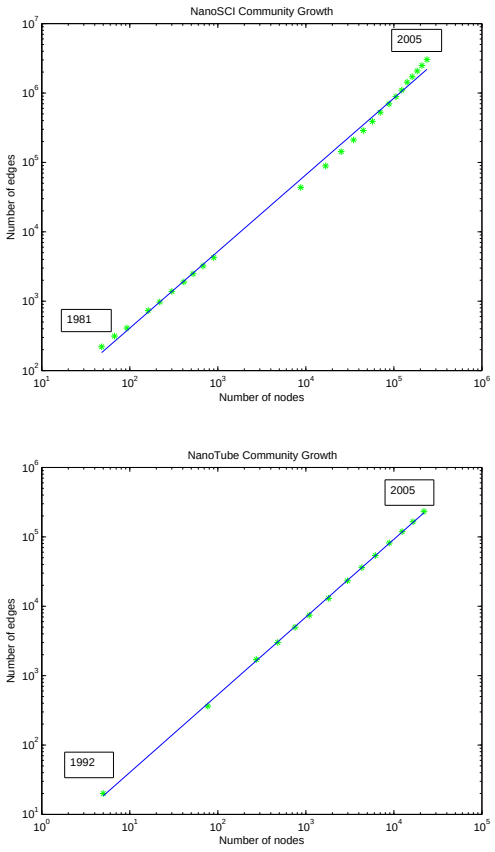

Fig. 1 The number of edges $E(t)$ versus number of nodes $V(t)$ for *NanoSCI* and *NanoTube*.

.

spectively. Note that *Fullenrene* and *NanoWire* are recently emerging communities and the number of researchers as of 2006 is less than 20000 according to the collected data. In this pilot study we focus on analysis of the $GCC(t)$ and will leave exploring the entire sysytem for future research. These results clearly illustrate the "shrinking diameter" phenomenon that has been reported in [16,18]. This is contradictory to conventional wisdom that would predict that the diameter of growing networks shall increase over time. Leskovec et al [18] developed a "forest fire" network growth model, in which the diameters can either increase or decrease over time by adjusting parameters of the model. In the following sections

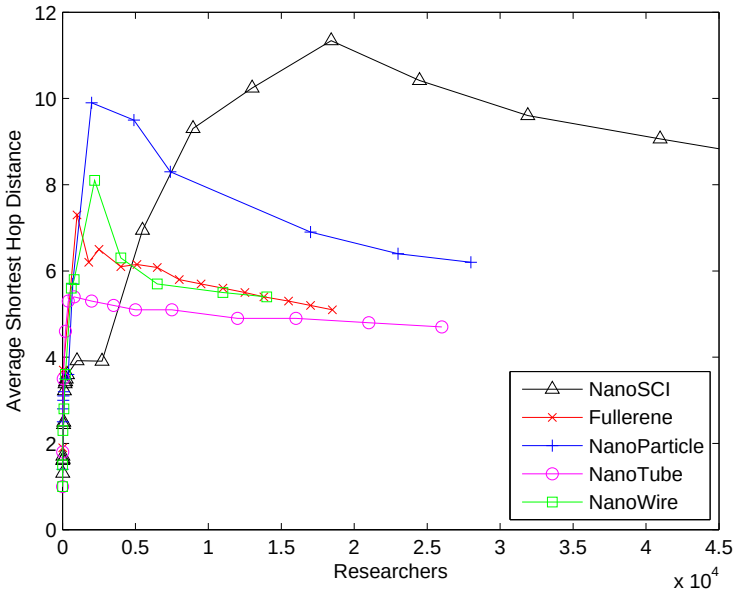

Fig. 2 The average shortest distance $\overline{r}(GCC(t))$ of $GCC(t)$ for *NanSCI*, *Fullerene*, *NanoParticle*, *NanoTube*, and *NanoWire*. Note that *Fullenrene* and *NanoWire* are recently emerging communities and the number of researchers as of 2006 is less than 20000 according to the collected data.

Table 3 Average clustering coefficient values for nanotecnology communities from year $1980 - 2006$

Dataset	NanoSCI	NanoTube	Fullerene	NanoParticle	NanoWire
$\overline{\aleph}(G(t))$	0.82	0.81	0.82	0.83	0.87

we show that this phenomenon can be attributed to both edge densification rates and the way that new connections are formed (growth models) in the evolving social networks.

The average clustering coefficient, $\overline{\aleph}(G(t))$, is an indicator for the modularity of networks. Table 3 shows the average clustering coefficient for *NanoSCI*, *NanoTube*, *Fullerene*, *NanoParticle*, and *NanoWire* communities. Clustering coefficient value of 0.81 for *NanoTube*, 0.82 for *Fullerene*, 0.83 for *NanoParticle*, and 0.87 for *NanoWire* communities are significantly higher than random networks, which is usually below

0.1 [2], and also higher than the clustering coefficient values for other collaboration networks reported [21]. In studying the structure of scientific collaboration networks [21] finds that the clustering coefficient varies between 0.066 for Medline (papers in biomedical research) to 0.43 for papers published in the Los Alamos Archive, to 0.73 for SPIRES (published papers and preprints in high-energy physics). Values of $\overline{\aleph}(G(t))$ obtained in this study being higher than the clustering coefficient for high-energy physics indicate even higher modularity of the nano-science communities. In Section 5, we will compare the average clustering coefficient of the simulation results of the several network growth models with these observations from the nanotechnology communities.

