Complete trails of coauthorship network evolution

Deokjae Lee,[1] K.-I. Goh,[2,*] B. Kahng,[1,3,†] and D. Kim[1]
[1]*Department of Physics and Astronomy, Seoul National University, Seoul 151-747, Korea*
[2]*Department of Physics, Korea University, Seoul 136-713, Korea*
[3]*School of Physics, Korea Institute for Advanced Study, Seoul 130-722, Korea*
(Received 17 February 2010; published 24 August 2010)

The rise and fall of a research field is the cumulative outcome of its intrinsic scientific value and social coordination among scientists. The structure of the social component is quantifiable by the social network of researchers linked via coauthorship relations, which can be tracked through digital records. Here, we use such coauthorship data in theoretical physics and study their complete evolutionary trail since inception, with a particular emphasis on the early transient stages. We find that the coauthorship networks evolve through three common major processes in time: the nucleation of small isolated components, the formation of a treelike giant component through cluster aggregation, and the entanglement of the network by large-scale loops. The giant component is constantly changing yet robust upon link degradations, forming the network's dynamic core. The observed patterns are successfully reproducible through a network model.

DOI: 10.1103/PhysRevE.82.026112 PACS number(s): 89.75.Hc, 89.75.Fb, 89.65.−s

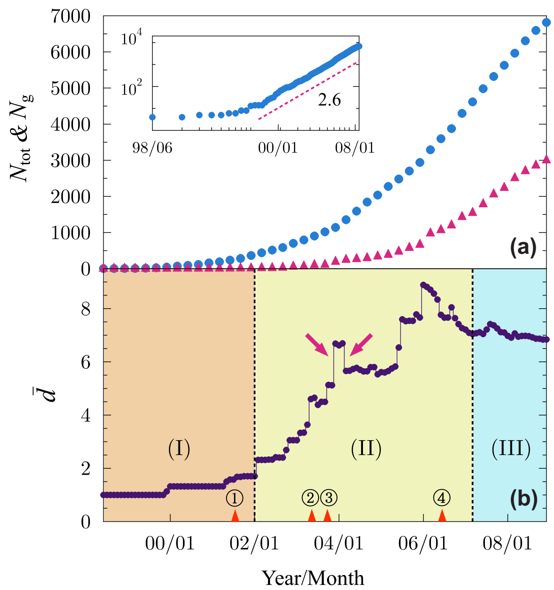
I. INTRODUCTION

At the brink of the 21st century, two papers heralding the beginning of a new science of network, which were for the small-world and the scale-free networks [1,2], were published. Since then, complex network research (CNR) has flourished, not only as an active research field but also as a common structural analytical framework by which the systems approach to various complex systems in the natural, social, and information sciences can potentially be unified [3–7]. Along with the emergence of a new research field, a new research community forms and evolves. Thus, the CNR provides a useful example to study the spreading of a research field in terms of the evolution of the social network behind it [8–11]. Specifically, we can quantify evolutionary patterns in its initial transient periods of the network formation, which has never been explicitly observed. We can also portray the route through which the coauthorship network reaches a fully grown state.

To achieve this goal, we apply the complex network theory to study how the CNR has developed since its inception: we first construct the coauthorship network in which nodes are researchers participating in the CNR and a link is made when two authors write a paper together [9]. The weight of the link is given by the number of papers coauthored. To track the evolution of the network, the two aforementioned papers [1,2] and three early review papers [12–14] have been chosen. They are the highest cited papers and regarded as pioneering papers in the CNR field. Next, we considered all the subsequent published papers citing any of these five papers to engage in the CNR. According to Web of Science, there are 5008 such papers with information on the list of authors and publication times measured in months, written by 6816 nonredundant authors (in terms of their last name and initials) for a period spanning 127 months from June 1998 to December 2008. The coauthorship network was constructed each month from the papers published up to that month [15]. This network is a growing weighted network. With the data, we have performed a detailed temporal analysis of the evolution of the large-scale structure of the network and discuss its social implications. We particularly emphasize the early stage of evolution, which has not been addressed in previous studies [16,17].

Our fine-scale temporal analysis in Secs. II–IV reveals a global structural transition of the network through three major regimes of (I) the nucleation of small isolated components, (II) the formation of a treelike giant component by cluster aggregation, and (III) the network entanglement by long-range loop formation. The network reaches the steady state in which the mean separation between two nodes stabilizes around a finite value. The locality constraint, that is, new links are formed much more locally than globally, played an important role in sustaining the network's treelike structure in regime (II). Here, by treelike structure, we mean that the network is dominated by short-range loops and devoid of long-range connections, thus becoming a tree when coarse grained into the network of supernodes, corresponding in this case to groups led by each principal investigator. This implies that most papers are made through in-group collaborations, even though researchers began sharing ideas through international conferences. If the locality effect were weak, the intermediate stage (II) would not appear. Moreover, such a treelike structure is a fractal and sustains even underneath the entangled network in the late regime (II). This structure is unveiled upon the removal of inactivated edges and has the same fractal dimension as in treelike structure. This implies that a hidden ordered structure with the same fractal dimension underlies in the evolution process. In Sec. V, a model is constructed based on the empirical findings and suggests that a structural transition in the real coauthorship network can be understood as a percolation transition in the growth parameter space. Finally, we will summarize the results and discuss their robustness and implications in Sec. VI.

