Smoluchowski Equation for Networks: Merger Induced Intermittent Giant Node Formation and Degree Gap

Hayato Goto¹ · Eduardo Viegas² · Henrik Jeldtoft Jensen^{1,2} ·
Hideki Takayasu^{1,3} · Misako Takayasu¹

Received: 7 March 2018 / Accepted: 25 May 2018 / Published online: 8 June 2018
© The Author(s) 2018

Abstract The dynamical phase diagram of a network undergoing annihilation, creation, and coagulation of nodes is found to exhibit two regimes controlled by the combined effect of preferential attachment for initiator and target nodes during coagulation and for link assignment to new nodes. The first regime exhibits smooth dynamics and power law degree distributions. In the second regime, giant degree nodes and gaps in the degree distribution are formed intermittently. Data for the Japanese firm network in 1994 and 2014 suggests that this network is moving towards the intermittent switching region.

Keywords Network stability · Complex systems · Econophysics · Network phase transitions · Stochastic processes & statistics · Sol–gel process

✉ Misako Takayasu
takayasu.m.aa@m.titech.ac.jp

Hayato Goto
hayato198634@gmail.com

Eduardo Viegas
e.viegas11@imperial.ac.uk

Henrik Jeldtoft Jensen
h.jensen@imperial.ac.uk

Hideki Takayasu
takayasu.h.aa@m.titech.ac.jp

¹ Institute of Innovative Research, Tokyo Institute of Technology, 4259, Nagatsuta-cho, Yokohama 226-8502, Japan

² Centre for Complexity Science and Department of Mathematics, Imperial College London, London SW7 2AZ, UK

³ Sony Computer Science Laboratories, 3-14-13, Higashi-Gotanda, Shinagawa-ku, Tokyo 141-0022, Japan

1 Introduction

The Smoluchowski equation was developed to describe the distribution of clusters containing specific numbers of particles formed through coagulation processes conserving mass [1]. This has been used widely in many fields such as colloid, aerosol, virus, fish school, and asteroid [2–14]. The approach was generalized by Miura et al. in [15] to analyze the formation of networks with application to the Japanese firm network. A model (MTT model) is proposed, which encompasses three processes of node creation, annihilation, and coagulation, together with the preferential attachment rule [16, 17], that is, a new link is more likely to be attached to a node having a larger number of links, so that a new node will be connected to node i , depending on the connectivity k_i^λ of that node. The third process is intended to represent mergers and acquisitions amongst firms. The behavior differs from the traditional Smoluchowski method since not only particles (i.e. nodes) but also links must be taken into account. Coagulation is studied from the viewpoint of the number of links (i.e. degree) connected to the nodes. The stationary state of the model exhibits a power law behavior for the degree distribution with exponents that depend on model parameters which allows the model to replicate the observations on the Japanese inter-firm transaction data for which the distribution of the number of transaction partners follow a power law with a cumulative exponent 1.3 ± 0.1 (for probability density, the exponent is 2.3 ± 0.1) [18] (see Appendix A). The previous study was restricted to replicating the distribution of the network by making use of parameter values for both preferential attachment exponents (i.e. for a newcomer at creation process and for an acquirer at coagulation process) as well as the occurrence probabilities of node creation, annihilation, and coagulation. Here, we broaden the scope in order to center on the dynamics of the network formation and the emerging phase transactions arising from such dynamics.

In this study, we generalize the MTT model and discover two new types of long time dynamics and we determine the locus of control parameters separating the different types of asymptotic dynamics. Our generalized MTT model includes a more detailed analysis of the merger process in which we distinguish between a firm (i.e. node) acquiring another firm, that is initiating a merger, and a firm being the target of the acquisition, i.e. being selected for merger and has been verified validation by comparing empirical analysis results with simulation results (see [19] for details of the validation). The first process happens with a probability proportional to k_a^α , where k_a is the degree of the acquiring node. The acquiring node selects the target for acquisition with a probability proportional to the degree k_t^β where k_t is the degree of the target. This represents a generalization of the original MTT model. We study this generalized process numerically and analytically using the Smoluchowski equation representing this process. To do that we have included a nonlinear coagulation effect by introducing a collision kernel proportional to $k_a^\alpha k_t^\beta$ [19]. We plot the phase diagram consisting of two axes, the vertical axis given by the preferential attachment exponent for newcomers (λ), and the horizontal axis given by the sum of coagulation kernel's exponents ($\alpha + \beta$). By observing temporal behaviors of the system for each point in this phase diagram through intensive numerical simulations, we are able to quantify the dynamical properties resulting from the maximum value of the difference of the sizes of between the largest and the second largest nodes, denoted as Diff. According to this value, we categorize the phase diagram into Type I and Type II, with marginal critical cases. Type I relates to the case where the value of Diff is small as shown in Fig. 1b and the size distribution is steady as shown in Fig. 1a. In contrast, Type II consists of the case that Diff is very large with intermittent appearance of giant degree node showing saw-teeth oscillations and the corresponding size distribution is not steady as typically shown in Fig. 1e and f. The marginal cases, typically shown in Fig.

1c and d are given by the value of Diff around 10^3 . The regions of Type I and II are separated by the marginal critical cases. We apply each value of model-parameter from the Japanese firm network data analysis [19] and find that the firm network during the last few decades has moved towards the Type II region, the region where a giant degree node will be formed.

2 Model Analysis

Time evolution of network growth is modeled by the following three processes which are chosen randomly at each time step.

Newcomer A new node with n links is added into the system. Each new link is randomly connected to an already existing node with a probability given by the preferential attachment rule [16, 17] with the exponent λ . The allocation of links to an existing node of degree k follows a probability that is proportional to k^λ . That is, the probability Π that a new node will be connected to node i depends on the connectivity k_i^λ of that node, so that $\Pi = k_i^\lambda / \sum_j k_j^\lambda = N k_i^\lambda / \langle k^\lambda \rangle$, where N is total number of nodes which fluctuates over time.