In order to explore the causes of such intriguing phenomena, including decreasing shortest distance and high clustering coefficient values, we propose a set of computational models that employ relatively simple growth schemes and explore how these growth mechanisms can affect the topological properties of the underlying temporal social networks.

5 Combining locality and global preferential attachment for modeling network growth

The attachedness factor has been traditionally considered as a principal factor in attracting new connections. In addition to this factor, we observe that nodes tend to connect to their peers within topologically proximity. For instance, Figure 3a shows the distribution of the shortest distance, as of time t_k, between nodes that establish the third category of connections at time $t_k + 1$ in the $GCC(t_k)$ of the pertaining social network versus the distribution of pair-wise shortest distance between the nodes in $GCC(t)$. Figure 3a shows the distribution of $r_{2005}(v_i, v_j)$ where $e_l = (v_i, v_j) \in \Delta E(2006)$ and $v_i, v_j \in GCC(2005)$ for the *NanoTube* community. We neglect those connections which had already been in the network in previous years. This figure demonstrates that there is remarkable disparity between the two distributions.

It indicates that new connections tend to be created between nodes in proximity. In particular, the vast majority of links are added between the nodes that are only two hops apart.

To demonstrate explicitly that nodes form new links inversely proportionally to the topological distance, we calculate the proportion $Fr(r) = M/N$, where N denotes the number of node pairs at distance r and M are the pairs among them that form new edges in the next time step. The results is shown in Figure 3b for *NanSCI* and *NanoTube* communities.

In a network a randomly selected node is connected to d other nodes through d links (edges) with probability $P(d)$ which is called vertex connectivity or degree distribution. We obtain the probability $P(d)$ for each of the nanoscience communities. The results for *NanoSCI* and *NanoTube* calculated in equidistant in logarithmic scale bins are plotted in Fig. 4a. Triangles mark the degree distribution of all nodes that exist in the *NanoSCI* network from its inception through the end of 2005. Crosses mark the degree distribution of all nodes that exist in the *NanoTube* network from its inception through the end of 2005. The tails of both of these distributions exhibit a behavior that is close to a power law. Networks that show such power law distribution are know as scale-free networks [3].

Barabasi and Albert [3,4] have suggested that power law distribution may apply to most of the networks of interest including social networks. They report that scientific collaboration networks in mathematics and neuroscience scale with power law exponent of 2.4 and 2.1, respectively. We find similar values of the power law exponent for nanoscience networks (see Table 4). Newman [21] report on the structure of scientific collaboration networks and find that collaboration networks in condensed matter physics, astrophysics, high-energy physics and computer science, all can be best fit with a power law form with exponential cut-off. Similarly to [21] we find that the degree distribution of the networks in nanosciences are best fit with a power law form with exponential cut-off

Table 4 Summary of results of the analysis. The p-values of the fit for all coefficients are less than 0.0001

Dataset	τ	d_c	R^2
NanoSCI			
Total in 2005	2.21	250	0.99
New in 2006	1.77	166.7	0.98
NanoTube			
Total in 2005	1.94	108.7	1.0
New in 2006	1.41	87.7	0.99

$$P(d) \sim d^{-\tau} e^{-d/d_c}, \tag{2}$$

where τ and d_c are constants whose values are listed in Table 4.

To demonstrate explicitly that nodes form new links proportionally to the degree of the nodes we calculate the proportion $Fd(d)$, as a ratio between the number of nodes that form new edges at a certain step and the number of nodes with the same degree that existed at the previous time step. The results for *NanoSCI* and *NanoTube* communities are shown in Fig. 4b.

Based on these observations, we propose a novel hybrid network growing scheme that incorporates a locality element into global preferential attachment. We compare this scheme with three existing models: (1) random growth model, (2) preferential attachment model, and (3) distance based growth model. In order to make a fair comparison, we parameterize these models and make the *edge densification rate* identical to real nanotechnology community growth rates. In Section 5.5, we discuss the simulation results of these four network growth models and compare them with the observations from nanotechnology communities. In the following sections, sections 5.1 and 5.2 discuss the random growth model and the

preferential attachment model; sections 5.3 and 5.4 describe the distance-based model and the proposed hybrid model. Finally, section 5.5 presents analysis and insights based on the simulation results.

5.1 A Random Growth Model

Callaway et al proposed a simple random growth model where one new node and at most one new edge are added at a time [6]. We slightly modify the model by parameterizing the number of the new edges as a function of the number of existing nodes. The number of existing nodes equals to the time-stamp t therefore the number of new edges is denoted as $\Delta E(t)$. This model is referred as *A-Random* model in this paper.

The model involves a cycle of three steps: (1) At each time step, add one new node to the graph. (2) Randomly select two nodes and create an undirected edge between them. (3) Repeat (2) for $\Delta E(t)$ times.