*kgoh@korea.ac.kr
†bkahng@snu.ac.kr



(c)

	Title	Place	Date
①	International conference on dynamical networks in complex systems	Kiel, Germany	July, 2001
②	Networks: Structure, Dynamics and Function	Santa Fe, USA	May, 2003
③	COSIN midterm conference on growing networks and graphs in statistical physics, finance, biology and social systems	Rome, Italy	September, 2003
④	NETSCI 2006	Bloomington, USA	June, 2006

FIG. 1. (Color online) Large-scale evolution of the CNR coauthorship network. (a) The total number of nodes N_{tot} (researchers; ∘) and the largest component size N_g (Δ) of the network as a function of time (quarterly). N_{tot} grows in a power-law fashion asymptotically with exponent 2.6 as shown in the inset. (b) The mean separation $\bar{d}$ between a pair of nodes in the giant component as a function of time. In regime (I), the giant component is small in size, and so is $\bar{d}$. It grows in regime (II) and exhibits intermittent jumps (indicated by the left arrow, for example) by merging with smaller but macroscopic components (illustrated in Fig. 2). It also drops suddenly from time to time (indicated by the right arrow, for example), by making a long-range link (illustrated in Fig. 2). In regime (III), the network gets entangled further and $\bar{d}$ remains almost constant, even though the size of the giant component grows steadily. (c) The list of the international conferences indicated in (b).

II. LARGE-SCALE NETWORK EVOLUTION AND STRUCTURAL PROPERTIES

Since its inception, CNR has grown steadily over the decade, and the pace of growth has not yet started to decelerate [Fig. 1(a)]. The largest connected component (giant component) of the coauthorship network has also grown in size and has reached almost a half of the total network [Fig. 1(a)]. The mean separation $\bar{d}$ between two nodes in the giant component becomes relatively stable to reach around 7 after passing the intermediate regime, during which it displays strong temporal fluctuations [Fig. 1(b)]. This stable behavior of $\bar{d} \approx 6-7$ is robust in other coauthorship networks [18–21].

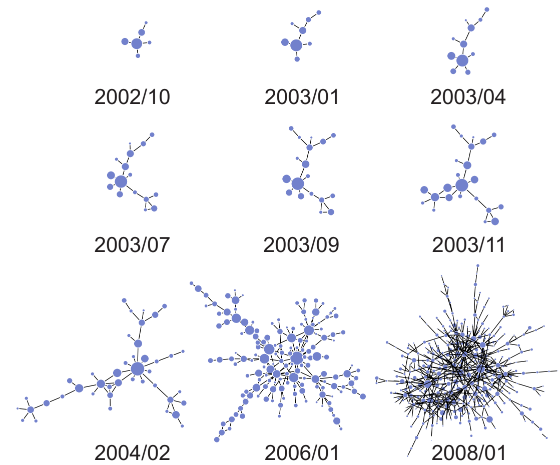


FIG. 2. (Color online) Structural evolution of coarse-grained network. The coarse-grained network obtained by the affinity propagation algorithm [22] is shown for various times. A supernode in the coarse-grained network represents a group of researchers identified by the algorithm, and its size corresponds to the number of researchers in the group. We note that until 2006, the coarse-grained network is effectively a tree, devoid of long-range loops in it.

The network has grown both by the expansion and merging of existing components and by the continuous introduction of new components. In the earliest stage [regime (I) in Fig. 1(b)], small-sized components nucleate independently and their number and size increase with time. Most of the currently highly connected nodes (researchers) have already appeared in this regime, playing the role of pioneers and contributing to the progress of the field. On the brink of the intermediate stage [regime (II) in Fig. 1(b)], the giant component is formed, which might be promoted by the first international conference exclusively devoted to CNR [(① in Figs. 1(b) and 1(c)]. In regime (II), the giant component grows in a treelike manner; it branches out more and deeper with the passage of time, but rarely establishes links between branches, as can be seen by coarse graining the network (Fig. 2) obtained by the affinity propagation algorithm [22]. This can also be seen quantitatively in the distribution of shortcut lengths, which is dominated by the peak at $d=2$ [Fig. 3(c)]. At this stage, the giant component is a fractal [23,24] [Figs. 3(a) and 3(b)], with the fractal dimension $d_f \approx 1.7$ measured by the recently introduced box-covering algorithm [26] and the mean branching ratio around unity [27]. With the passage of time, such dynamics continue and the giant component and mean separation $\bar{d}$ gradually increase. Component sizes become inhomogeneous in the growth process (Fig. 4). Such a steady growth may be promoted by large-scale international conferences such as the one held in Santa Fe with the purpose of bringing together scientists from diverse disciplines to discuss network science problems across fields [② in Figs. 1(b) and 1(c)]. However, there are a few intermittent jumps [e.g., the left arrow in Fig. 1(b)], resulting from the merging of smaller but macroscopic components with the largest one [Fig. 5(a)]. A large-scale loop does not appear until 2004, and it is formed by the long-range interbranch link [Fig. 5(b)]. Such a long-range loop formation can be