Annihilation A node, chosen randomly with uniform probability, is removed along with all links connected to this node.

Coagulation A pair of nodes, an acquiring node a and a target node t , are chosen randomly taking into account their link numbers to become proportional to the coagulation kernel, $K(k_a, k_t) \propto k_a^\alpha k_t^\beta$. Node t is absorbed by the node a . In this process the k_t links emerging from the target node are transferred to the acquiring node. The merger (coagulation) process conserves the sum of the link numbers except overlapping links, that is, multiple links between the same nodes is not allowed within the simulation.

First, we study the model by use of Monte Carlo simulations (see Appendix B for the method). The initial state consists of 10,000 nodes with zero links. We then develop the network by applying the three processes: newcomer, annihilation and coagulation (merger and acquisition) with the rates r_n , r_a and r_c . By comparing with the Japanese firm network in 2014, see [19, 20], the rates are chosen as 0.5 : 0.4 : 0.1. The simulations are done as described in previous studies [15, 19]. We are particularly interested in the evolution of the degree distribution and the appearance of a gap between the link numbers of the high degree nodes. Figure 1 shows the salient results of the simulations for three different choices of values for the three exponents (λ, α, β) , namely $(1.0, 0.8, 0.6)$, $(1.0, 1.1, 0.7)$, and $(1.0, 1.4, 0.8)$. In panels a, c and e the black squares, pink circles, and light-green triangles indicate the distribution found after 1.5×10^5 , 2.5×10^5 , and 3.5×10^5 time steps, respectively. Panel b, d and f show the corresponding time dependence (in semi-log) of the number of links k for the nodes with the largest degree (red), second largest degree (green), 10-th largest degree (blue), 100-th largest degree (light blue) and 1,000-th largest degree (gray) and for the median (yellow). We notice that as α and β are increased, the temporal fluctuation becomes larger.

For the smaller values of α and β the nodes of highest degree ranks have link numbers of the same order of magnitude as shown in panel a and b. We call this Type I behavior. A change occurs as the α and β exceed certain threshold values, this is shown in panel e and f. The time dependent degree distribution develops a form that we approximate below by a power law for probability density $c_k(\tau) \propto k^{-\gamma}$ with an approximate exponent $\gamma = 2.3 \pm 0.1$, where $c_k(\tau)$ is probability density of number of nodes with link k . Concurrently, a significant gap appears intermittently between the number of links of the node with the highest degree and the link number of the other nodes of lower rank, see panel f in Fig. 1. We call this Type II behavior. A noticeable decrease occurs in the link number of the nodes with rank smaller

Fig. 1 **a, c, e** Cumulative distribution of link number k in logarithm scales, with parameter-sets $(\lambda, \alpha, \beta) = (1.0, 0.8, 0.6)$ as **(a)**, $(1.0, 1.1, 0.7)$ as **(c)** and $(1.0, 1.4, 0.8)$ as **(e)**, respectively. The rates for each of the processes are $r_n : r_a : r_c = 0.5 : 0.4 : 0.1$. The black squares, pink circles, and light-green triangles indicate the distribution found after 150, 000 step, 250, 000 step, and 350, 000 step, respectively in **a, c**, and 210, 000 step, 240, 000 step, and 280, 000 step, respectively in **e**. The dashed-lines indicate slope $-1.3 = -\gamma + 1$. The distribution shown by each of the coloured symbols in **a, c**, and **e** relate to the times given by the vertical lines in **b, d**, and **f** of the same colors. **b, d, f** Simulated time dependence of the link number k corresponding to the distributions **a, c** and **e**, respectively. Each line shows 1st (red), 2nd (green), 10th (blue), 100th (light blue), 1,000th (gray) and median (yellow) link number k

than the largest, while the largest rapidly increases its link number. The gap between 1st and 2nd grow quickly, then, stays at a certain level for a while and the 1st disappears suddenly by annihilation in the case of Type II. This categorization of Type I and II is available numerically for any initial condition after a brief transient state. This behavior will be discussed below in relation to Eq. (2).

3 Smoluchowski Equation

As the dynamics switches from Type I to Type II, the intermittent appearance of a node with giant degree after the appearance of power-law degree distribution becomes evident. This is similar to the formation of the gel in the sol–gel phase transition, which is an aggregation phenomena that has a long history in statistical physics. Within this process, it is possible to observe the existence of a steady state which obeys a robust power law distribution in the aggregation system with continuous injection and removal near the gelation point [4, 5, 8], so to understand better the nature of the change between the two behaviors we establish a master equation in the spirit of Smoluchowski's equation used to describe sol–gel phase transitions which gives the time evolution of $c_k(\tau)$. The following distinct processes and their probabilities describe the dynamics:

Increase of the number of nodes with k links

- (i) $\frac{(k-1)^\lambda+1}{\langle k^\lambda \rangle+1} c_{k-1}(\tau)$; for connecting one new link to a node with $k-1$ links so that the resulting node holds k links.
- (ii) $\delta_{k,4}$; for directly creating a node of 4 links (δ denotes Kronecker delta). This constant is derived by the ratio between the total number of transaction partners (c. 4.1 million) and the total number of firms (c. 1.1 million) observed within the Japanese firm network real data in 2014 [19].
- (iii) $(k+1)c_{k+1}(\tau)$; for annihilating a node which is connecting to a node of $k+1$ links.
- (iv) $\sum_{k'=0}^k \frac{k'^\alpha+1}{\langle k^\alpha \rangle+1} c_{k'}(\tau) \frac{(k-k')^\beta+1}{\langle k^\beta \rangle+1} c_{k-k'}(\tau)$; for merging a node of k' links as an acquirer and that of $k-k'$ links as a target.