At time t, there will be t vertices and on average $e(t) = \sum_t \Delta e(t)$ edges, where $e(t)$ is defined as $|E(t)|$. Let $c_k(t)$ be the expected number of vertices with degree k at time t. The number of isolated vertices $c_0(t)$ will increase by 1 at each time step, but decrease on average by $2\Delta e(t)\frac{c_0(t)}{t}$, the probability that a degree zero vertex is randomly chosen as one of the ends of a new edge. Thus:

$$c_0(t+1) = c_0(t) + 1 - 2\Delta e(t+1)\frac{c_0(t)}{t} \tag{3}$$

Similarly, the expected number of degree k vertices ($k > 0$) will increase on average by an amount proportional to the probability that a degree $k-1$ vertex is chosen for attachment by a new edge, and decrease by an amount proportional to the probability that a degree k vertex is chosen. This gives:

$$c_k(t+1) = c_k(t) + 2\Delta e(t+1)\frac{c_{k-1}(t+1)}{t} - 2\Delta e(t+1)\frac{c_k(t)}{t+1} \tag{4}$$

Note that the above equations neglect the possibility that an edge links a vertex to itself. This means the equations are only approximate at short times, but they become exact in the limit $t- > \infty$ because the probability that any vertex is chosen twice decreases like t^{-2}.

5.2 Parameterized preferential attachment model (PPAM)

This section describes a simple parameterized preferential attachment model (PPAM), in which a new vertex and l new edges ($l = \Delta e(t)$) are added into the network at each time step. In building a new connection, we specify that (1) we randomly select a start node v_i, and (2) the probability that a node v_j is selected as the end node of the new edge is

$$p_t(v_j) = \frac{d_t^{'}(v_j)}{\sum_k d_t^{'}(v_k)}$$

where

$$d_t^{'}(v_i) = d_t(v_i) + 1 \tag{5}$$

Thus,

$$\sum_{k=1}^{t} d_t^{'}(v_k) = \sum_{k=1}^{t} d_t(v_k) + t = 2e(t) + t \tag{6}$$

Therefore, the likelihood of a node n_s connecting to another node n_e only depends on their degree. Note that by using $d_t^{'}(v_i)$, the model allows zero-degree nodes to be selected as the end node.

$$c_0(t+1) = c_0(t) + 1 - \Delta e(t+1)(\frac{c_0(t)}{t} + \frac{c_0(t)}{2e(t)+t}) \tag{7}$$

The probability $SA_{k-1}(t+1)$ that a degree $k-1$ vertex is selected for attachment by a new edge at time $t+1$ is:

$$SA_{k-1}(t+1) = c_{k-1}(t)(\frac{1}{t} + \frac{k-1}{2e(t)+t}) \tag{8}$$

Similarly, the probability $SA_k(t+1)$ that a degree k vertex is chosen for attachment by a new edge at time $t+1$ is:

$$SA_k(t+1) = c_k(t)(\frac{1}{t} + \frac{k}{2e(t)+t}) \tag{9}$$

Hence, the degree distribution is determined by

$$c_k(t+1) = c_k(t) + \Delta e(t+1)(SA_{k-1}(t+1) - SA_k(t+1)). \tag{10}$$

5.2.1 Degree-Product based PPAM (DP-PPAM) Barabasi et al. extended the preferential attachment model to take into account the degree product of both nodes in the network. Thus we also compare our distance based models to the following updated degree-product based PPAM model. In particular, in building a new connection, we specify that (1) the start node v_i is selcted based on its degree; the probability is defined as $\frac{d_i^{'}}{\sum_k d_k^{'}}$; and (2) the probability that a node v_j is selected as the end node of the new edge is

$$p_t(v_j) = \frac{d_t^{'}(v_j)}{\sum_k d_t^{'}(v_k)}$$

where

$$d_t^{'}(v_i) = d_t(v_i) + 1 \tag{11}$$

5.3 Distance-based Growth (DG) model

This section describes a simple proximity-based growth model in which the likelihood of building a connection between two nodes only depends on their proximity. Note that the proximity between two individual nodes can be evalulated by a variety of measures, including shortest distance, CFEC [12]. In this paper we use shortest distance to measure the proximity between two nodes. In the growth process, a new vertex and l edges ($l = \Delta E(t)$) are added to the graph at each time step. The two vertices of a new edge are determined in the following way: (1) one node n_s is selected uniformly from the graph as

the start vertex of the new edge; (2) the probability that a node v_p is selected as the end vertex of the new edge is:

$$p_t(v_p) = \frac{\frac{1}{r_t^{'}(v_p,v_s)}}{\sum_p \frac{1}{r_t^{'}(v_p,v_s)}} \tag{12}$$

where

$$r_t^{'}(v_p, v_s) = \begin{cases} R & \text{if } r_t(v_p, v_s) = \infty \\ r_t(v_p, v_s) & \text{otherwise.} \end{cases} \tag{13}$$

Equation 13 specifies that if two nodes are disconnected, the distance between them is a large number R. This way, the probability of building a new connection between any two nodes is non-zero.