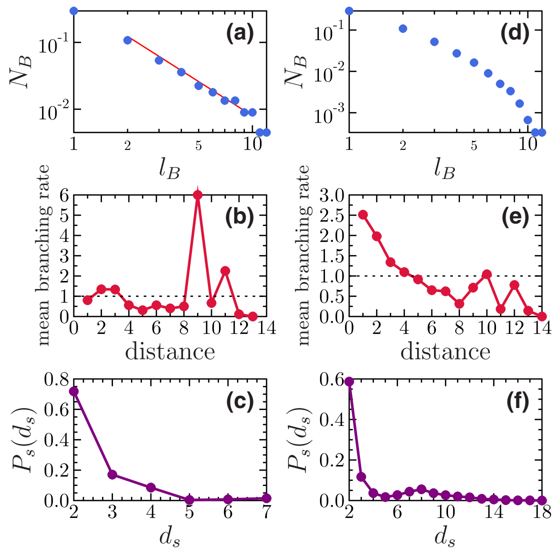


FIG. 3. (Color online) Fractal analysis of the giant component. (a) Box-covering analysis of the giant component in February 2004. $N_B(\ell_B)$, the number of boxes needed to cover an object with box size ℓ_B, decays algebraically in ℓ_B as $N_B(\ell_B) \sim \ell_B^{-d_B}$ with $d_B \approx 1.7$, indicating that the network is a fractal. The guideline has a slope of −1.7, drawn for the eye. (b) Mean branching rate as a function of distance from a root node. To measure this, one identifies the skeleton [25] of the network and its largest hub as the root node. Then we calculate the average branching number in the skeleton as a function of distance from the root. The plateau around the unity value indicates that its branching pattern is that of a critical branching tree, which is a fractal. (c) The distribution of shortcut lengths. By decomposing the network into the skeleton and the residual links, we define the distance along the skeleton between two nodes connected by each residual link as the shortcut length. The shortcut length distribution of the February 2004 network has a peak at $d_s = 2$ and decrease monotonically as d_s increases. In (d)–(f), we measure same quantities for the network in December 2008, finding that (d) the box number decreases exponentially with the box size, (e) the mean branching ratio does not exhibit a plateau but decreases steadily as a function of the distance from the root, and (f) the shortcut length distribution becomes bimodal with an additional peak at $d_s = 8$ due to the long-range links. These results indicate that the network is no longer a fractal in December 2008.

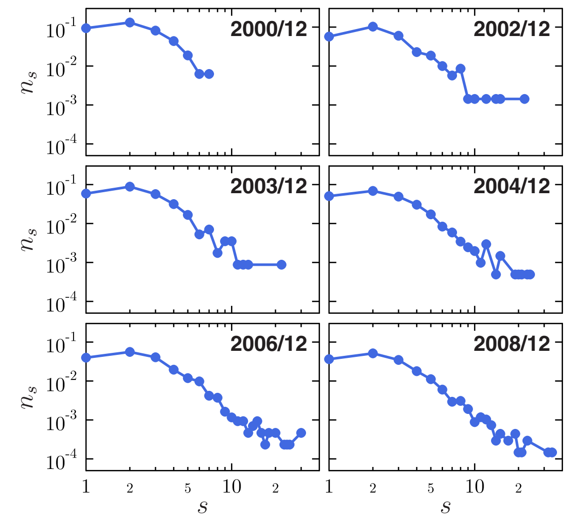


FIG. 4. (Color online) Component size distribution. Distribution of the component sizes excluding the giant component is shown in time. We can see that the tail of the distribution becomes heavier as time goes on, meaning that the heterogeneity in component sizes increases.

monitored by the sudden drop in the mean separation of the giant component [i.e., the right arrow in Fig. 1(b)]. This change can also be monitored by examining the size of largest biconnected component. Furthermore, before this long-range loop formation the ratio between the largest biconnected component size to the largest singly connected component size tends to decrease, implying the treelike growth of the largest component during the period. This event is a consequence of the first major multinational project devoted to CNR in Europe (COSIN) [③ in Figs. 1(b) and 1(c)]. Another prominent example is the peak in January 2006 [Figs. 5(c) and 5(d)]. Since then, an increasing number of large-scale loops have been formed, resulting in an increasingly entangled and interwoven giant component structure and the network has made the transition into regime (III), in which the network properties such as the mean separation become stable, despite the steady growth of the giant component. This transition into a stable research field may be epitomized by the establishment of a *regular* international conference gathering researchers from various multidisciplinary fields [International Workshop and Conference on Network Science (NetSci); ④ in Figs. 1(b) and 1(c)].

Temporal evolution of motif contents reveals the global structural changes from a different viewpoint. We observed that the motifs with one triangle begin to be significant approximately from the beginning of the intermediate regime (II), whereas the motifs containing two triangles do so around June 2005, at which the mean separation exhibits a drastic jump. In this way, the temporal evolution of motif contents is related to that of the mean separation and complements the global evolution picture.

The coauthorship network exhibits heavy-tailed behaviors in the degree (number of links a node is connected to) and strength (the sum of the weights of links a node has) distributions, which become robust over time [Figs. 6(a) and 6(b)]. The degree-degree correlation within the giant component is almost neutral or weakly assortative, in contrast with the assortative behavior observed for the full network and other social networks [Figs. 6(c) and 6(d)] [31,32].

III. EFFECT OF LINK DEGRADATIONS

Social ties decay in strength over time in the absence of reinforcement. Coauthorship links may be no longer active if the collaboration ceased long ago. Thus, to ensure that the generic features remain robust, it is informative to examine how the overall network structure is affected in the presence of a link degradation process. The central question would be whether the giant component persists to support the integration of the research field. To this end, for each month, we removed all the links that had not been reactivated during the previous two years—a typical postdoctoral contract period.