Decrease of the number of nodes with k links

- (v) $\frac{k^\lambda+1}{\langle k^\lambda \rangle+1} c_k(\tau)$; for creating a node which is connecting to a node of k links.
- (vi) $kc_k(\tau)$; for annihilating a node which is connecting to a node of k links.
- (vii) $c_k(\tau)$; for directly annihilating a node of k links.
- (viii) $\sum_{k'=0}^\infty \frac{k'^\alpha+1}{\langle k^\alpha \rangle+1} c_{k'}(\tau) \frac{k^\beta+1}{\langle k^\beta \rangle+1} c_k(\tau)$; for merging a node of k' links as an acquirer and that of k links as a target.
- (ix) $\sum_{k'=0}^\infty \frac{k'^\alpha+1}{\langle k^\alpha \rangle+1} c_{k'}(\tau) \frac{k'^\beta+1}{\langle k^\beta \rangle+1} c_k(\tau)$; for merging between a node of k links as an acquirer and that of k' links as a target.

In order to allow isolated nodes (degree zero nodes) to be able to connect to newcomers and to participate in the merger events, we have added one to the degree of the involved nodes in the numerator and denominator of the kernels describing these processes. The resulting master equation for the time evolution of c_k referring to the previous study [15, 21] is given by;

$$\begin{aligned}
 & (\langle N \rangle + r_n - r_a - r_c) c_k(\tau + 1) - \langle N \rangle c_k(\tau) \\
 &= r_n \left[\frac{(k-1)^\lambda + 1}{\langle k^\lambda \rangle + 1} c_{k-1}(\tau) - \frac{k^\lambda + 1}{\langle k^\lambda \rangle + 1} c_k(\tau) + \delta_{k,4} \right] \\
 &+ r_a \left[(k+1)c_{k+1}(\tau) - kc_k(\tau) - c_k(\tau) \right] \\
 &+ r_c \left[\sum_{k'=0}^k \frac{k'^\alpha + 1}{\langle k^\alpha \rangle + 1} c_{k'}(\tau) \frac{(k-k')^\beta + 1}{\langle k^\beta \rangle + 1} c_{k-k'}(\tau) \right. \\
 &\quad \left. - \sum_{k'=0}^\infty \frac{k'^\alpha + 1}{\langle k^\alpha \rangle + 1} c_{k'}(\tau) \frac{k^\beta + 1}{\langle k^\beta \rangle + 1} c_k(\tau) - \sum_{k'=0}^\infty \frac{k^\alpha + 1}{\langle k^\alpha \rangle + 1} c_k(\tau) \frac{k'^\beta + 1}{\langle k^\beta \rangle + 1} c_{k'}(\tau) \right]. \quad (1)
 \end{aligned}$$

Note that $\langle N \rangle$ represents the mean (i.e. ensemble average) of total number of nodes, which $\langle N \rangle = \sum_{s=1}^L N_s(\tau)/L$ where $N_s(\tau)$ the s -th sample's total number of nodes at time τ and L the total number of samples, so that $\langle N \rangle_{c_k}(\tau)$ is the mean number of nodes with degree k at time τ . Moreover, the mean number of nodes added (or deleted) to the network per time is $r_n - r_a - r_c$. Therefore, the mean number of nodes with degree k after one step is $\langle N \rangle_{c_k}(\tau + 1) + (r_n - r_a - r_c)c_k(\tau + 1)$, where $c_k(\tau + 1)$ is the new value of $c_k(\tau)$.

We now describe the analytic criteria we use to determine the boundary between Type I and Type II dynamics. We follow previous studies [22–24] of the Smoluchowski coagulation equation in view of sol–gel phase transition. The formation of the gel can be identified from the time dependence of the moments restricted by a given upper-bound link number Q , i.e. $M_{1,Q}(\tau) \equiv \sum_{k=1}^Q k c_k$. The gelation transition, i.e. the formation of a giant node, is determined by the behavior of the derivative $\Lambda(Q) \equiv dM_{1,Q}/d\tau$. The gel phase appears when $\Lambda(Q_1) > 0$ while $\Lambda(Q_2) < 0$, where Q_1 and Q_2 denote the link sizes of the largest node and the second largest node, respectively [23] (see Appendix C for a strong divergence of a size of a node of maximum degree Q_1 in finite time which is effected by aggregating between large nodes). For theoretical evaluation of Eq. (1), we pay attention to the boundary between Type I and Type II dynamics and we assume that the degree distribution follows a power law $c_k(\tau) = H(\tau)k^{-\gamma}$, based on findings from previous studies [3–5, 8, 22–24]. Further more we assume stationarity for the averaged number of nodes by choosing $r_n = 1/2$ and $r_a = 1/2 - r_c$, we have the following evaluation (see Appendix D for detail derivation).

$$\Lambda(Q) \approx 1 - \frac{Q^{\lambda-\gamma+1}}{\langle k^\lambda \rangle + 1} H(\tau) + (1/2 - r_c) \left[H(\tau) [Q^{2-\gamma} + (1-\gamma)Q^{1-\gamma}] - 2\langle k \rangle \right] - 2r_c [H(\tau)]^2 Q^{3+\alpha+\beta-2\gamma} \bar{B}. \quad (2)$$

Here $\bar{B} = [B(\bar{\alpha} + 1, \bar{\beta} + 1)/\bar{\beta}] + [B(\bar{\beta} + 1, \bar{\alpha} + 1)/\bar{\alpha}]$ and B expresses beta function, where $\bar{\alpha} = \alpha - \gamma + 2$, $\bar{\beta} = \beta - \gamma + 2$. We identify the boundary of the transition between Type I and Type II as the locus of the exponents λ and $\alpha + \beta$ for which the time derivative given in Eq. (2) can become negative in the limit of large Q . We note that in Eq. (2) $\bar{B}$ is a constant and the normalization factor $H(\tau)$ is essentially independent of the number of nodes for $\gamma > 2$. So the sign of $\Lambda(Q)$ for large Q is estimated by the exponents of Q of the terms in the right hand side of Eq. (2). We fix γ at 2.3 ± 0.1 , suggested by the simulations and by the real value estimated for the Japanese firm network [18]. Then, Eq. (2) derives the following two conditions for $\Lambda(Q)$.