5.4 A hybrid distance and degree growth (DDG) model

This section describes a hybrid *DDG* model in which the distance and attachedness (i.e. degree) factors are combined to determine the likelihood that a node is selected to form a new edge. Similar to the aforementioned models, one new vertex and $l = \Delta E(t)$ edges are added at each time step in the graph growth process. The *DDG* model specifies that the start node v_s is selected randomly from the graph, but the probability that a node v_p is selected as the end node of the new edge is

$$p_t(v_p) = \frac{\frac{d_t^{'}(v_p)}{r_t^{'}(v_p,v_s)}}{\sum_p \frac{d_t^{'}(v_p)}{r_t^{'}(v_p,v_s)}} \tag{14}$$

where $r_t(p(v_p, v_s))$ is defined by equation 13 and $d_t(v_p)$ is defined by equation 11.

5.5 Analysis and simulations

In order to evaluate the proposed network growth model, we compare the simulation results of the hybrid model together with those of the other three models with observations regarding *NanoSCI*, *NanoParticle*,

and *NanoTube* communities. As a comparison, we also calculate the topological properties of pure random graphs with the same number of nodes and edges as a baseline to compare with other approaches. This baseline approach is denoted as *PureRandom* in this section. As a reminder, *A-Random* refers to the quasi-random approach described in Section 5.1; *PPAM* refers to the previously described preferential attachment model. In the rest of this section, we analyze and compare these network growth methods from two important perspectives: (1) the average shortest distance of the networks generated by these models over time, and (2) the average clustering coefficient of these networks.

5.5.1 Temporal distance analysis In this section we compare the average shortest distance between the different models described previously in this Section. Figures 5, 6, 7, and 8 illustrate the results of a set of experiments obtained by varying the densification rates and growth dynamics. In particular, Figure 5 shows the average shortest distance versus time using the same node growth rates ($\Delta|V(t)|$) and the same *edge densification rate* ($\Delta\chi(t) = 3.39 * |V(t)|^{0.12}$). Figure 7 shows the temporal patterns of average shortest distance for *A-Random* approaches at different edge densification rates ($\Delta\chi(t) = \{0.02 * |V(t)|^{0.69}; 0.02 * |V(t)|^{0.6}; 3.39 * |V(t)|^{0.12}; 0.02 * |V(t)|^{0.5}\}$ respectively. Figure 7 indicates that the kinetic properties of average shortest distance metric are closely related to the ratio of edge growth versus node growth. When the *edge densification rate* is $\Delta\chi(t) = 3.39 * |V(t)|^{0.12}$, the average shortest distances of *A-Random* increase slowly over time. However, when the *edge densification rate* is $\Delta\chi(t) = 0.02 * |V(t)|^{0.69}$, the average distance for *A-Random* decreases over time after a sharp increase in early stage of network growth. Similar results are observed for other growth models as well. In general when the edge density is higher, the average shortest distance for these models at a particular time point is smaller. Figure 5, however, shows an interesting result from the perspective of modeling the growth dynamics of nanotechnology community. By adopting the edge densification rates of *NanoSCI* community

(i.e, $\Delta\chi(t) = 3.39 * |V(t)|^{0.12}$), the simulation results of these models resulted in rather different growth behavior.

The average shortest distance of the network generated by the hybrid *DDG* model decreases over time after an early-stage increase, which is similar to what we observed in the actual nanotechnology community (see Figure 2). In contrast, the average shortest distance of the networks generated by the random growth models and the preferential attachment model (PPAM) increases over time. Results for a modification of the preferential attachment model which assumes that two nodes connect with each other proportionally to the product of their degrees are plotted in Figure 6. The average shortest distance is first increasing and then keeps almost constant value similar to the the average shortest distance of the local distance-based model (*DG*) which increases first, then oscillates around a convergence point. The average shortest distance produced by the DDG model better describes the observations than the preferential attachment degree product model. Figure 8 demonstrates that the average shortest distance first increases and then decreases with increasing the size of the **GCC**. We also noted that the effect of adding a local factor into the global attachment-based model reduces the average shortest distance, while the purely locality-based model results in a larger average shortest distance than does the global attachment-based model. This growth behavior suggests that there is a synergistic effect in proximity and attachedness factors

5.5.2 Clustering coefficient analysis A node's clustering coefficient measures the connectivity among this node's neighbors. Large cluster coefficient values indicate that the neighbors of the node in question are well connected to each other. Table 5 shows the clustering coefficient values for networks generated by different models using the edge densification rate of the *NanoSCI* community (i.e. $\Delta\chi(t) = 3.39 * |V(t)|^{0.12}$). This table indicates that the hybrid *DDG* model clearly has significantly higher clustering coefficient values than the preferential attachment, the A-Random growth, and the local distance-based

Table 5 Average clustering coefficient values for proposed computational models. The clustering coefficient values are the average of clustering coefficient values for snapshots obtained over 5000 time cycles.

Dataset	A-Random	*DG*	*PPAM*	*DDG*
$\overline{\aleph}(G(t))$	0.02	0.05	0.03	0.45

model. More importantly, the simulation results suggest that the hybrid *DDG* model generates networks with a clustering coefficient that is much closer to those of the collaborative network of the nanotechnology communities, shown in Table 3.

In conclusion, based on the simulation results, we observe that the hybrid *DDG* model is able to produce networks with clustering coefficient values closer to what was observed in the nanotechnology community. Furthermore, it generates networks whose average shortest distance decreases over time when the edge densification rate of the *NanoSCI* community is used. Hence it is more suitable as a model for the collaborative network of the nanotechnology community than are either the global preferential attachment model or the local distance-based model.