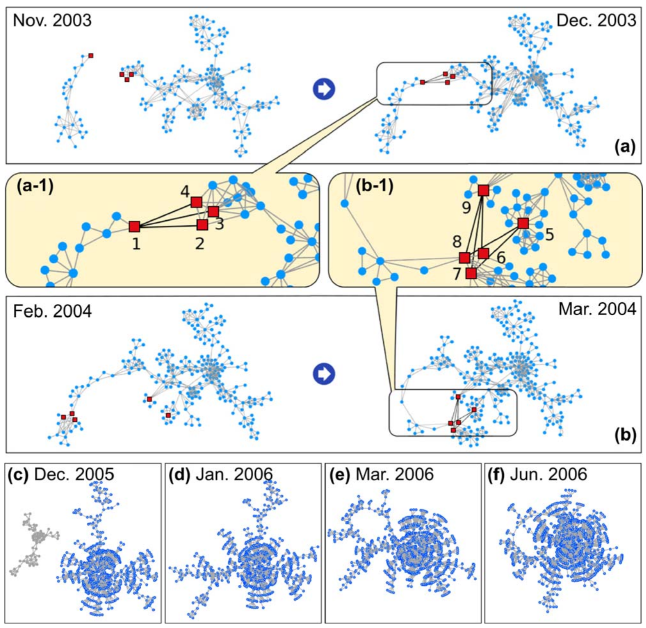


FIG. 5. (Color online) Large-scale changes in the giant component structure. (a) Between November and December 2003, the largest component grows abruptly by merging with a smaller but macroscopic component, leading to a big jump in the mean separation [indicated by left arrow in Fig. 1(b)]. Also shown is the zoomed-in version of the merging process (a-1). This event corresponds to the publication of the paper [28] with four authors (labeled as authors 1–4 with square nodes) from the two components. (b) Later between February and March 2004, the largest component acquires interbranch link to form a long-range loop, giving rise to a sudden drop of the mean separation [right arrow in Fig. 1(b)]. The zoomed-in version of the looping process is also shown (b-1). This event is driven by the publication of two papers [29,30] with three (nodes 5, 7, and 8) and four authors (nodes 6–9), respectively, all of whom were already in the giant component and members of COSIN project. The links involved in the processes mentioned above are in thick lines. (c)–(f) Network changes around the peak in $\bar{d}$ in January 2006. (c) A macroscopically large component has grown in December 2005, and (d) it merges to the giant component in January 2006, yielding a sudden increase in $\bar{d}$. (e) Then a large-scale loop connecting two long branches forms in March 2006, and (f) many such interbranch links are made in June 2006, yielding a sudden decrease in $\bar{d}$.

The link degradation process significantly affects network configurations because many links become inactive in the end [Fig. 7(a)]; for example, 86% of the links formed up to the year 2006 eventually disappeared before the end of 2008 according to the two-year inactivation rule. However, the giant component not only persists upon degradation, but is also more stable, in the sense that its relative size S has been stable at ≈10% of the total network since 2000 [Fig. 7(c)]. At the same time, the link-degraded giant component (LDGC) is highly dynamic, in that its members constantly change over time. At the end of 2008, 1195 nodes formed the LDGC, among which only 272 were the LDGC members in December 2006 composed of 727. This indicates that the CNR is still a vigorous field [17]. The LDGC exhibits a treelike structure throughout the observation period, implying that such a treelike spanning component structure exists to provide a dynamic backbone underlying the complex original interwoven network. Furthermore, the LDGC [Fig. 7(a)] topologically resembles the original giant component in regime (II) [Fig. 5(a)], and their fractal dimensions are the same as $d_f \approx 1.7(1)$ [Figs. 3(d) and 7(b)].

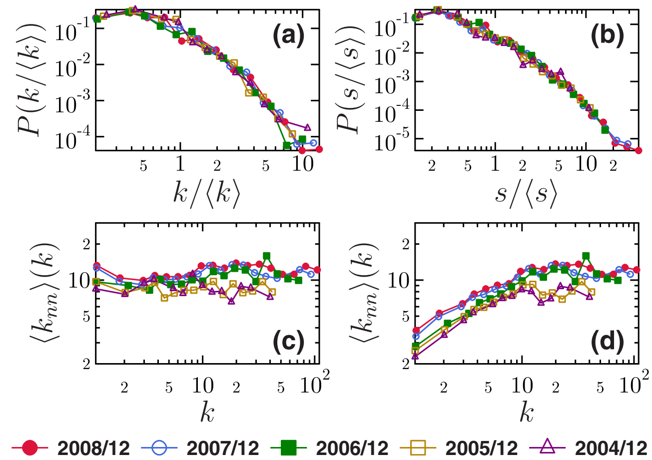


FIG. 6. (Color online) Basic network characteristics. (a) The degree distribution and (b) the strength distribution of the giant component. They are heavy tailed and stationary over time when rescaled by the average value. (c) and (d) The average nearest-neighbor degree function $\langle k_{nn}\rangle$ (c) of the giant component and (d) of the full network as a function of degree k. In contrast to the full network exhibiting the assortative mixing behavior, a typical pattern of social networks, the giant component is almost neutrally mixed over time.

Link degradation properties indicate the future prospects of the research field. Based on the evolutionary trajectory observed, the CNR has passed its initial transient growth period and has now settled into a steady growth regime with stationary topological properties such as the degree distribution $P_k(k)$ and the mean separation $\bar{d}$, even taking the link degradation process into consideration. Moreover, the link degradation process in the coauthorship network occurs in a manner that is consistent with the so-called asymmetric disassembly [33], where the probability of link degradation decreases with the degree of the connecting node [Fig. 7(d)]. Given that asymmetric disassembly provides structural robustness in a declining network [33], the current steady growth of the CNR coauthorship network backed up by the asymmetric disassembly implicates the integration and stability of research discipline in the future, even after it eventually enters into the network saturation stage.

IV. MICROSCOPIC LINK DYNAMICS

To understand the microscopic mechanisms responsible for the large-scale evolution pattern observed, we focus on the link dynamics. We categorize each new link into five classes depending on the nature of the nodes that it connects and measure their relative frequencies in the link dynamics [Fig. 8(a)]. They are (i) the duplicate link, connecting two nodes already linked; (ii) the intracomponent link, connect-