- (A) $r_c = 0 \rightarrow \lambda - \gamma + 1 > 0 \rightarrow \lambda > 1.3 \pm 0.1$,
- (B) $1/2 > r_c > 0 \rightarrow (\lambda - \gamma + 1 > 0)$ and $(3 + \alpha + \beta - 2\gamma > 0) \rightarrow (\lambda > 1.3 \pm 0.1)$ and $(\alpha + \beta > 1.6 \pm 0.2)$.

We stress that this is the prediction for a node of giant degree to be able to form and we take this as an estimate of the boundary between the two types of dynamics observed in the simulations. We are going to check the validity of the analytic estimations, (A) and (B), by introducing phase diagrams as follows. To detect the onset of Type II dynamics, corresponding to the intermittent formation of a giant node, we measure the maximum difference between the first and second highest number of links, Q_1 and Q_2 during 500,000 Monte Carlo steps of the processes newcomer, annihilation and coagulation. We pay attention to the following three cases. (a) $r_n = 0.5$, $r_a = 0.49$, $r_c = 0.01$, as a representative case that the effect of coagulation is very weak and the condition (A) should holds. (b) $r_n = 0.5$, $r_a = 0.01$, $r_c = 0.49$, as a representative case that the effect of coagulation is very strong and the condition

Fig. 2 Maximum difference between the first and second highest number of links. The dashed lines refer to the estimate of the boundary for the onset of Type II dynamics given by **a** and **b**. The generation rates $r_n : r_a : r_c$ are **a** $0.5 : 0.49 : 0.01$ with the dashed line referring to **a**, **(b)** $0.5 : 0.01 : 0.49$ with the dashed lines referring to **b**, **(c)** $0.5 : 0.4 : 0.1$ with the dashed lines referring to **b**, with the black circles showing the parameters $(\lambda, \alpha, \beta) = (1.0, 0.8, 0.6)$, $(1.0, 1.1, 0.7)$, and $(1.0, 1.4, 0.8)$. The first two set of values correspond to the Japanese firm network data for 1994 and 2014, and the last is a simple linear extrapolation to 2034 assuming steady growth of the exponents

(B) should holds. (c) $r_n = 0.5, r_a = 0.4, r_c = 0.1$, as a representative case that corresponds to the observed rates for Japanese business firms in the real world [19]. Each figure in Fig. 2 shows the phase diagram of the corresponding case using a colored heatmap showing the difference $Q_1 - Q_2$ with the theoretical boundary given by the dotted lines.

We notice that when the rate for merger, or coagulation, is very low, Fig. 2 panel a, the phase boundary obtained from the analytic analysis of the time dependence of the moments coincides with the observed change in $Q_1 - Q_2$ in the simulations. For larger values of the rate r_c for merger events, the analytic prediction deviates quantitatively from the simulated boundary between the two types of dynamics. This is natural since the Smoluchowski is a mean field approach and allowing mergers increase significantly the stochastic fluctuations in the dynamics.

4 Discussion

We have analytically and numerically studied the threshold condition for the formation of a giant component, or gelation, in a network subject to the three processes of creation, annihilation and merger.

Analysis of the Japanese firms network shows that the preferential attachment exponent λ remains constant at about 1. In contrast to the α and β exponent describing the merger,

which show a significant change with time. Specifically analysis of 1994 and 2014 leads to the following estimates of the exponents $(\lambda, \alpha, \beta) = (1.0, 0.8, 0.6)$ and $(1.0, 1.1, 0.7)$, respectively [19]. This analysis also shows acceleration in the number of firms with large numbers of links over the last twenty years. This corresponds to the growth of the exponents α, β as a result of substantial merger and acquisitions activities. Therefore, our theoretical analysis indicates that the dynamics of the inter-firm network in 1994 and 2014 is likely to be located near and at the threshold for the onset of giant node formation, respectively. Extrapolation of this trend, see Fig. 2, suggests that the ecosystem of Japanese firms is moving into the Type II region and that intermittent formation and disappearance of giant components is likely to upset the economy with increasing intensity in the near future. Such behavior may obviously be consequential for the sustainability [25] and health of the Japanese economy. The results of our work, therefore, provide quantitatively measurable indicators that can aid assessment of the evolution and the stability of an economical ecosystem.

Conceptually, appearance of giant degree nodes in evolving network systems may look similar to the study developed by et al. [26]. They analytically showed appearance of giant nodes by the effect of stronger preferential attachment of newly added nodes in the limit of infinite time. By contrast, we add two more effects, coagulation of nodes and spontaneous annihilation, and find that the effect of coagulation of nodes can be so strong for some parameter region that we can observe tendency that the link number of the largest degree node diverges in a finite time corresponding to the physical phenomenon of gelation, that never occurs by the preferential attachment alone.

Additionally, the appearance of saw-teeth oscillation in Fig. 1f may also have apparent similarities to the study developed by Ball et al. [27] and Matveev et al. [28], where steady oscillations in the aggregation systems with removal of cluster or fragmentation of cluster into monomers is detected. In particular, Matveev et al. [28] obtained an analytical solution for the steady state cluster size distribution with numerical confirmation of the relaxation of the size distribution to this steady-state form. However, both of these studies aim to evaluate the steady oscillatory behavior of cluster size distribution, with focus on the kinetics of irreversible aggregation and clustering phenomena of the Smoluchowski coagulation equation. In contrast, our study focuses on the sol–gel phase transition, and the impact (separately and interactively) of each fundamental process of newcomers, annihilations, and coagulations, since these are the fundamental cornerstones to the understanding of business transaction networks and systems. As a result, our findings are unique and cannot be derived by these earlier studies. In addition, there is an important methodological difference within our work as the oscillation period after gelation is depended on the random event of the gel-node's removal, so that the dynamics are nontrivial.