6 Conclusions and Future Work

The explosive development of nanotechnology research calls for in-depth scientometric study of this field. In this paper we conducted a case study on the evolving nanotechnology collaboration networks and concluded that both locality and attachedness play significant roles in the social network growth process. In particular, this paper expands the definition of temporal social networks and demonstrates that a science-based collaboration network is similar to other real-world social networks. Furthermore, the nanotechnology collaboration networks studied exhibit an intriguing set of static and dynamic properties. Inspired by the observations that in collaboration networks new connections tend to be augmented between nodes in proximity, we explored both locality and attachedness factors in growing networks

and proposed two distance-based computational growth schemes, namely *DG* and *DDG*. The *DG* model considers only the locality element while the *DDG* is a hybrid model that factors into both locality and attachedness elements. We discovered that the dynamic patterns of average shortest distance and clustering coefficient are closely connected to the edge densification rates as well as specific growth mechanisms. In addition, we discovered that when we use when the edge densification rate of the *NanoSCI* community, (1) clustering coefficient rates of the *DDG* model were closer to those of the nanotechnology community, and (2) the *DDG* model exhibited decreasing average shortest distance phenomenon, which was observed in the collaborative network of the nanotechnology community as well. These simulation results suggest that the hybrid approach that combines locality and attachedness can better characterized the growth of the nanotechnology community.

The result of this study also inspires us to investigate related questions in our future work. For instance, what are the general characteristics of social networks that are best characterized by a hybrid network growth model? How can variations in hybrid growth models be compared to each other? What insights can be obtained from different hybrid approaches to model network growth? Future research that address these and other related questions can not only improve our understanding about the dynamic behavior of network growth, but also lay the foundation for providing deeper insights on social network analysis.

References

1. http://scientific.thomson.com/products/sci. 2006.
2. R. Albert and A.-L. Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 2002.
3. A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509, 1999.

4. A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *PHYSICA A*, 311:3, 2002.

5. B. Bollobs. *Random Graphs, Second Edition*. Cambridge University Press, 2001.

6. D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz. Are randomly grown graphs really random? *Physical Review E*, 64:041902, 2001.

7. J. Davidsen, H. Ebel, and S. Bornholdt. Emergence of a small world from local interactions: Modeling acquaintance networks. *Physical Review Letters*, 88:128701, 2002.

8. E. M. Jin, M. Girvan, and M. E. J. Newman. The structure of growing social networks. *Physical Review E*, 64:046132, 2001.

9. R. N. Kostoff, J. A. Stump1, D. Johnson1, J. S. Murday, C. G. Lau, and W. M. Tolles. The structure and infrastructure of the global nanotechnology literature, 2006.

10. L. Krapivsky and S. Redner. Organization of growing random networks. *Physical Review E*, 63:066123, 2001.

11. R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 611–617, New York, NY, USA, 2006. ACM Press.

12. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. The web as a graph. In *PODS*, pages 1–10, 2000.

13. J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *KDD '05: Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187, New York, NY, USA, 2005. ACM Press.

14. D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. *Journal of American Society for Information Science and Technology*, 58(7):1019–1031, 2007.

15. S. A. Morris and M. L. Goldstein. Manifestation of research teams in journal literature: a growth model of papers, authors, collaboration, coauthorship, weak ties and lotka's law. *Journal of American Society for Information Science and Technology*, 58(12):1764–1782, 2007.

16. M. E. Newman. The structure of scientific collaboration networks. *Proc Natl Acad Sci U S A*, 98:404–409, Jan 2001.
17. M. E. Newman. Coauthorship networks and patterns of scientific collaboration. *Proc Natl Acad Sci U S A*, 101 Suppl 1:5200–5205, April 2004.
18. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–1555, August 2002.
19. A. Vazquez. Knowing a network by walking on it: emergence of scaling, Sep 2000.

References

1. http://scientific.thomson.com/products/sci. 2006.
2. R. Albert and A.-L. Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 2002.
3. A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509, 1999.
4. A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A*, 311:3, 2002.
5. B. Bollobs. *Random Graphs, Second Edition*. Cambridge University Press, 2001.
6. D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz. Are randomly grown graphs really random? *Physical Review E*, 64:041902, 2001.
7. J. Davidsen, H. Ebel, and S. Bornholdt. Emergence of a small world from local interactions: Modeling acquaintance networks. *Physical Review Letters*, 88:128701, 2002.
8. R. Guimera, B. Uzzi, J. Spiro, and L. A. N. Amaral. Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. *Science*, 308(5722):697–702, 2005.
9. D. B. Horn, T. A. Finholt, J. P. Birnholtz, D. Motwani, and S. Jayaraman. Six degrees of jonathan grudin: a social network analysis of the evolution and impact of cscw research. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 582–591, New York, NY, USA, 2004. ACM.

10. E. M. Jin, M. Girvan, and M. E. J. Newman. The structure of growing social networks. *Physical Review E*, 64:046132, 2001.

11. J.Jost and M. P. Joy. Evolving networks with distance preferences. *Physical Review E*, 66, 2002.

12. Y. Koren, S. C. North, and C. Volinsky. Measuring and extracting proximity in networks. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 245–255, New York, NY, USA, 2006. ACM.