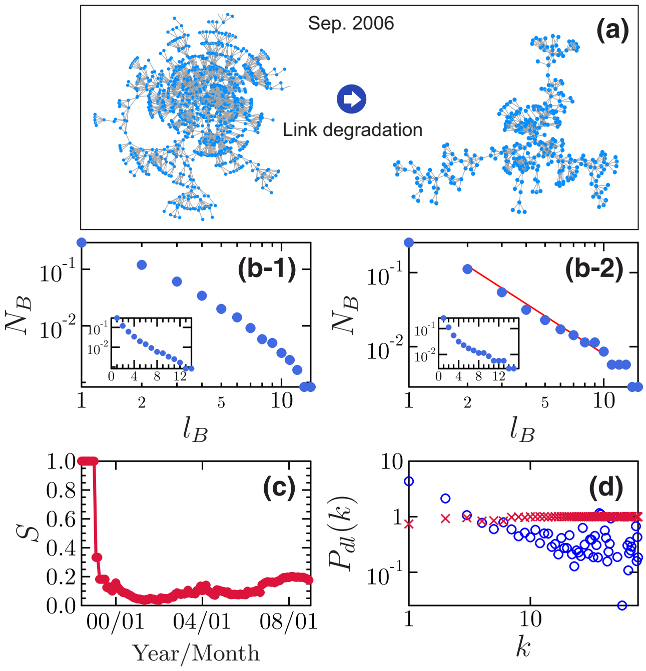


FIG. 7. (Color online) Link degradations. (a) Snapshots of the giant component without (left) and with (right) link degradations in September 2006. The network with link degradation (right) lacks long-range loops and thus is more treelike, while that without the link degradation (left) is much more entangled by long-range loops. (b) Box-covering analysis of September 2006 network both in double-logarithmic (main) and in semilogarithmic (insets) scales. (b-1) Without link degradation N_B decreases exponentially with ℓ_B; thus, it is not a fractal. (b-2) On the other hand, N_B decays algebraically in ℓ_B as $N_B(\ell_B) \sim \ell_B^{-d_B}$ with $d_B \approx 1.7$ for the LDGC, implying its fractal structure. The guideline in (b-2) has a slope of -1.7, drawn for the eye. (c) The relative size S of the LDGC of the network over time. S becomes stable around 0.1–0.2 after 2001. (d) The relative probability of link deletion $P_{dl}(k)$ in the CNR coauthorship network ($\circ$) with respect to random deletions ($\times$), as a function of the node degree k. The decreasing behavior in k indicates the link removal occurs via asymmetric disassembly [33].

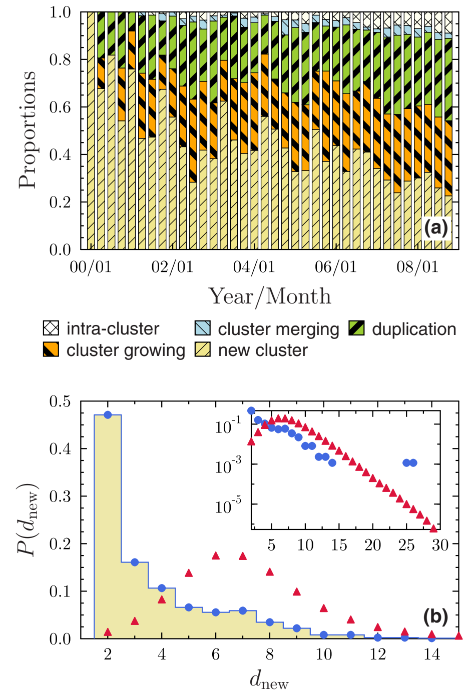


FIG. 8. (Color online) Microscopic link dynamics. (a) The proportions of new links color coded by the five link categories in time (see text for details). (b) The distribution $P(d_{\text{new}})$ of the separation of existing nodes connected by the new links (blue circle supported by light orange bar), compared with that for linking a random pair of nodes (red triangle). The strong enrichment in low d_{new} is indicative of the importance of locality constraint in the link dynamics. Inset: the same plot in the semilogarithmic scale.

ing two unlinked nodes in the same component; (iii) the cluster-growing link, connecting an existing node in a component to a new node, thereby resulting in an incremental growth of the component; (iv) the cluster-merging link, connecting two nodes in different components; and (v) the new-cluster link, connecting two new nodes to introduce a new component.

Among them, the duplicate link, the cluster-growing link, and the new-cluster link are found to be of high frequency, each constituting approximately a quarter to a third of all the links. The remaining two classes, the cluster-merging link and the intracluster link, are far less frequent, comprising 2.8% and 4.7% of all links, respectively. Although infrequent, the latter two classes are the driving forces of major large-scale structural changes: the former provides the punctuated growth of components by merging existing macroscopic components, while the latter can introduce long-range loops that entangle the connectivity structure.

We found that the existing nodes play an equally important role as that of the new nodes: among nodes connected by new links, approximately half of them are existing nodes (51%). We also found a clear signature of the locality effect (i.e., the tendency of nodes in proximity to link with each other), which manifests itself as a strong enrichment of the links connecting nodes at shorter separations, compared to random linkages without such a locality constraint. For example, about 47% of the links between existing nodes are found to be separated by two links before linking, that is, they connect "friends of a friend," compared to 1.5% for random linkages [Fig. 8(b)]. It is this locality-constrained link formation that is responsible for the treelike growth dominating the early structure of the network.