Acknowledgements This study is supported by the Grant-in-Aid for Scientific Research (C), Grant Number 24540395 and (B), Grant Number 26310207, HPCI System Research Project (Project ID:hp160261), and Strategic International Research Cooperative Program, Japan Science and Technology Agency (JST). We would like to express our appreciation to the Center for Teikoku Databank Advanced Data Analysis and Modeling, Tokyo Institute of Technology for providing the datasets.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix A: Cumulative Distribution of Firm Size

Figure 3 shows the cumulative distributions of firm size in Japan. As for the distribution of the annual sales in Panel a, an approximated power law exponent slightly decreases from 1.1 (1995) to 0.9 (2015). This indicates that large size of firms grow larger and middle-small size of firms shrink in their sizes; the gap widens. As for the distribution of the number of business transactions in Panel b, an approximated power law exponent keeps 1.3 from 1995 to 2015. However, a difference between the first and second highest number of business transactions grows larger. This also indicates that the firm-ecosystem comes close to the gelation phase.

Appendix B: Monte Carlo Simulation

As for our numerical analysis, we conducted Monte Carlo simulation as follows;

Step1 Start with N_0 firms without any business transactions ($N_0 = 10,000$ in Figs. 1 and 2).

Step2 Choose one of the following three events stochastically.

Newcomer A new node with n links is added into the system. Each new link is randomly connected to an already existing node with a probability given by the preferential attachment rule [16, 17] with exponent λ . We decided that each new node has four links in order to keep the rate between the number of firms and the number of transaction partners [19]. This condition is not essential, and the main results do not change by replacing the numbers of newcomer's links by random numbers with the same average, for example.

Annihilation A node, chosen randomly with uniform probability, is removed along with all links connected to this node. This is because age of firms follow exponential distribution; it is roughly consistent with the simple assumption that a firm disappears randomly following a Poisson process [15, 19].

Coagulation A pair of nodes, an acquiring node a and a target node t , are chosen randomly taking into account their link numbers to become proportional to the

Fig. 3 Cumulative distribution of firm size in Japan in log–log plot. Each panel **a** and **b** shows the distribution of annual sales (million yen) and number of business transactions. Each red, green, and blue line shows the distribution in 1995, 2005, and 2015, respectively. Panel **a** each upper and lower dashed-line shows an approximated power law distribution with slope 0.9 for 2015 and 1.1 for 1995, respectively. Panel **b** dashed-line shows an approximated power law distribution with slope 1.3

Fig. 4 Panel **a** shows the three basic processes for nodes: newcomers, annihilations, and coagulations in our simulation model. Panel **b** shows time step flow chart of our simulation model

coagulation kernel, $K(k_a, k_t) \propto k_a^\alpha k_t^\beta$ [19]. The merger (coagulation) process conserves the sum of the link numbers except overlapping links.

Step3 Repeat Step2 for 5, 000, 000 times.

Figure 4 shows our algorithm by schematic diagram. The occurrence probabilities of newcomers, annihilations, and coagulations are denoted by r_n , r_a , and r_c , respectively, satisfying $r_n + r_a + r_c = 1$ (systems are quasi-stationally states when $r_n = r_a + r_c$ in the number of nodes).

Appendix C: Simulated Time Dependence of Link Number k in Open Systems

Figure 5 shows the simulated time dependence of link number k in an open system with, which nodes may be injected and be deleted (i.e. enter and exit), initial 10, 000 nodes. Panel a the size of the node of maximum degree Q_1 . Panel b the moment $M_{1,Q_2}(\tau)$, where Q_2 is the size of the 2nd largest, and Panel c contains $dM_{1,Q_2}/d\tau$.

The rates $r_n : r_a : r_c$ are 0.5 : 0.0 : 0.5. Red line, green line, and blue line show the parameters $(\lambda, \alpha, \beta) = (1.0, 0.8, 0.6)$, $(1.0, 1.1, 0.7)$, and $(1.0, 1.4, 0.8)$, respectively. Each colored area shows the critical point of gelation τ_c as given by the criteria $dM_{1,Q_2}/d\tau = 0$. The time required to reach the onset of gelation τ_c depends on the size of parameters and there would be parameter threshold effects as phase transition in the systems. We note that decrease of the size of the second largest, Q_2 , is caused both by annihilation of the nodes which are already connecting to Q_2 and by less attachment of newcomers to this node as the largest node attracts most of newcomers by the preferential attachment with its dominant size. Namely, the effect of annihilation plays an important role to realize the critical time, τ_c , at a finite time, while gelation occurs only asymptotically in the infinite time in the model of Krapivsky et al. [26].

Fig. 5 Simulated time dependence of link number k in open systems with initial 10,000 nodes. Panel **a** contains the number of links of the node with maximum degree Q_1 . Panel **b** shows $M_{1,Q_2}(\tau)$, and Panel **c** contains $dM_{1,Q_2}/d\tau$. The rates $r_n : r_a : r_c = 0.5 : 0.0 : 0.5$. Each red, green, and blue line shows the parameter-set $(\lambda, \alpha, \beta) = (1.0, 0.8, 0.6)$, $(1.0, 1.1, 0.7)$, and $(1.0, 1.4, 0.8)$, respectively. Each colored area shows the onset of gelation τ_c according to the criteria $dM_{1,Q_2}/d\tau = 0$