13. G. Kossinets and D. J. Watts. Empirical analysis of an evolving social network. *science*, 311:88–90, 2006.

14. R. N. Kostoff, J. A. Stump1, D. Johnson1, J. S. Murday, C. G. Lau, and W. M. Tolles. The structure and infrastructure of the global nanotechnology literature, 2006.

15. L. Krapivsky and S. Redner. Organization of growing random networks. *Physical Review E*, 63:066123, 2001.

16. R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In *KDD '06: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 611–617, New York, NY, USA, 2006. ACM Press.

17. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. The web as a graph. In *PODS*, pages 1–10, 2000.

18. J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *KDD '05: Proceeding of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, pages 177–187, New York, NY, USA, 2005. ACM Press.

19. D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. *Journal of American Society for Information Science and Technology*, 58(7):1019–1031, 2007.

20. S. A. Morris and M. L. Goldstein. Manifestation of research teams in journal literature: a growth model of papers, authors, collaboration, coauthorship, weak ties and lotka's law. *Journal of American Society for Information Science and Technology*, 58(12):1764–1782, 2007.

21. M. E. Newman. The structure of scientific collaboration networks. *Proceedings of National Academy Sciences U S A*, 98:404–409, Jan 2001.

22. M. E. Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of National Academy Sciences U S A*, 101 Suppl 1:5200–5205, April 2004.

23. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–1555, August 2002.

24. A. Vazquez. Knowing a network by walking on it: emergence of scaling, Sep 2000.

25. D. J. Watts, P. S. Dodds, and M. E. J. Newman. Identity and search in social networks. *Science*, 296:1302–1305, 2002.

26. Q. Xuan, Y. Li, and T.-J. Wu. A local-world network model based on inter-node correlation degree. *Physica A Statistical Mechanics and its Applications*, 378:561–572, May 2007.

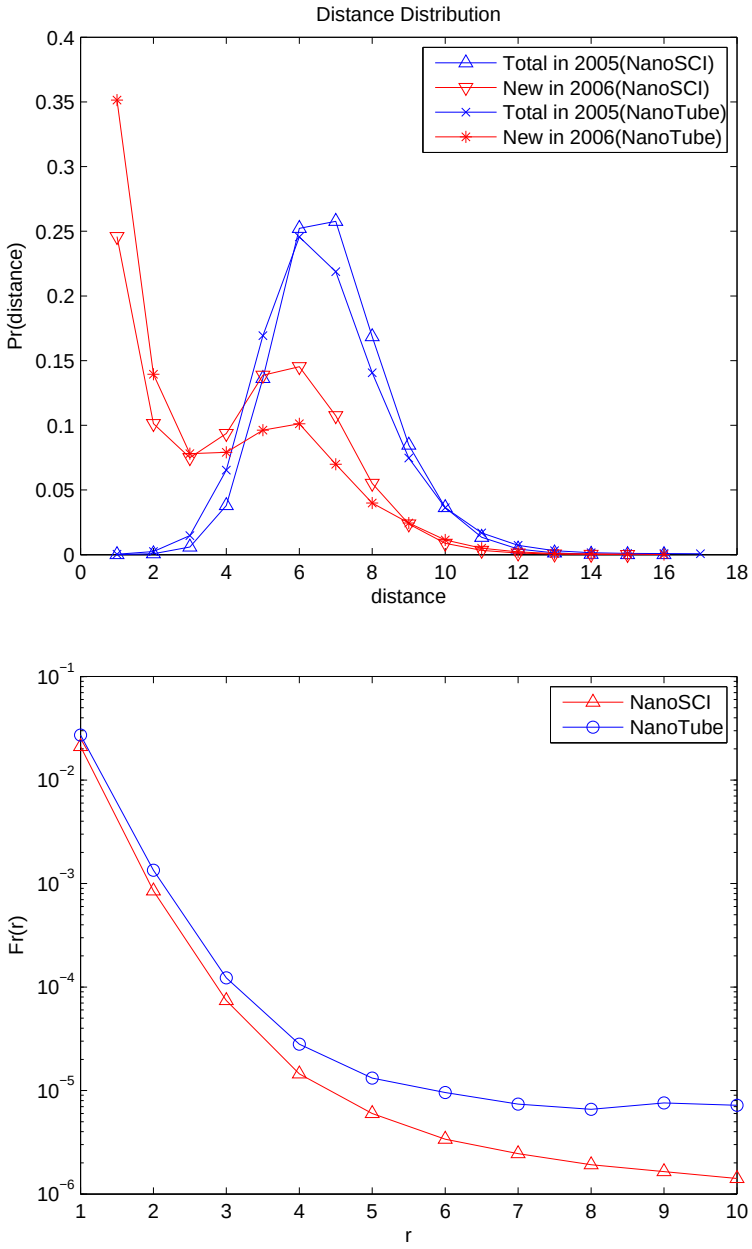

Fig. 3 (a) The locality preference of selecting neighbors in the NanoTube community. The left curve shows the distribution of shortest distance values between those nodes that are in GCC(2005) and build direct collaboration relation between each other in year 2006 (i.e. $r_{2005}(v_i, v_j)$ where $(v_i, v_j) \in \Delta E(2006) and v_i, v_j \in GCC(2005)$). The right curve shows the distribution of the pair-wise shortest distribution for all nodes in $GCC(2005)$. (b) Proportion of node pairs $Fr(r)$ with certain distance forming new edges in a new time step.