V. NETWORK EVOLUTION MODEL

We model the coauthorship network evolution by incorporating the observed microscopic link dynamics. The network model is built upon a number of previous network models [34–38], with relevant growth ingredients such as the preferential attachment-based internal link formations [35], the triad formation due to locality constraint [36], and the team-based evolution [37]. Although all these ingredients are found relevant in the coauthorship network evolution, none of them alone can account for the whole process. Therefore, we combined ingredients from these previous models and incorporate them into the combined model with additional parameters for the relative frequencies of these processes. In

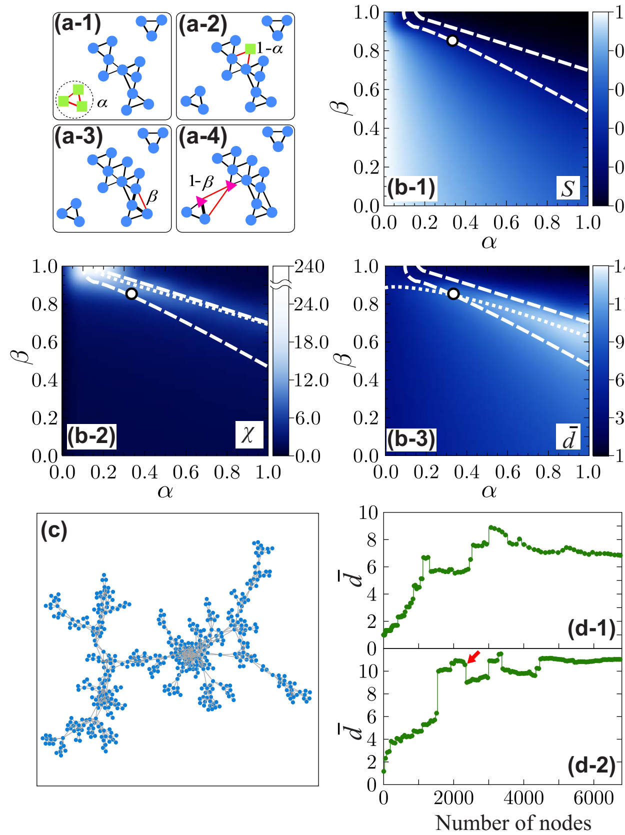


FIG. 9. (Color online) The network evolution model. (a) Schematic illustration of the evolution rules of the network model (see text for details). Square nodes represent the new nodes and the triangles are the selected nodes in the rule. (b) Density plot in the (α,β) parameter space of the relative size of (b-1) the giant component S, (b-2) the mean finite-component size χ, and (b-3) the mean separation of the largest component $\bar{d}$ for the network model. Numerical simulations are performed at the 0.025 intervals for both parameters and are averaged over 100 independent runs. The approximate phase boundaries with $S=0.1$ and $S=0.3$ are drawn in dashed lines for guidance and the empirical parameter set for the complex network coauthorship network is indicated by a white dot. The line along the peak in (b-2) χ and (b-3) $\bar{d}$, respectively, is drawn with the dotted line. (c) Snapshot of the largest component of the model network obtained with the empirical parameter set ($\alpha=0.33,\beta=0.88$), at the point indicated by the arrow in (d-2) with $N_{tot}=972$ and $N_g=336$, exhibiting a treelike topology. (d) Mean separation $\bar{d}$ between a pair of nodes in the giant component (d-1) of the empirical network and (d-2) of a model network with the empirical parameter set as a function of N_{tot}.

this combined network evolution model which is schematically depicted in Fig. 9(a), the model network evolves by a node dynamics and a link dynamics, according to the following rules applied at each time step: (i) With the probability α, a new component with three connected nodes (a triangle) is introduced, representing a new group. The average number of authors per paper is observed to be approximately three [Fig. 10(a)]. (ii) With the complementary probability $1-\alpha$, a new node is added, and then it selects an existing component in proportion to the component size and connects to a node in the component chosen in proportion to the degree (preferential attachment [2]), as well as to a randomly chosen neighbor of that node, forming a triangle. (iii) Independently, with the probability β, a randomly chosen node links to a random neighbor of its neighbor (a friends' friend). (iv) With the complementary probability $1-\beta$, a triangle is formed by a random pair of nodes with a separation larger than 2 and one of their nearest neighbors. Here, the parameters α and β control the influx of new components and the strength of the locality effect, respectively.

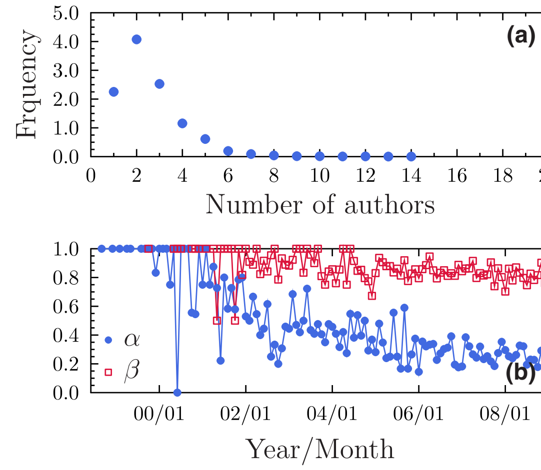


FIG. 10. (Color online) Empirical model parameters. (a) Distribution of the number of authors per paper calculated from all 5008 papers is plotted. The average value is obtained to be 2.9. Note that we excluded the papers with more than 20 authors. (b) The network model parameters α and β are measured from the CNR publication data. The parameter α changes slowly in time, with the global average being ≈ 0.33, while the parameter β is relatively stable as ≈ 0.85.

Model simulation results

We run the network model up to the size $N=6800$ for various growth parameters α and β and calculate the general characteristics of the model network, specifically the fraction of nodes in the giant component S and the average size of the finite (nongiant) components χ used in the percolation study. Here, $S\equiv 1-\Sigma_s s n_s$ and $\chi\equiv\Sigma_s s^2 n_s/\Sigma_s s n_s$, where n_s is the fraction of s-sized component and the summation runs over finite components. In addition, we measure the mean separation $\bar{d}$ between two nodes in the giant component, which may be analogous to the correlation length used in the percolation theory. S decreases monotonically in both α and β [Fig. 9(b-1)], quite abruptly in the region bounded with dashed lines. χ [Fig. 9(b-2)] exhibits a peak behavior along the same region, denoted by dotted line. Both behaviors suggest a percolation transition-type event occurring across the region in the parameter space. The mean separation $\bar{d}$ [Fig. 9(b-3)] also displays a peak behavior along the same region in a large-α regime, denoted by dotted line, establishing that the giant component is treelike in the percolation transition region.