Appendix D: The Master Equation

We consider the version of the Smoluchowski equation for the link number k , $c_k(\tau)$ relevant to the MTT model [15]. For the coagulation kernel we use the empirical form $K(k_a, k_t) \propto k_a^\alpha k_t^\beta$ [19] and get Eq. (1). The rates for creation, annihilation, and coagulation are denoted by r_n , r_a , and r_c , respectively, we have the following equation approximately by assuming a quasi-stationary state for the mean node numbers, which is realized by the set of parameters, $r_n = 1/2$, $r_a = 1/2 - r_c$.

$$\begin{aligned}
 \langle N \rangle dc_k/d\tau = & \frac{1}{\langle k^\lambda \rangle + 1} \left[[(k-1)^\lambda + 1]c_{k-1} - (k^\lambda + 1)c_k \right] \\
 & + (1/2 - r_c)(k+1)(c_{k+1} - c_k) \\
 & + \frac{r_c}{(\langle k^\alpha \rangle + 1)(\langle k^\beta \rangle + 1)} \left[\sum_{k'=0}^k (k'^\alpha + 1)[(k-k')^\beta + 1]c_{k'}c_{k-k'} \right. \\
 & \left. - \sum_{k'=0}^{\infty} (k'^\alpha + 1)(k^\beta + 1)c_{k'}c_k - \sum_{k'=0}^{\infty} (k^\alpha + 1)(k'^\beta + 1)c_k c_{k'} \right]. \quad (3)
 \end{aligned}$$

For simplicity we write $c_k(\tau)$ simply as c_k . By use of symmetries, Eq. (3) is rewritten as

$$\begin{aligned}
\langle N \rangle dc_k/d\tau &= \frac{1}{\langle k^\lambda \rangle + 1} \left[[(k-1)^\lambda + 1]c_{k-1} - (k^\lambda + 1)c_k \right] \\
&\quad + (1/2 - r_c)(k+1)(c_{k+1} - c_k) \\
&\quad + \frac{r_c}{(\langle k^\alpha \rangle + 1)(\langle k^\beta \rangle + 1)} \left[\sum_{k_a+k_t=k} (k_a^\alpha k_t^\beta + k_a^\beta k_t^\alpha) c_{k_a} c_{k_t} \right. \\
&\quad \left. - 2 \sum_{k_t=0}^{\infty} (k^\alpha k_t^\beta + k^\beta k_t^\alpha) c_k c_{k_t} \right]. \quad (4)
\end{aligned}$$

By changing the scale factors for time τ and distribution c_k as $\langle N \rangle \tau$, c_k and $(\langle k^\alpha \rangle + 1)(\langle k^\beta \rangle + 1)c_k$, we have the following simpler representation of Eq. (4), which does not change the results.

$$\begin{aligned}
dc_k/d\tau &= \frac{1}{\langle k^\lambda \rangle + 1} \left[[(k-1)^\lambda + 1]c_{k-1} - (k^\lambda + 1)c_k \right] \\
&\quad + (1/2 - r_c)(k+1)(c_{k+1} - c_k) \\
&\quad + r_c \left[\sum_{k_a+k_t=k} (k_a^\alpha k_t^\beta + k_a^\beta k_t^\alpha) c_{k_a} c_{k_t} - 2 \sum_{k_t=0}^{\infty} (k^\alpha k_t^\beta + k^\beta k_t^\alpha) c_k c_{k_t} \right]. \quad (5)
\end{aligned}$$

To study the criteria in Eq. (2), we consider $M_{\lambda,Q} \equiv \sum_{k=0}^Q k^\lambda c_k$ [22,23]. We multiply both sides by k and sum these from 0 to Q in Eq. (5), where Q is maximum size of k . This way we obtain for the creation processes

$$\begin{aligned}
&\frac{1}{\langle k^\lambda \rangle + 1} \sum_{k=0}^Q k \left[[(k-1)^\lambda + 1]c_{k-1} - (k^\lambda + 1)c_k \right] \\
&= \frac{1}{\langle k^\lambda \rangle + 1} \left[\sum_{k=0}^{Q-1} (k+1)(k^\lambda + 1)c_k - \sum_{k=0}^Q (k^{\lambda+1} + k)c_k \right] \\
&= \frac{1}{\langle k^\lambda \rangle + 1} \left[\sum_{k=0}^{Q-1} (k^{\lambda+1} + k^\lambda + k + 1)c_k - M_{\lambda+1,Q} - M_{1,Q} \right] \\
&= \frac{1}{\langle k^\lambda \rangle + 1} (M_{\lambda+1,Q-1} + M_{\lambda,Q-1} + M_{1,Q-1} + M_{0,Q-1} - M_{\lambda+1,Q} - M_{1,Q}) \\
&= \frac{1}{\langle k^\lambda \rangle + 1} (M_{\lambda,Q-1} + M_{0,Q-1} - Q^{\lambda+1}c_Q - Qc_Q), \quad (6)
\end{aligned}$$

and for the annihilation processes

$$\begin{aligned}
&(1/2 - r_c) \sum_{k=0}^Q k(k+1)(c_{k+1} - c_k) \\
&= (1/2 - r_c) \left[\sum_{k=0}^Q [(k+1)^2 - (k+1)]c_{k+1} - \sum_{k=0}^Q (k^2 + k)c_k \right] \\
&= (1/2 - r_c) \left[\sum_{k=0}^{Q+1} (k^2 - k)c_k - M_{2,Q} - M_{1,Q} \right]
\end{aligned}$$

$$\begin{aligned}
&= (1/2 - r_c) \left[\sum_{k=0}^Q (k^2 - k) c_k + [(Q+1)^2 - (Q+1)] c_{Q+1} - M_{2,Q} - M_{1,Q} \right] \\
&= (1/2 - r_c) [M_{2,Q} - M_{1,Q} + Q(Q+1) c_{Q+1} - M_{2,Q} - M_{1,Q}] \\
&= (1/2 - r_c) [Q(Q+1) c_{Q+1} - 2M_{1,Q}]. \tag{7}
\end{aligned}$$