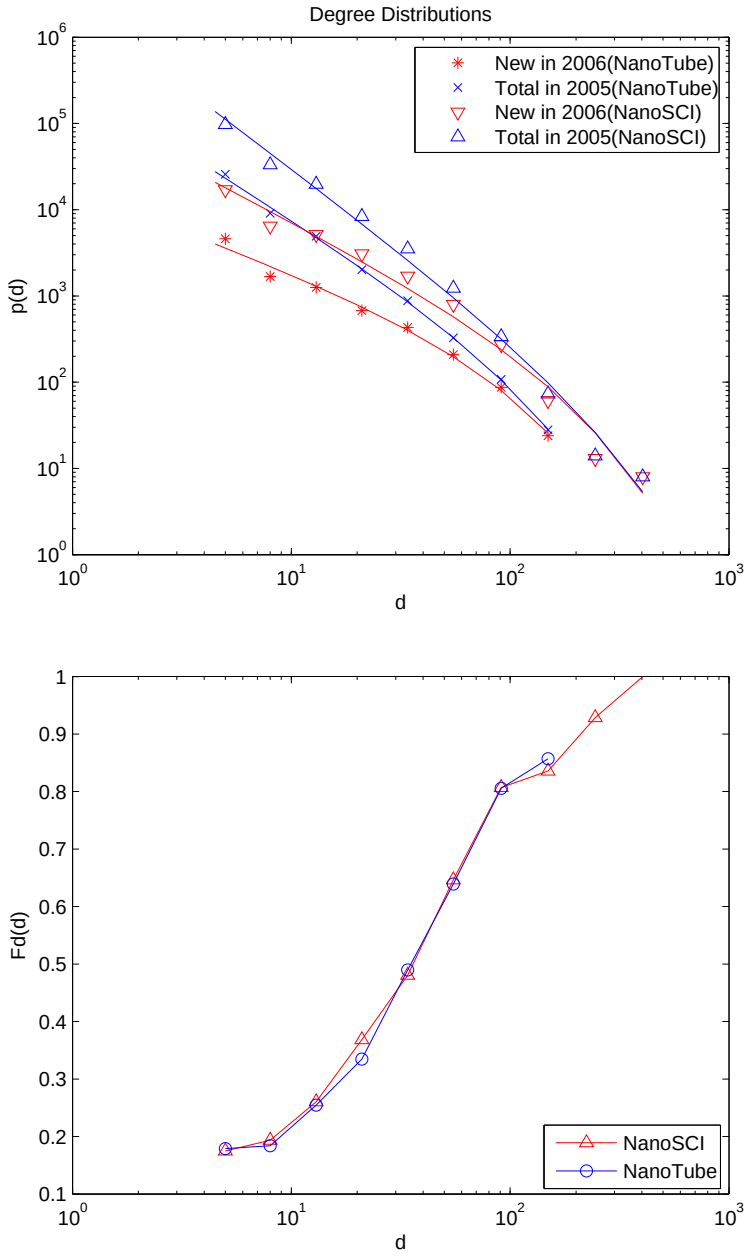

Fig. 4 (a) Degree distributions of the *NanoSCI* and *NanoTube* networks for all nodes through the end of 2005 (triangles and crosses, respectively) and for nodes that occur in 2006 only (inverted triangles and asterick, respectively). Each set of symbols is fitted with a power law with exponential cut-off. See the text for discussion and consult Table 4 for the values of the parameters of the fit. (b) Proportion of node $Fd(d)$ with certain degree attracting new edges in a new time step.

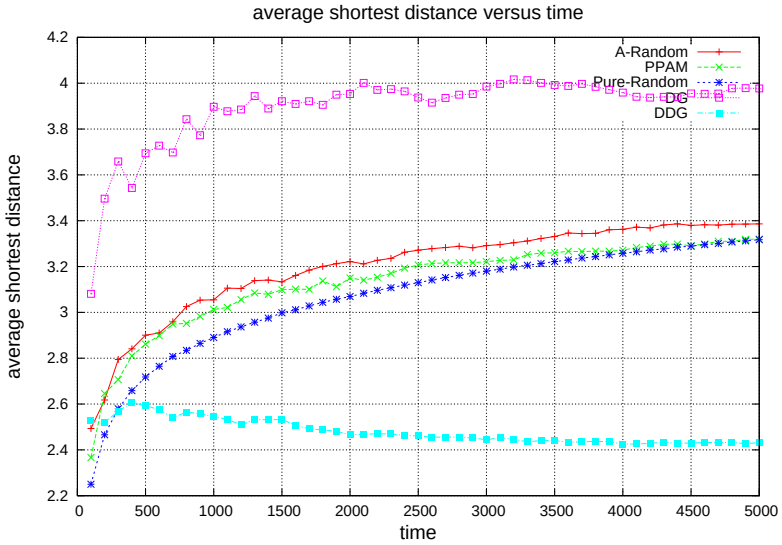

Fig. 5 The average shortest distance versus time. The *edge densification rate* is fixed at $\Delta\chi(t) = 3.39 * |V(t)|^{0.12}$