Having understood the generic behavior of the network model, where does the real coauthorship network reside in the parameter space? We measure the parameter set from the

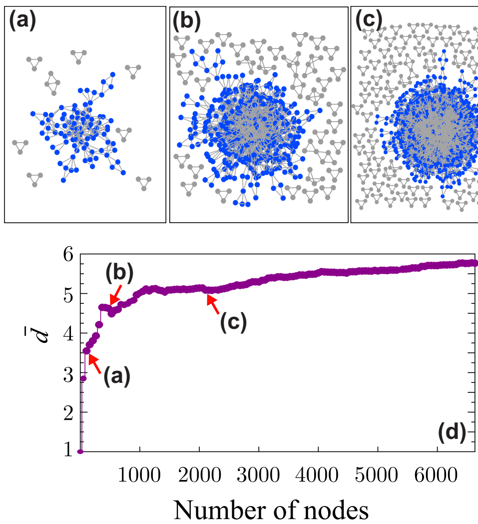


FIG. 11. (Color online) Evolution of the model network without locality constraint. (a)–(c) The network snapshots and (d) the mean separation $\bar{d}$ over time obtained from a typical run of the network model with parameters $\alpha=0.33$ and $\beta=0.2$. Very weak locality constraint is imposed to the model network with this low value of β, and it is unable to maintain the initial treelike growth stage but gets interwoven immediately. Therefore, one cannot observe the macroscopic-scale cluster aggregation process except the incremental growth of the giant component.

empirical data, finding that while β is steady at $\beta_{em}\approx 0.85$, α_{em} depends on time [Fig. 10(b)]. We estimate it roughly by taking the average over time to be $\alpha_{em}\approx 0.33$. Interestingly, the measured parameter set [indicated by a white dot in Fig. 9(b)] is located within the percolation transition region. This implies that the network achieves a balance between the continuous influx of new isolated components and the formation of global connectivity. The treelike giant component may be rooted from the fact that most research groups tend to work independently and rarely collaborate with other group members; however, the hub group members perform out-of-group collaborations more actively by locating themselves near the center of the network.

The configurations generated from the network model with the empirically measured parameter set successfully reproduce the observed general large-scale structural features, such as the treelike growth of the giant component [Fig. 9(c)] as well as a strong fluctuation in the mean separation $\bar{d}$ in the early time regime, followed by its stabilization through the network entanglement by long-range loops in the later time regime [Fig. 9(d)]. Outside the empirical parameter point, the model network evolves in different ways. Notably, when the locality effect is weak (small β), the network is unable to maintain the initial treelike growth stage and quickly forms a hairball-like interwoven structure, without the macroscopic-scale cluster aggregation process, as in most of the unconstrained random growth models (Fig. 11). Furthermore, generalization of the model with a more complex component structure such as a mixture of dimer and trimer in the growth rule does not affect the main results. Thus, the current simple network model appears to address the essential mechanisms underlying the evolution of real networks in a minimal way.

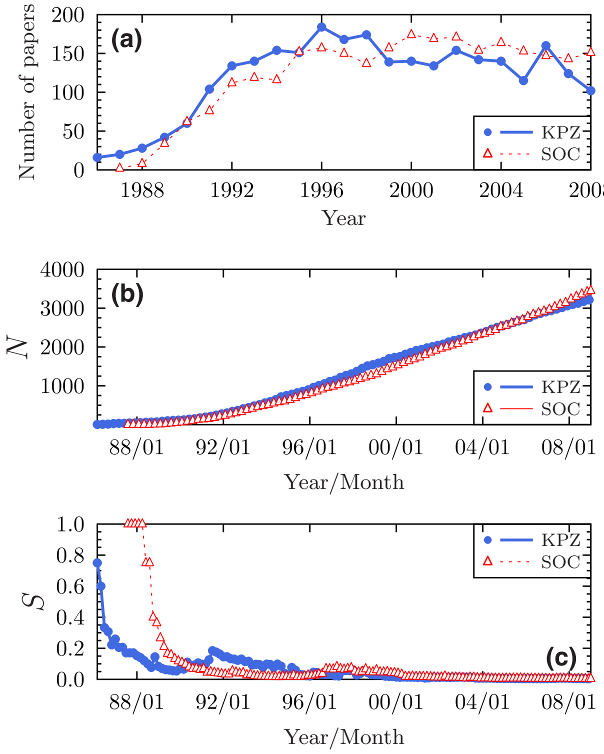


FIG. 12. (Color online) Coauthorship network evolution of declining subjects. (a) Plot of the numbers of papers published each year versus time for two fields, fractal surface growth obtained with Ref. [20], and self-organized criticality with Ref. [21] as the seed papers. They show steady-state behaviors after around the year 1996. (b) Plot of the numbers of distinct authors contributed to the papers in (a) versus time for the two subjects, which show monotonically increasing behaviors. New authors steadily join the coauthorship networks. (c) Plot of the relative sizes S of the LDGC of the coauthorship networks in both fields compared with the CNR network; S of these two fields degenerates below 0.01, an order of magnitude smaller, meaning that the giant component is practically no longer present.