Finally for the coagulation processes;

$$\begin{aligned}
&r_c \left[\sum_{k=0}^Q k \sum_{k_a+k_t=k} (k_a^\alpha k_t^\beta + k_a^\beta k_t^\alpha) c_{k_a} c_{k_t} - 2 \sum_{k=0}^Q k \sum_{k_t=0}^{\infty} (k^\alpha k_t^\beta + k^\beta k_t^\alpha) c_k c_{k_t} \right] \\
&= r_c \left[\sum_{k=0}^Q \sum_{k_a+k_t=k} (k_a + k_t) (k_a^\alpha k_t^\beta + k_a^\beta k_t^\alpha) c_{k_a} c_{k_t} - 2 \sum_{k_a=0}^Q \sum_{k_t=0}^{\infty} k_a (k_a^\alpha k_t^\beta + k_a^\beta k_t^\alpha) c_{k_a} c_{k_t} \right] \\
&= r_c \left[\sum_{k_a=0}^Q \sum_{k_t=0}^{Q-k_a} 2k_a (k_a^\alpha k_t^\beta + k_a^\beta k_t^\alpha) c_{k_a} c_{k_t} - 2 \sum_{k_a=0}^Q \sum_{k_t=0}^{\infty} k_a (k_a^\alpha k_t^\beta + k_a^\beta k_t^\alpha) c_{k_a} c_{k_t} \right] \\
&= -2r_c \sum_{k_a=0}^Q \sum_{k_t=Q-k_a+1}^{\infty} k_a (k_a^\alpha k_t^\beta + k_a^\beta k_t^\alpha) c_{k_a} c_{k_t}. \tag{8}
\end{aligned}$$

By combining Eqs. (6), (7) and (8), we get Eq. (9).

$$\begin{aligned}
dM_{1,Q}/d\tau &= \frac{1}{\langle k^\lambda \rangle + 1} (M_{\lambda,Q-1} + M_{0,Q-1} - Q^{\lambda+1} c_Q - Q c_Q) \\
&\quad + (1/2 - r_c) [Q(Q+1) c_{Q+1} - 2M_{1,Q}] \\
&\quad - 2r_c \sum_{k_a=0}^Q \sum_{k_t=Q-k_a+1}^{\infty} k_a (k_a^\alpha k_t^\beta + k_a^\beta k_t^\alpha) c_{k_a} c_{k_t}. \tag{9}
\end{aligned}$$

Substitution of $c_k = H(\tau)k^{-\gamma}$ at the gelation point [24] leads to Eqs. (10), (11) and (12) for $k \gg 1$. The creation processes;

$$\begin{aligned}
&\frac{1}{\langle k^\lambda \rangle + 1} (M_{\lambda,Q-1} + M_{0,Q-1} - Q^{\lambda+1} H(\tau) Q^{-\gamma} - Q H(\tau) Q^{-\gamma}) \\
&\approx \frac{1}{\langle k^\lambda \rangle + 1} [\langle k^\lambda \rangle + 1 - Q(Q^\lambda + 1) H(\tau) Q^{-\gamma}] \\
&\approx 1 - \frac{Q^{\lambda-\gamma+1}}{\langle k^\lambda \rangle + 1} H(\tau). \tag{10}
\end{aligned}$$

The annihilation processes;

$$\begin{aligned}
&(1/2 - r_c) [Q(Q+1) H(\tau) (Q+1)^{-\gamma} - 2M_{1,Q}] \\
&= (1/2 - r_c) \left[H(\tau) [(Q+1)^{2-\gamma} - (Q+1)^{1-\gamma}] - 2\langle k \rangle \right] \\
&= (1/2 - r_c) \left[H(\tau) [Q^{2-\gamma} + (1-\gamma) Q^{1-\gamma}] - 2\langle k \rangle \right]. \tag{11}
\end{aligned}$$

The coagulation processes;

$$\begin{aligned}
 & -2r_c \sum_{k_a=0}^Q \sum_{k_t=Q-k_a+1}^{\infty} k_a (k_a^{\alpha} k_t^{\beta} + k_a^{\beta} k_t^{\alpha}) H(\tau) k_a^{-\gamma} H(\tau) k_t^{-\gamma} \\
 & = -2r_c [H(\tau)]^2 \int_0^Q dk_a \int_{Q-k_a+1}^{\infty} dk_t k_a (k_a^{\alpha} k_t^{\beta} + k_a^{\beta} k_t^{\alpha}) (k_a k_t)^{-\gamma} \\
 & \approx -2r_c [H(\tau)]^2 \int_0^1 Q dx \int_{1-x}^{\infty} Q dy Q x [(Qx)^{\alpha} (Qy)^{\beta} + (Qx)^{\beta} (Qy)^{\alpha}] (Qx Qy)^{-\gamma} \\
 & = -2r_c [H(\tau)]^2 Q^{3+\alpha+\beta-2\gamma} \int_0^1 dx \int_{1-x}^{\infty} dy x (x^{\alpha} y^{\beta} + x^{\beta} y^{\alpha}) (xy)^{-\gamma}. \quad (12)
 \end{aligned}$$

Combining Eqs. (10), (11) and (12), we get Eq. (13).

$$\begin{aligned}
 dM_{1,Q}/d\tau \approx & 1 - \frac{Q^{\lambda-\gamma+1}}{\langle k^{\lambda} \rangle + 1} H(\tau) \\
 & + (1/2 - r_c) \left[H(\tau) [Q^{2-\gamma} + (1-\gamma)Q^{1-\gamma}] - 2\langle k \rangle \right] \\
 & - 2r_c [H(\tau)]^2 Q^{3+\alpha+\beta-2\gamma} \int_0^1 dx \int_{1-x}^{\infty} dy x (x^{\alpha} y^{\beta} + x^{\beta} y^{\alpha}) (xy)^{-\gamma}. \quad (13)
 \end{aligned}$$

The third term of Eq. (13) can be replaced by the beta function, B , as follows;

$$\int_0^1 dx \int_{1-x}^{\infty} dy x (x^{\alpha} y^{\beta} + x^{\beta} y^{\alpha}) (xy)^{-\gamma} = \frac{B(\bar{\alpha} + 1, \bar{\beta} + 1)}{\bar{\beta}} + \frac{B(\bar{\beta} + 1, \bar{\alpha} + 1)}{\bar{\alpha}}, \quad (14)$$

where $\bar{\alpha} = \alpha - \gamma + 2$ and $\bar{\beta} = \beta - \gamma + 2$.