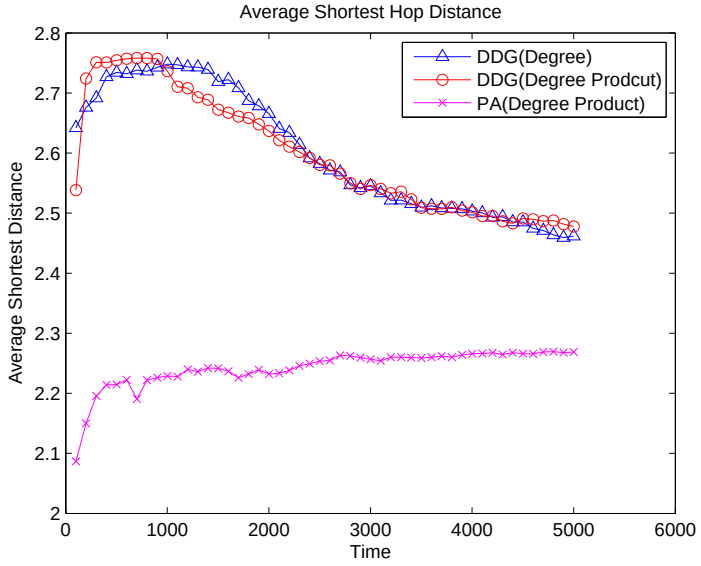

Fig. 6 The average shortest distance versus time simulation results for three models; triangles mark the result for *DDG* model, the circles denote the result for *DDG* model with degree-product and the asterisks mark the degree-based based preferential attachment model

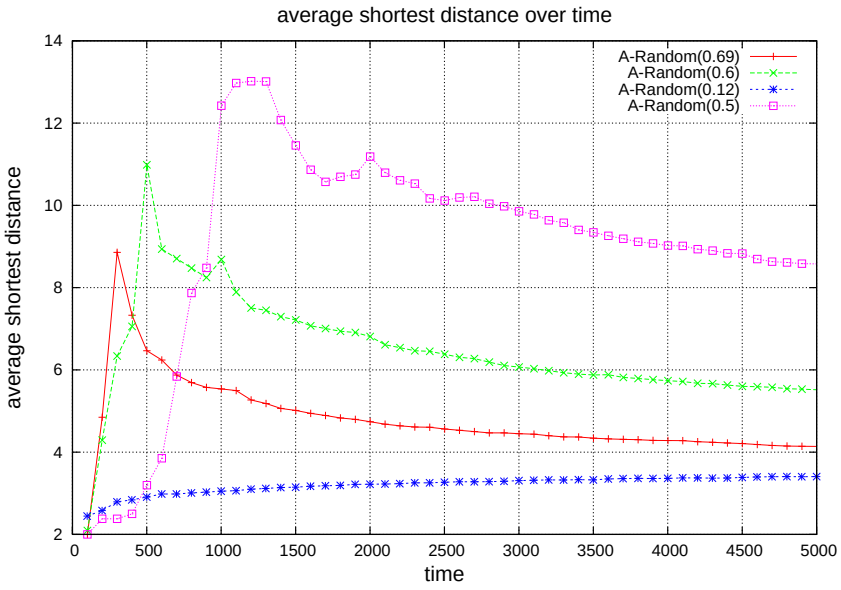

Fig. 7 The average shortest distance versus time ($\Delta\chi(t) = \{0.02 * |V(t)|^{0.69}; 0.02 * |V(t)|^{0.6}; 3.39 * |V(t)|^{0.12}; 0.02 * |V(t)|^{0.5}\}$ respectively)

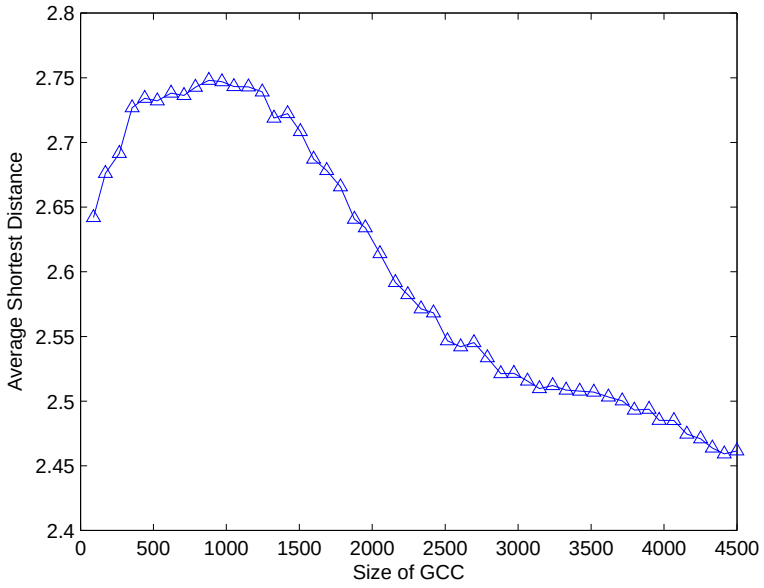

Fig. 8 The average shortest distance versus GCC size for DDG-degree algorithm