VI. DISCUSSION

In the past ten years, more than 5000 papers have been published on the subject of the CNR by approximately 6800 researchers. Using the Web of Science database, we have traced those papers for 127 months. The evolution exhibits the percolation transition through which the giant component forms to establish global connectivity under the continuous supply of noninbred new members into a society. The coauthorship network currently appears to maintain diversity without sacrificing the internal evolution within groups, by a moderate value of model parameter $\alpha_{em}\approx 0.33$. However, in order to form a stable research discipline, both the growth and the connectedness are important. In this respect, the existence of the giant component spanning the system, even with link degradation, would represent the maturity of the subject in that it allows the exchange of ideas through the body of the community by occasional unconstrained collaborations that overcome the prevailing locality effect. Such a formation of a giant component would correspond to the emergence of the so-called invisible college [37]. However, such a dominant college supported by the core treelike giant component is shown to be highly dynamic, raising the possibility of continuous diversity and variations in the leading ideas and research trends. What is remarkable is that these evolutionary patterns appear robustly in other systems. In addition to the CNR, we chose two new topics in theoretical high-energy physics, the anti-DeSitter–conformal field theory duality conjecture [18] and the Randall-Sundrum model [19], and confirmed that their evolutionary pattern is similar to what we discussed for the CNR. On the other hand, when we consider a research field that is fading away, we get different patterns. For example, the core giant component spanning the coauthorship network disappears after link degradation, as observed in the cases of the fractal surface growth and the self-organization criticality, triggered by papers [20,21], respectively (Fig. 12). These two subjects have passed their heydays in the 1990s, and now their LDGC has degenerated even though new papers are published in a steady rate and the new comers still continue to enter the fields. Thus, the relative size of the coauthorship LDGC may be an indicator of the current state of a research field. Even though there may be other subject-specific factors that are responsible for some of the observed properties and therefore comparison between different fields has to be interpreted with care, our finding that they exhibit many shared patterns suggests the validity of common evolutionary mechanisms explored in this work.

We have shown that two microscopic mechanisms, the continuous influx of new nodes and groups and link formations strongly constrained by the locality effect, underlay the observed coauthorship evolution pattern. Moreover, the current state of the network was found to be nearly critical from the perspective of percolation theory. An important remaining question is how the system has located itself in such a delicate state. One appealing answer might be that it has done so in a self-organized way, calling for further studies in this direction.

ACKNOWLEDGMENTS

This work was supported by Mid-career Researcher Program through NRF grants funded by the MEST Grants No. 2010-0015066 (B.K.) and No. 2009-0080801 (K.-I.G.), and by KRCF (D.K.).

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[3] M. E. J. Newman, A.-L. Barabasi, and D. J. Watts, *Structure and Dynamics of Networks* (Princeton University Press, Princeton, NJ, 2006).
[4] G. Caldarelli, *Scale-Free Networks* (Oxford University Press, Oxford, 2007).
[5] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, England, 2004).
[6] B. O. Palsson, *Systems Biology: Properties of Reconstructed Networks* (Cambridge University Press, Cambridge, England, 2006).
[7] M. O. Jackson, *Social and Economic Networks* (Princeton University Press, Princeton, NJ, 2008).
[8] D. Lazer *et al.*, Science **323**, 721 (2009).
[9] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[10] S. Eubank, H. Guclu, V. S. A. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, Nature (London) **429**, 180 (2004).
[11] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, Proc. Natl. Acad. Sci. U.S.A. **104**, 7332 (2007).
[12] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[13] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[14] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[15] We excluded 11 papers with more than 20 authors, most of which were from large-scale biology experiments such as the human genome sequencing project.
[16] H. Jeong, Z. Néda, and A.-L. Barabási, EPL **61**, 567 (2003).
[17] G. Palla, A.-L. Barabási, and T. Vicsek, Nature (London) **446**, 664 (2007).
[18] J. M. Maldacena, Adv. Theor. Math. Phys. **2**, 231 (1998); Int. J. Theor. Phys. **38**, 1113 (1999); S. S. Gubser, I. R. Klebanov, and A. M. Polyakov, Phys. Lett. B **428**, 105 (1998); E. Witten, Adv. Theor. Math. Phys. **2**, 253 (1998).
[19] L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999).
[20] M. Kardar, G. Parisi, and Y.-C. Zhang, Phys. Rev. Lett. **56**, 889 (1986).
[21] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).
[22] B. J. Frey and D. Dueck, Science **315**, 972 (2007).
[23] C. Song, S. Havlin, and H. A. Makse, Nature (London) **433**, 392 (2005).
[24] K.-I. Goh, G. Salvi, B. Kahng, and D. Kim, Phys. Rev. Lett. **96**, 018701 (2006).
[25] D.-H. Kim, J.-D. Noh, and H. Jeong, Phys. Rev. E **70**, 046126 (2004).
[26] J. S. Kim, K.-I. Goh, B. Kahng, and D. Kim, Chaos **17**, 026116 (2007).
[27] J. S. Kim, K.-I. Goh, G. Salvi, E. Oh, B. Kahng, and D. Kim, Phys. Rev. E **75**, 016110 (2007).
[28] M. Llas, P. M. Gleiser, J. M. Lopez, and A. Diaz-Guilera, Phys. Rev. E **68**, 066101 (2003).
[29] G. Caldarelli, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **38**, 183 (2004).
[30] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).
[31] M. C. González, H. J. Herrmann, J. Kertész, and T. Vicsek, Physica A **379**, 307 (2007).
[32] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. Argollo de Menezes, K. Kaski, A.-L. Barabási, and J. Kertész, New J. Phys. **9**, 179 (2007).

[33] S. Saavedra, F. Reed-Tsochas, and B. Uzzi, Proc. Natl. Acad. Sci. U.S.A. **105**, 16466 (2008).
[34] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz, Phys. Rev. E **64**, 041902 (2001).
[35] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. E **64**, 066110 (2001).
[36] P. Holme and B. J. Kim, Phys. Rev. E **65**, 026107 (2002).
[37] R. Guimera, B. Uzzi, J. Spiro, and A. N. Amaral, Science **308**, 697 (2005).
[38] J. M. Kumpula, J.-P. Onnela, J. Saramäki, K. Kaski, and J. Kertész, Phys. Rev. Lett. **99**, 228701 (2007).