References

1. Smoluchowski, M.V.: Drei vortrage uber diffusion, brownsche bewegung und koagulation von kolloidteilchen. *Z. Phys.* **17**, 557 (1916)
2. Vicsek, T., Family, F.: Dynamic scaling for aggregation of clusters. *Phys. Rev. Lett.* **52**(19), 1669 (1984)
3. Van Dongen, P.G.J., Ernst, M.H.: On the occurrence of a gelation transition in Smoluchowski's coagulation equation. *J. Stat. Phys.* **44**(5–6), 785–792 (1986)
4. Takayasu, H., Nishikawa, I., Tasaki, H.: Power-law mass distribution of aggregation systems with injection. *Phys. Rev. A* **37**(8), 3110 (1988)
5. Takayasu, H.: Steady-state distribution of generalized aggregation system with injection. *Phys. Rev. Lett.* **63**(23), 2563 (1989)
6. Broide, M.L., Cohen, R.J.: Experimental evidence of dynamic scaling in colloidal aggregation. *Phys. Rev. Lett.* **64**(17), 2026 (1990)
7. Koch, W., Friedlander, S.K.: The effect of particle coalescence on the surface area of a coagulating aerosol. *J. Colloid Interface Sci.* **140**(2), 419–427 (1990)
8. Takayasu, H., Takayasu, M., Provata, A., Huber, G.: Statistical properties of aggregation with injection. *J. Stat. Phys.* **65**(3–4), 725–745 (1991)
9. Grant, S.B.: Virus coagulation in aqueous environments. *Environ. Sci. Technol.* **28**(5), 928–933 (1994)
10. Niwa, H.: School size statistics of fish. *J. Theor. Biol.* **195**(3), 351–361 (1998)
11. Davies, S.C., King, J.R., Wattis, J.A.D.: The Smoluchowski coagulation equations with continuous injection. *J. Phys. A* **32**(44), 7745 (1999)

12. Dullemond, C.P., Dominik, C.: Dust coagulation in protoplanetary disks: a rapid depletion of small grains. *Astron. Astrophys.* **434**(3), 971–986 (2005)
13. Beales, P.A., Vanderlick, T.K.: Specific binding of different vesicle populations by the hybridization of membrane-anchored DNA. *J. Phys. Chem. A.* **111**(49), 12372–12380 (2007)
14. Cho, Y.S., Kahng, B., Kim, D.: Cluster aggregation model for discontinuous percolation transitions. *Phys. Rev. E* **81**(3), 030103 (2010)
15. Miura, W., Takayasu, H., Takayasu, M.: Effect of coagulation of nodes in an evolving complex network. *Phys. Rev. Lett.* **108**(16), 168701 (2012)
16. Simon, H.A.: On a class of skew distribution functions. *Biometrika* **42**(3/4), 425–440 (1955)
17. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. *Science* **286**(5439), 509–512 (1999)
18. Takayasu, M., Sameshima, S., Ohnishi, T., Ikeda, Y., Takayasu, H., Watanabe, K.: Massive Economics Data Analysis by Econophysics Methods-The case of companies' network structure. In: *Earth Simulator Center Report No. vol. 263*, pp. 263–268 (2008)
19. Goto, H., Viegas, E., Jensen, H.J., Takayasu, H., Takayasu, M.: Appearance of unstable monopoly state caused by selective and concentrative mergers in business networks. *Sci. Rep.* **7**(1), 5064 (2017)
20. Goto, H., Takayasu, H., Takayasu, M.: Empirical analysis of firm-dynamics on japanese interfirm trade network. *Social Modeling and Simulation plus Econophysics Colloquium 2014*, In: *Proceeding of the International Conference*, pp. 195–204 (2014)
21. Moore, C., Ghoshal, G., Newman, M.E.: Exact solutions for models of evolving networks with addition and deletion of nodes. *Phys. Rev. E* **74**(3), 036121 (2006)
22. White, W.H.: A global existence theorem for Smoluchowski's coagulation equations. *Proc. Am. Math. Soc.* **80**(2), 273–276 (1980)
23. Hendriks, E.M., Ernst, M.H., Ziff, R.M.: Coagulation equations with gelation. *J. Stat. Phys.* **31**(3), 519–563 (1983)
24. Van Dongen, P.G.J., Ernst, M.H.: Dynamic scaling in the kinetics of clustering. *Phys. Rev. Lett.* **54**(13), 1396 (1985)
25. Ives, A.R., Carpenter, S.R.: Stability and diversity of ecosystems. *Science* **317**(5834), 58–62 (2007)
26. Krapivsky, P.L., Redner, S., Leyvraz, F.: Connectivity of growing random networks. *Phys. Rev. Lett.* **85**(21), 4629 (2000)
27. Ball, R.C., Connaughton, C., Jones, P.P., Rajesh, R., Zaboronski, O.: Collective oscillations in irreversible coagulation driven by monomer inputs and large-cluster outputs. *Phys. Rev. Lett.* **109**(16), 168304 (2012)
28. Matveev, S.A., Krapivsky, P.L., Smirnov, A.P., Tyrtshnikov, E.E., Brilliantov, N.V.: Oscillations in aggregation-shattering processes. *Phys. Rev. Lett.* **119**(26), 260601 (2017)